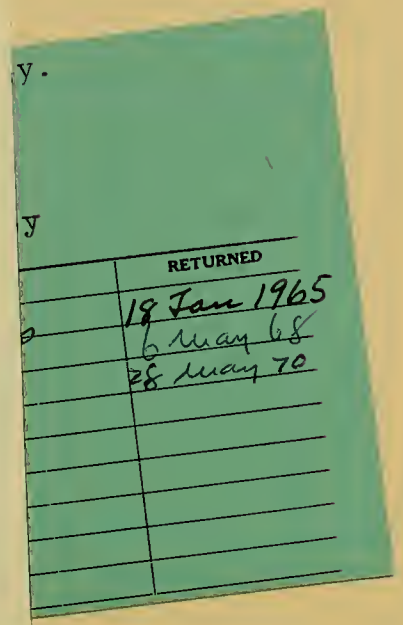
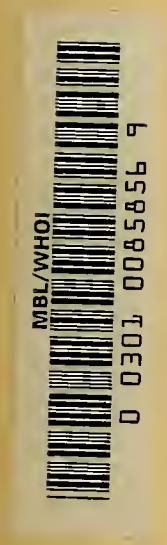

y.
y
RETURNED
18 Jan 1965
6 May 68
28 May 70

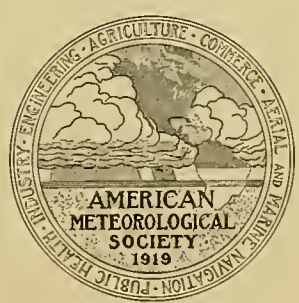


COMPENDIUM OF METEOROLOGY

COMPENDIUM
OF
METEOROLOGY

Prepared under the Direction of the
Committee on the Compendium of Meteorology
H. R. BYERS H. E. LANDSBERG H. WEXLER
B. HAURWITZ A. F. SPILHAUS H. C. WILLETT
H. G. HOUGHTON, Chairman

Edited by
THOMAS F. MALONE



AMERICAN METEOROLOGICAL SOCIETY
BOSTON, MASSACHUSETTS
1951

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U.S.A.

PREFACE

The purpose of the *Compendium of Meteorology* is to take stock of the present position of meteorology, to summarize and appraise the knowledge which untiring research has been able to wrest from nature during past years, and to indicate the avenues of further study and research which need to be explored in order to extend the frontiers of our knowledge. Perhaps it is appropriate that this stocktaking should be made as we enter the second half of the twentieth century, for surely no one can read the pages which follow without experiencing the feeling that we are on the threshold of an exciting era of meteorological history in which significant advancements are possible toward a better understanding of the physical laws which govern the workings of the atmosphere. That this progress will not be made without some difficulty is quite apparent from the number of unsolved problems which still remain as a challenge to the research worker in spite of the centuries of study which have been devoted to the nature and behavior of the atmosphere. If this book will have clarified and defined these problems, it will have fulfilled the purpose of those who planned it.

The desirability of a survey of the current state of meteorology became apparent during the years following World War II, when research effort was being greatly intensified not only in meteorology but also in other fields of pure and applied science in which the importance of meteorological factors was coming into recognition. The idea that meteorologists and atmospheric physicists from all over the world might combine their efforts to prepare a work of this nature took definite shape in 1948 when the Geophysics Research Division of the Air Force Cambridge Research Center invited the American Meteorological Society to draw up plans for a book in which specialists in the several fields of meteorology would appraise the state of knowledge in their respective specialties. The general scope of the work was decided upon by representatives of the Society and the Geophysics Research Division, and support and sponsorship was provided by the latter under Contract No. W 28-099 ac-399 with the Society. It is understood, however, that the recommendations and conclusions presented in the articles which follow do not necessarily represent those of the sponsoring agency.

Capt. H. T. Orville, U.S.N. (Ret.), president of the Society from 1948 to 1950, appointed the Committee on the Compendium of Meteorology, under the chairmanship of Professor H. G. Houghton, to organize and supervise this undertaking. This committee sought and obtained suggestions from many eminent meteorologists and, in a series of meetings in the latter part of 1948, formulated the specific nature of the present work. One hundred and two authors were commissioned in 1949 to prepare the one hundred and eight articles which comprise this book. In most cases, more than one author was invited to contribute on a single broad topic. This was done intentionally, despite some slight duplication, to insure the presentation of specialized aspects of certain general subjects and to provide ample opportunity for the exposition of different viewpoints.

A logical grouping of papers on related topics has resulted in a division of the book into twenty-five sections. Since the composition and physics of the atmosphere are fundamental to a consideration of meteorological problems, the first part of the book is concerned with the field generally referred to as physical meteorology. Then comes a discussion of the upper atmosphere—a topic in which the interests of meteorologists and physicists are now converging—followed by a section which deals with extraterrestrial effects on the atmosphere and with the meteorology of other planets. The section in which is presented a general discussion of the dynamics of the atmosphere is followed by three sections which treat various aspects of the primary, secondary, and tertiary circulations, respectively. These papers provide a logical introduction to the treatment of synoptic meteorology and weather forecasting and to the discussions of the meteorology of the tropical and polar regions and the section on climatology. Hydrometeorology, marine meteorology, biological and chemical meteorology, and atmospheric pollution are fields in which the interests of meteorologists meet those of hydrologists, oceanographers, biologists and chemists, and engineers, and these topics are treated in that order. The topics of clouds, fog, and aircraft icing have been included in a single section because of their obvious relationship to one another. The discussions of meteorological instruments and laboratory investigations and the theory of radiometeorology and microseisms and their applications to meteorological problems constitute the final sections.

References to the literature are indicated by bracketed numbers in the text. A list of references is given at the end of each article. Some care has been exercised to insure complete and accurate bibliographic information. Abbreviations for the titles of periodicals have, with a few minor modifications, followed the convenient system used in the second edition of *A World List of Scientific Periodicals*, published by the Oxford University Press in 1934. When successive entries have one or more authors in common, dashes have been used to replace the name, or names, given in the preceding entry.

It is a pleasure to acknowledge the interest and cooperation, quite apart from the financial support, of the Geo-

physics Research Division in this undertaking. The Division's library was kindly made available for use in the editorial work and the friendly counsel and suggestions of Division personnel have been most helpful. Sincere appreciation is due to Professor Hans Neuberger for carefully checking and editing the papers translated from the German. Miss Eleanor Richmond and Mr. W. Lawrence Gates have labored long and faithfully in systematizing and checking the lists of references and in assisting with the editing and proofreading. The assistance of Mr. Jean Le Corbeiller in the editorial work and proofreading is also gratefully acknowledged, as is the careful proofreading by Mrs. William Greene and Mrs. Israel Kopp. Mr. Kenneth C. Spengler, executive secretary of the American Meteorological Society, his assistant, Mrs. Holt Ashley, and their staff have capably handled all of the administrative matters connected with this work. Preparation of the illustrations for publication has been efficiently accomplished by Mr. Chester Jancewicz.

BOSTON, MASSACHUSETTS T. F. M.

August 1951

TABLE OF CONTENTS

COMPOSITION OF THE ATMOSPHERE

The Composition of Atmospheric Air *by E. Glueckauf* 3

RADIATION

Solar Radiant Energy and Its Modification by the Earth and Its Atmosphere *by Sigmund Fritz* 13
Long-Wave Radiation *by Fritz Möller* 34
Actinometric Measurements *by Anders Ångström* 50

METEOROLOGICAL OPTICS

General Meteorological Optics *by Hans Neuberger* 61
Polarization of Skylight *by Zdeněk Sekera* 79
Visibility in Meteorology *by W. E. Knowles Middleton* 91

ATMOSPHERIC ELECTRICITY

Universal Aspects of Atmospheric Electricity *by O. H. Gish* 101
Ions in the Atmosphere *by G. R. Wait and W. D. Parkinson* 120
Precipitation Electricity *by Ross Gunn* 128
The Lightning Discharge *by J. H. Hagenguth* 136
Instruments and Methods for the Measurement of Atmospheric Electricity *by H. Israël* 144
Radioactivity of the Atmosphere *by H. Israël* 155

CLOUD PHYSICS

On the Physics of Clouds and Precipitation *by Henry G. Houghton* 165
Nuclei of Atmospheric Condensation *by Christian Junge* 182
The Physics of Ice Clouds and Mixed Clouds *by F. H. Ludlam* 192
Thermodynamics of Clouds *by Fritz Möller* 199
The Formation of Ice Crystals *by Ukichiro Nakaya* 207
Snow and Its Relationship to Experimental Meteorology *by Vincent J. Schaefer* 221
Relation of Artificial Cloud-Modification to the Production of Precipitation *by Richard D. Coons and Ross Gunn* 235

THE UPPER ATMOSPHERE

General Aspects of Upper Atmospheric Physics *by S. K. Mitra* 245
Photochemical Processes in the Upper Atmosphere and Resultant Composition *by Sidney Chapman* 262
Ozone in the Atmosphere *by F. W. Paul Götz* 275
Radiative Temperature Changes in the Ozone Layer *by Richard A. Craig* 292
Temperatures and Pressures in the Upper Atmosphere *by Homer E. Newell, Jr.* 303
Water Vapour in the Upper Air *by G. M. B. Dobson and A. W. Brewer* 311
Diffusion in the Upper Atmosphere *by Heinz Lettau* 320
The Ionosphere *by S. L. Seaton* 334
Night-Sky Radiations from the Upper Atmosphere *by E. O. Hulburt* 341
Aurorae and Magnetic Storms *by L. Harang* 347
Meteors as Probes of the Upper Atmosphere *by Fred L. Whipple* 356
Sound Propagation in the Atmosphere *by B. Gutenberg* 366

COSMICAL METEOROLOGY

Solar Energy Variations As a Possible Cause of Anomalous Weather Changes *by Richard A. Craig and H. C. Willett* 379
The Atmospheres of the Other Planets *by S. L. Hess and H. A. Panofsky* 391

DYNAMICS OF THE ATMOSPHERE

The Perturbation Equations in Meteorology *by B. Haurwitz* 401
The Solution of Nonlinear Meteorological Problems by the Method of Characteristics *by John C. Freeman* 421
Hydrodynamic Instability *by Jacques M. Van Mieghem* 434
Stability Properties of Large-Scale Atmospheric Disturbances *by R. Fjørtoft* 454

The Quantitative Theory of Cyclone Development *by E. T. Eady* 464
Dynamic Forecasting by Numerical Process *by J. G. Charney* 470
Energy Equations *by James E. Miller* 483
Atmospheric Turbulence and Diffusion *by O. G. Sutton* 492
Atmospheric Tides and Oscillations *by Sydney Chapman* 510
Application of the Thermodynamics of Open Systems to Meteorology *by Jacques M. Van Mieghem* 531

THE GENERAL CIRCULATION

The Physical Basis for the General Circulation *by Victor P. Starr* 541
Observational Studies of General Circulation Patterns *by Jerome Namias and Philip F. Clapp* 551
Applications of Energy Principles to the General Circulation *by Victor P. Starr* 568

MECHANICS OF PRESSURE SYSTEMS

Extratropical Cyclones *by J. Bjerknes* 577
The Aerology of Extratropical Disturbances *by E. Palmén* 599
Anticyclones *by H. Wexler* 621
Mechanism of Pressure Change *by James M. Austin* 630
Large-Scale Vertical Velocity and Divergence *by H. A. Panofsky* 639
The Instability Line *by J. R. Fulks* 647

LOCAL CIRCULATIONS

Local Winds *by Friedrich Defant* 655
Tornadoes and Related Phenomena *by Edward M. Brooks* 673
Thunderstorms *by Horace R. Byers* 681
Cumulus Convection and Entrainment *by James M. Austin* 694

OBSERVATIONS AND ANALYSIS

World Weather Network *by Athelstan F. Spilhaus* 705
Models and Techniques of Synoptic Representation *by John C. Bellamy* 711
Meteorological Analysis in the Middle Latitudes *by V. J. Oliver and M. B. Oliver* 715

WEATHER FORECASTING

The Forecast Problem *by H. C. Willett* 731
Short-Range Weather Forecasting *by Gordon E. Dunn* 747
A Procedure of Short-Range Weather Forecasting *by Robert C. Bundgaard* 766
Objective Weather Forecasting *by R. A. Allen and E. M. Vernon* 796
General Aspects of Extended-Range Forecasting *by Jerome Namias* 802
Extended-Range Weather Forecasting *by Franz Baur* 814
Extended-Range Forecasting by Weather Types *by Robert D. Elliott* 834
Verification of Weather Forecasts *by Glenn W. Brier and Roger A. Allen* 841
Application of Statistical Methods to Weather Forecasting *by George P. Wadsworth* 849

TROPICAL METEOROLOGY

Tropical Meteorology *by C. E. Palmer* 859
Equatorial Meteorology *by A. Grimes* 881
Tropical Cyclones *by Gordon E. Dunn* 887
Aerology of Tropical Storms *by Herbert Riehl* 902

POLAR METEOROLOGY

Antarctic Atmospheric Circulation *by Arnold Court* 917
Arctic Meteorology *by Herbert G. Dorsey, Jr.* 942
Some Climatological Problems of the Arctic and Sub-Arctic *by F. Kenneth Hare* 952

CLIMATOLOGY

Climate—The Synthesis of Weather *by C. S. Durst* 967
Applied Climatology *by Helmut E. Landsberg and Woodrow C. Jacobs* 976
Microclimatology *by Rudolf Geiger* 993

Geological and Historical Aspects of Climatic Change *by C. E. P. Brooks* 1004
Climatic Implications of Glacier Research *by Richard Foster Flint* 1019
Tree-Ring Indices of Rainfall, Temperature, and River Flow *by Edmund Schulman* 1024

HYDROMETEOROLOGY

Hydrometeorology in the United States *by Robert D. Fletcher* 1033
The Hydrologic Cycle and Its Relation to Meteorology—River Forecasting *by Ray K. Linsley* 1048

MARINE METEOROLOGY

Large-Scale Aspects of Energy Transformation over the Oceans *by Woodrow C. Jacobs* 1057
Evaporation from the Oceans *by H. U. Sverdrup* 1071
Forecasting Ocean Waves *by W. H. Munk and R. S. Arthur* 1082
Ocean Waves as a Meteorological Tool *by W. H. Munk* 1090

BIOLOGICAL AND CHEMICAL METEOROLOGY

Aerobiology *by Woodrow C. Jacobs* 1103
Physical Aspects of Human Bioclimatology *by Konrad J. K. Buettner* 1112
Some Problems of Atmospheric Chemistry *by H. Cauer* 1126

ATMOSPHERIC POLLUTION

Atmospheric Pollution *by E. Wendell Hewson* 1139

CLOUDS, FOG, AND AIRCRAFT ICING

The Classification of Cloud Forms *by Wallace E. Howell* 1161
The Use of Clouds in Forecasting *by Charles F. Brooks* 1167
Fog *by Joseph J. George* 1179
Physical and Operational Aspects of Aircraft Icing *by Lewis A. Rodert* 1190
Meteorological Aspects of Aircraft Icing *by William Lewis* 1197

METEOROLOGICAL INSTRUMENTS

Instruments and Techniques for Meteorological Measurements *by Michael Ference, Jr.* 1207
Aircraft Meteorological Instruments *by Alan C. Bemis* 1223

LABORATORY INVESTIGATIONS

Experimental Analogies to Atmospheric Motions *by Dave Fultz* 1235
Model Techniques in Meteorological Research *by Hunter Rouse* 1249
Experimental Cloud Formation *by Sir David Brunt* 1255

RADIOMETEOROLOGY

Radar Storm Observation *by Myron G. H. Ligda* 1265
Theory and Observation of Radar Storm Detection *by Raymond Wexler* 1283
Meteorological Aspects of Propagation Problems *by H. G. Booker* 1290
Sferics *by R. C. Wanta* 1297

MICROSEISMS

Observations and Theory of Microseisms *by B. Gutenberg* 1303
Practical Application of Microseisms to Forecasting *by James B. Macelwane, S. J.* 1312

INDEX 1317

COMPOSITION OF THE ATMOSPHERE

The Composition of Atmospheric Air *by E. Glueckauf* 3

THE COMPOSITION OF ATMOSPHERIC AIR

By E. GLUECKAUF

Atomic Energy Research Establishment, Harwell, England

For many reasons it is desirable to have a complete knowledge of the composition of the atmosphere as regards both the molecular species present and their absolute quantities. This applies not only to the main components, but also to rare species of polyatomic molecules whose importance in the radiation balance is often quite out of proportion to their actual quantity. Any observed variation in the composition of air—with time, with geographical location, with height, with the seasons, or with meteorological conditions—seriously affects our conception of the processes in the atmosphere. It follows from this that the present article must lay its emphasis on facts in order to see where our knowledge is still inadequate.

SURVEYS OF VARIATIONS IN THE COMPOSITION OF THE ATMOSPHERE

Oxygen (O_2). The first "international" investigation of the O_2 content with adequate equipment was carried out as early as 1852 by Regnault [18] (see Table I). He concluded that atmospheric air presents only small variations.

TABLE I. SURVEY OF OXYGEN CONTENT
(*After Regnault* [18])

Location	Average O_2 (per cent)
Montpellier	20.95 ± 0.00
Lyon	20.94 ± 0.01
Normandy	20.95
Berlin	20.96 ± 0.01
Mediterranean*	20.94 ± 0.01
Atlantic	20.94 ± 0.01
East Indian Ocean	†
Arctic Ocean	20.91 ± 0.01
Paris	20.96 ± 0.01
Average	20.944

* Two samples from Algiers have been omitted.
† Greatly varying, mostly low.

Sixty years later, in a survey involving many hundreds of precision analyses, Benedict [c. 16][1] proved conclusively that during a period of about two years, which involved a great variety of weather conditions, no material variations occurred in the O_2 content of air at the Nutrition Laboratory of the Carnegie Institution of Washington. A statistical analysis of Benedict's data gives a standard deviation of ±0.006 per cent, while the repeated analysis of a bottled air sample gave a standard deviation of only ±0.0025 per cent. It is likely that the higher standard deviation of the former originated from minute variations in the sampling procedure.

Similar precision was obtained by Krogh [c. 16] whose analyses from October 4, 1917, to January 25, 1918, showed a standard deviation of ±0.005 per cent for O_2, ±0.0025 per cent for CO_2, and ±0.0020 per cent for O_2 plus CO_2, all in uncontaminated air. The absolute values for O_2 and CO_2 in dry air obtained by Krogh after careful calibrations are shown in Table II. This last figure—or the corresponding figure of 79.0215 per cent for the content of "atmospheric nitrogen" (*i.e.*, N_2 + rare gases)—is considered by Krogh to be a geophysical constant which does not vary more than indicated by the standard deviation observed (±0.0020 per cent).

TABLE II. OXYGEN AND CARBON DIOXIDE VALUES IN DRY AIR
(*After Krogh* [c. 16])

Constituent	Content (per cent)	
	Experiment I	Experiment II
CO_2	0.0305	0.0300
O_2	20.9480	20.9485
$O_2 + CO_2$	20.9785	20.9785

The absolute value of the O_2 content of air obtained by Benedict, taking into account a small correction resulting from the formation of CO from the pyrogallate, was later computed by Benedict and by Haldane as 20.952 per cent and by Krogh [c. 16] as 20.954 per cent.

While Benedict's experiments prove conclusively that only trifling changes occur in air observations at one locality, it is more difficult to show that there is uniformity over all the earth's surface. For such a survey it is necessary to bottle air samples, and changes of one or two parts in ten thousand can easily occur during transit of the air sample. Table III gives a summary of Benedict's analyses of air of various origins. It is not likely that significance can be attached to these small variations.

TABLE III. SURVEY OF OXYGEN CONTENTS
(*After Benedict* [c. 16])

Origin of air	No. of samples	O_2 (corrected) (per cent)
Boston, Mass.	212	20.952
Ocean air (Montreal to Liverpool)	7	20.950
Ocean air (Genoa to Boston)	36	20.946
Pikes Peak*	5	20.945

* The data for August 14, 1911 which showed abnormally low values have been omitted.

1. A "c" is used in front of reference numbers to indicate that detailed reference is given in the paper cited.

An oxygen deficiency in antarctic air, claimed by Lockhart and Court [c. 12], cannot yet be regarded as well established, since no check analyses with normal air were carried out.

In spite of Paneth's conservative estimate [16] that the second decimal is still not exactly known, the author feels inclined, in view of the agreement of the two best surveys with a recent redetermination by M. Shepherd [c. 16] which gave 20.945 per cent, to recommend an absolute value of 20.946 ± 0.002 per cent as the most likely figure for the O_2 content of uncontaminated air, in combination with an average CO_2 value of 0.033 per cent.

Apart from possible minor changes resulting from the greater solubility of O_2 than of N_2 in ocean water, all major variations of the O_2 content must result from the combustion of fuel, from the respiratory exchange of organisms, or from the assimilation of CO_2 in plants. The first process does not result in more than local changes of the O_2 content, while the latter two processes, though locally altering the CO_2/O_2 ratio, leave their sum unchanged.

Carbon Dioxide (CO_2). Though the extensive investigations of Benedict and of Krogh suggest that the CO_2 content of atmospheric air over land does not vary except within very narrow limits, significant variations of the CO_2 content have been observed by workers both before and after them.

In particular, Callendar [5] has drawn attention to an increase in the CO_2 content during the last fifty years which is best demonstrated in Fig. 1. This in-

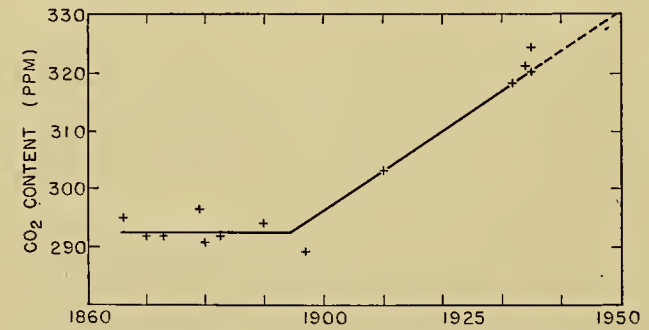


Fig. 1.—Increase of CO_2 in air. (*After Callendar* [5].)

crease in the total atmosphere of about 30 ppm (parts per million) represents a quantity of CO_2 (2×10^{11} tons) which is approximately equal to the amount resulting from the combustion of fuels produced during this period. This implies that not much of this excess CO_2 has been lost to the ocean water, an assumption which is justified in view of the fact that, apart from a thin agitated surface layer, the transport of CO_2 inside the water proceeds by diffusion and is very slow.

Variations of the CO_2 Content over the Sea. The variations of CO_2 over the sea are now well understood (Buch, Wattenberg [c. 5]). Because of an excess of strongly basic cations over strongly acid anions in sea water, CO_2 is soluble in sea water not only as dissolved CO_2, but also in the form of carbonate and bicarbonate ions, the quantities being roughly of the order 1:8:150. The result of this is that one litre of sea water contains about 150 times as much CO_2 as the same volume of air.

For a given content of total CO_2, the equilibrium pressure varies considerably with the water temperature. To give an example: For water with a chloride content of 1.95 per cent and a total CO_2 of 2.07×10^{-3} mol l^{-1}, the CO_2 pressures[2] in air at equilibrium at 0C, 10C, 20C, and 30C are 1.6, 2.5, 3.6, and 5.1 $\times$ 10^{-4} atm, respectively. Thus, far from having an equalizing effect on the CO_2 content of the air, as was believed during the last century and the earlier part of this century, changes in the surface temperature of the sea upset the otherwise comparatively constant CO_2 content of air. This explains the low values of the CO_2 content found near the polar regions. The lowest value (1.52×10^{-4} atm) was observed near Spitsbergen by Buch [c. 5]. This value corresponds roughly to the equilibrium value at 0C.

In the most northerly regions, particularly over polar ice, the CO_2 content is again normal, a feature which was explained by Buch on the basis of Bjerknes' scheme of the air circulation over the Atlantic. According to the latter, air in the Arctic is more or less constantly descending, and since it has by-passed at great height the cold-water regions on its way from the south, its CO_2 content would be expected to be near that of the temperate zones. Buch found 2.57 and 2.91 $\times 10^{-4}$ atm.

Equally complicated is the situation in regions where masses of water rise to the surface from greater depths. The CO_2 content of sea water, after falling slightly in the first 50 m below the surface (due to CO_2 assimilation), rises to a maximum at about 500-m depth where the CO_2 pressure may be as much as 11×10^{-4} atm (obviously due to decay of organic matter). If these CO_2-rich water masses rise to the surface (as observed near the west coast of Africa), the CO_2 content of air may locally rise to 7×10^{-4} atm. Similarly high values have been observed by Krogh [14] in the vicinity of West Greenland, and by Moss [c. 14] at latitude 82°27′N, though in these two cases the origin of the CO_2 was not traced, and the effect may possibly, but not necessarily, be spurious.

These phenomena apparently do not affect the air masses to a very great depth. Near Petsamo, Finland, arctic air (range 297 to 313 ppm) differs unmistakably from continental and tropical air (range 319 to 335 ppm), so that in this region the CO_2 content can serve as an indicator for the origin of the air masses. However, these differences become smaller as we go farther south. Thus the difference is still appreciable at Kew, England, where, on the average, subtropical air contains 19 ppm more CO_2 than polar and maritime air; but the mean difference is only 8 ppm near Dieppe, France, and Gembloux, Belgium [c. 5].

---

2. The CO_2 pressure in atmospheres is very nearly, but not quite, identical with the "parts per volume" unit, *i.e.*, 10^{-4} atm $\simeq$ 100 ppm.

The contaminations of CO_2 caused by large towns are also fairly localised. At Kew, about 6 miles west of the centre of London, on the average only 27 ppm more CO_2 are found in easterly than in westerly winds.

Argon (*A*). The constancy of the composition of air with respect to its minor constituents has been investigated in the cases of *A* and *He*. The former was investigated by Moissan [c. 16], who, after chemical removal of O_2, N_2, and CO_2 by calcium metal, measured the remaining rare gases and obtained for the *A* content the figures shown in Table IV. The standard deviation of these analyses (excepting the value of 0.949 over the Atlantic Ocean which must be considered erroneous) is ±0.002 per cent, or 0.2 per cent of the *A* content, and within these limits there are no significant variations.

TABLE IV. SURVEY OF ARGON CONTENTS OF AIR
(*After Moissan* [c. 16])

Location	*A* (per cent)
Odessa	0.935
St. Petersburg (Leningrad)	0.933
Athens	0.935
Ionian Sea (37°N, 15°E)	0.936
Vienna	0.938
Berlin	0.932
Venice	0.936
Mt. Blanc	0.935
Paris	0.934
London	0.933
Atlantic Ocean (37°N, 24°W)	0.932
Atlantic Ocean (43°N, 22°W)	(0.949)
Average (omitting the last value)	0.9343 ± 0.0006

Helium (*He*). Similar results were also obtained with *He* in spite of the fact that vast quantities of *He* constantly escape from the earth's crust, particularly from oil fields in the United States. It has been estimated that between eight and thirty million cubic metres of *He* are generated annually by radioactive processes. The amounts of *He* added to the atmosphere in this way are balanced by losses of *He* into the void of the universe, because *He*, owing to its lightness, is not permanently retained by the gravitational field of the earth.

A world survey covering all continents and oceans from the Arctic to the Antarctic [12] showed no significant deviations even comparatively near oil fields in the United States. The value obtained was 5.239 ± 0.002 ppm with a standard deviation of ±0.008 ppm [11]. It is apparent that the turbulence of the troposphere quickly eliminates any nonuniformity resulting from localised *He* discharge.

ATMOSPHERIC GASES OF CONSTANT PERCENTAGE

While surveys over large areas have been carried out only for the four gases O_2, CO_2, *A*, and *He*, the constancy in the total percentage of these gases makes it plausible that other gases too must have a constant total percentage, unless they are subject to vapour-pressure equilibria (as is H_2O) or to radiation equilibria (as is O_3), or are simply the result of some industrial activity (as are SO_2 and I_2).

Nitrogen (N_2). For the analysis of N_2 no direct precision method has been discovered. However, as the sum of nitrogen and rare gases (called "atmospheric nitrogen" by Krogh) has a constant value of 79.0215 per cent, and as this sum is constant to at least ±0.002 per cent, N_2 too must be constant to the same degree.

Neon (*Ne*). The most reliable data for the *Ne* content of air are (1) a single analysis by Watson [c. 11] who found 18.2 ppm, (2) three analyses by Glueckauf [11] who found a mean of 18.21 ± 0.04 ppm, and (3) recent analyses by Chackett, Paneth, and Wilson [6] who found 18.1 ± 0.08 ppm.

Krypton (*Kr*) and Xenon (*Xe*). Of the figures in the literature [c. 16], those by Moureu and Lepape and by Damköhler appear to be the most reliable. As these values are accurate only to ±10 per cent of their value, the *Kr* and *Xe* contents have recently been redetermined with a much higher accuracy by the author of this article. The figures in parts per million by volume are:

Moureu + Lepape (1926) *Kr:* 1.0 ± 0.1, *Xe:* 0.09 ± 0.01
Damköhler (1935) *Kr:* 1.08 ± 0.1, *Xe:* 0.08 ± 0.01
Glueckauf + Kitt (unpublished) *Kr:* 1.14 ± 0.01, *Xe:* 0.087 ± 0.001

Nitrous Oxide (N_2O). The presence of nitrous oxide in atmospheric air was discovered by Adel [15, Chap. 10] by means of an absorption band at 7.8 μ in the solar spectrum. Since then further atmospheric absorption bands have been discovered at 3.9 μ, 4.5 μ, and 8.6 μ. The recent chemical analysis by Slobod and Krogh [20] of N_2O in air at ground level gave a value of 0.5 ± 0.1 ppm, which is in agreement with the spectroscopic data.

Methane (CH_4). Methane was found by spectroscopic identification of its absorption band in sunlight modified by the passage through the earth's atmosphere over Columbus, Ohio (Migeotte), over Flagstaff, Arizona (Adel), and over Pontiac, Michigan (McMath Observatory) [15, Chap. 10]. From the latter data the CH_4 content of air has been estimated to be about 1.2 ppm (by weight), that is, about 2.2×10^{-6} by volume. It is possible that this figure is somewhat high, as during the process of distillation of air it is found that the *Kr* and *Xe* fraction contains only an amount of CH_4 roughly equal to that of these gases (1.2×10^{-6} by volume).

We are faced with the question of the origin of this CH_4 which is constantly destroyed by the ozone in atmospheric air. As a constant source of this CH_4 we may consider either decay of biological products, or gas escaping from oil wells, or both. The question of the relative extent of these two processes can be decided by determining the content of radiocarbon (^{14}C) in the CH_4 of air. Methane from biological sources contains 0.95×10^{-12} g of ^{14}C per g of *C*, while mineral CH_4 is inactive.

At the suggestion of the author, Prof. F. W. Libby at the University of Chicago analysed the ^{14}C content of

atmospheric methane and found it to coincide with that of biological methane. The analysis also places an upper limit of about 200 years on the mean lifetime of the methane in the atmosphere and suggests an annual production of upwards of 10^7 tons of methane from biological sources.

Hydrogen (H_2). The H_2 content of air is known only approximately. Paneth [16] concludes that it is a constant constituent of atmospheric air and that its amount can be assumed to be about 5×10^{-7} by volume. Recent analyses by the author and G. P. Kitt gave varying amounts of H_2 upwards of 0.4 ppm, and investigations are in progress to see whether these variations occur in the free air or are due to local contamination.

Summary. The figures believed to be most reliable for the constituents of dry air are listed in Table V.

TABLE V. NONVARIABLE COMPONENTS OF ATMOSPHERIC AIR

Constituent	Content (per cent)	Content (ppm)
N_2	78.084 ± 0.004	
O_2	20.946 ± 0.002	
CO_2*	0.033 ± 0.001	
A	0.934 ± 0.001	
Ne		18.18 ± 0.04
He		5.24 ± 0.004
Kr		1.14 ± 0.01
Xe		0.087 ± 0.001
H_2		0.5
CH_4		2
N_2O		0.5 ± 0.1

* Extrapolated to 1950 according to Fig. 1.

CONSTITUENTS OF VARIABLE CONCENTRATIONS
(Excluding Water Vapour)

A rough survey of some of these constituents is given in Table VI.

TABLE VI. VARIABLE CONSTITUENTS OF DRY ATMOSPHERIC AIR

Constituent	Origin	Proportion in ground air (range)
O_3	Ultraviolet radiation	0 to 0.07 ppm (summer) 0 to 0.02 ppm (winter)
SO_2	Industrial	0 to 1 ppm
NO_2	Industrial	0 to 0.02 ppm
CH_2O	Biological or oxydation of CH_4	Uncertain
I_2	Industrial	Up to 10^{-4} g m^{-3}
$NaCl$	Sea spray	Order of 10^{-4} g m^{-3}
NH_3	Industrial	0 to trace
CO	Industrial	0 to trace

Ozone (O_3). The bulk of atmospheric O_3 is contained in the stratosphere, where it is produced by the ultraviolet radiation from the sun. The problems connected with its production and occurrence there form the subject of a separate article in this Compendium.[3] We shall deal here only with observations of O_3 near the surface.

3. Consult "Ozone in the Atmosphere" by F. W. P. Götz, pp. 275–291.

Methods of Determination. The main difficulty in the determination of O_3 in atmospheric air lies in the fact that simple chemical reactions are not specific for O_3 and that gases like H_2O_2, NO_2, and SO_2 interfere with the chemical determination, the first two by increasing, the last by decreasing the analytical result. However, these gases occur only in the vicinity of human habitation. On the other hand, the spectroscopic investigation of surface air [7, 13], though accurate, is so time-consuming (due to the necessary photometry of the spectrograms) that it does not lend itself to routine observations. This also applies to the chemical method of Edgar and Paneth [c. 12] which relies on the separation of O_3 from all other gases by low-temperature adsorption on silica gel. However, two accurate methods have been evolved which give reliable results in a comparatively short time and thus make possible large numbers of determinations under quickly changing meteorological conditions. (See V. H. Regener [c. 17] and Glueckauf, Heal, Martin, and Paneth [c. 12].)

The results obtained so far can generally be explained on the basis that O_3 which is produced in the higher regions—mostly in the stratosphere and possibly some just below the tropopause—reaches the ground level through the turbulence of the air, and on its way down is gradually diminished and eventually destroyed by oxidisable materials of an organic and inorganic character.

Diurnal Variations. On days with little turbulence, the ground O_3 found during the day usually disappears at nightfall because of the increased stability of the air, but it remains unaffected at higher wind velocities.

Annual Variations. Pronounced maxima (7×10^{-8} by volume about May) and minima (2×10^{-8} about November) of the ground O_3 have been found by many observers. The fact that these annual variations are greater than those of the total O_3 may be due to the greater instability of the atmosphere during the summer months. There are some indications that at higher latitudes (*e.g.*, Abisko, Lapland) the high summer values of ground O_3 appear later, if at all.

Geographic Variations. Next to nothing is known about the geographical distribution of O_3 over continents and oceans. Almost all determinations have been carried out between the latitudes 45°N to 68°N over land. Usher and Rao [22], however, reported the absence of ground O_3 in India. This result may not be reliable, but there is obviously wide scope for further investigations.

Ozone during Depressions. The usually high wind velocity and atmospheric turbulence during depressions result in high O_3 contents (subject to the seasonal variations). However, low values of O_3 were found at Durham, England [10], even at high wind velocities in advance of warm fronts, and during occlusions of the cold-front type. Apparently under these conditions inversions are formed which restrict the turbulent interchange of air masses near the ground with the O_3 produced in higher regions. It is to be expected that such phenomena will be greatly reduced in regions with low industrial contamination (*e.g.*, over the

oceans), but no experimental data are available to check this argument.

Ozone during Thunderstorms. Dobson [c. 10] has observed many cases where the total O_3 increases during thunderstorms and during the passage of thunderclouds. There can be little doubt that these changes occur in the tropospheric air and are caused by electric phenomena. A continuous measurement [10] during a thunderstorm gave no indication of any abnormal increase of O_3 near the ground and on this occasion, at least, the O_3-bearing air masses did not reach ground level. Other observers [13], however, found abnormally high O_3 values in ground air on some thundery days.

Vertical Distribution of Ozone in the Troposphere. An increase of the O_3 content with height in the troposphere is shown by the data of Chalonge, Götz, and Vassy [7], who found an average of 1.7×10^{-8} at Lauterbrunnen, Switzerland (800 m), and 3.0×10^{-8} at Jungfraujoch, Switzerland (3450 m).

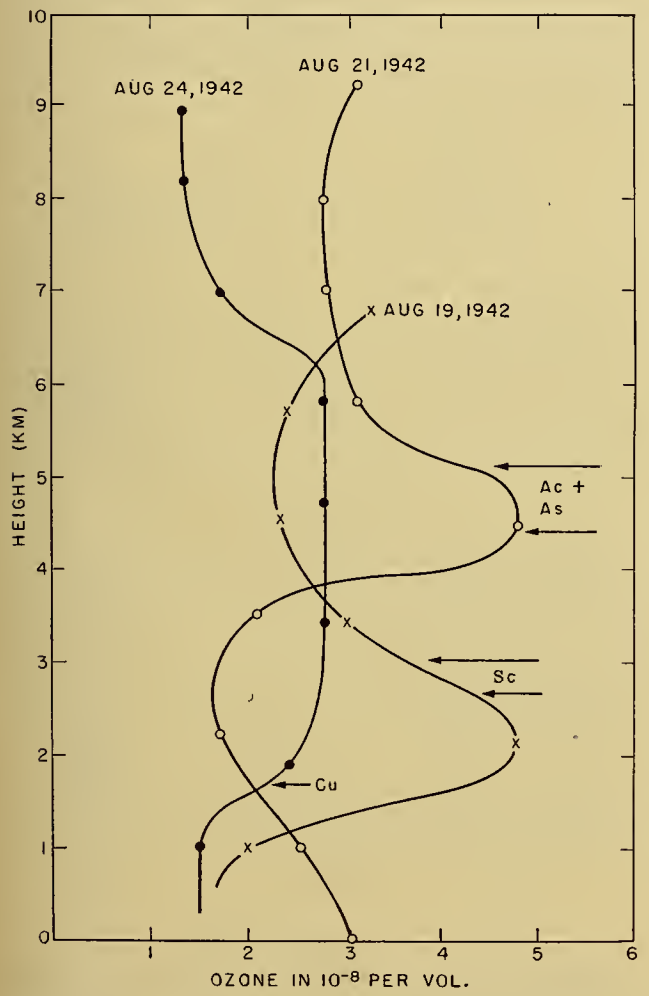


Fig. 2.—Vertical distribution of ozone in the troposphere. (*After Ehmert* [c. 17].)

Ozone determinations made in aircraft by Ehmert [c. 17] show a variety of features (see Fig. 2). In one case the air above cloud level has O_3 contents which, if measured in volume per volume, are independent of height. This would be expected if the mixing ratio is kept constant by turbulence. Much less obvious is the fact that the O_3 content is high in cloudy regions and reaches a maximum just below cloud level. Regener [17] explains these maxima as produced by advection, which may sometimes be the case. However, the repeated occurrence of stratified clouds near such an O_3 maximum seems to indicate that under these conditions O_3 may be produced by phenomena of an electrical nature.

Sulphur Dioxide (SO_2). The quantities of SO_2 found in air vary greatly according to the nearness of towns and the turbulence of the air. To give a few examples: An average of 0.033 ppm was found at the Boyce Thompson Institute (about 15 miles from New York City); at Chicago an average of from 0.06 to 0.27 ppm was found in residential districts and from 0.4 to 0.5 ppm in manufacturing districts. In the presence of O_3, air may be expected to be free of SO_2.

Nitrogen Dioxide (NO_2). No systematic determinations of this constituent seem to have been made. This is largely because of the small quantity present in air and the difficulty of analysis. A very reliable method of NO_2 analysis in air, based on the use of 2:4 xylen-l-ol was used by Edgar and Paneth [c. 12]. From the fact that on a large number of days the NO_2 content found in this way was below the threshold of sensitivity (5×10^{-10}), one is tempted to conclude that NO_2 is not a normal constituent of air. This may be due essentially to its high solubility in water. In large towns, however, where NO_2 occurs as a by-product from the combustion of nitrogenous matter, varying quantities up to 2×10^{-8} were found by a number of authors.

Ammonia (NH_3). The presence of minute amounts of NH_3 in atmospheric air over Michigan has been claimed by Mohler, Goldberg, and McMath [c. 15, Chap. 10] from absorption bands in the 2-μ region. But, as Migeotte and Chapman [c. 15, Chap. 10] have pointed out, the 10.5-μ fundamental band of NH_3 is much more suitable for testing the presence of atmospheric NH_3, and no evidence could be found in this region of absorption either above Flagstaff, Arizona, or above Columbus, Ohio. Moreover, NH_3 is very soluble in water and thus is not likely to be retained in the air for any lengthy period.

Carbon Monoxide (CO). This gas was observed spectroscopically by Migeotte [c. 1] over Columbus, Ohio, as well as on the Jungfraujoch (3580 m altitude) but, as none could be found by Adel over Flagstaff, Arizona [1], it is not yet certain whether it is a permanent constituent of the atmosphere.

THE UPPER ATMOSPHERE

The composition of air in the upper atmosphere is of considerable interest as an indicator of whether or not large-scale mixing of air masses takes place in the stratosphere. It is often assumed that the absence of a systematic temperature gradient in the stratosphere is incompatible with large-scale mixing. In the absence of turbulent mixing, however, diffusive separation of the gases should take place and the lighter gases should

become relatively more abundant with increasing height.

The most direct method of finding the level where diffusive separation begins is the chemical analysis of air samples taken at great heights, since the gravitational equilibrium should cause the O_2 content to decrease by 2 per cent per kilometre and the *He* content to increase by 14 per cent per kilometre. Air samples from the stratosphere have been obtained by manned and unmanned balloon flights. The results of these analyses are shown in Table VII.

TABLE VII. COMPOSITION OF AIR IN THE STRATOSPHERE

Height (km)	Helium variation (per cent)	Oxygen variation (per cent)	Source of data*
9–17		0	*c*
14.5		−0.14	*d*
16.5	+0.5 ± 0.5		*e*
18.0	+0.35 ± 0.1		*e*
18.5	+0.7 ± 0.3	0	*e,a*
18.5		−0.38	*d*
19.0	+0.55 ± 0.15	−0.24	*e,d*
21.0	+6.9 ± 0.7		*e*
21.5		−0.24	*b*
22.0	+4.1 ± 0.2		*e*
22.0	+1.95 ± 0.15		*e*
22.5	+5.1 ± 0.6		*e*
22.5	+1.9 ± 0.3		*e*
23.5	+4.0 ± 0.3		*e*
23.5	+0.3 ± 0.15		*e*
24		−0.86	*d*
25	+2.1 ± 0.3		*e*
28–29		−2.5	*d*

* Manned balloon flights
(*a*) Prokofiev, 1933
(*b*) *Explorer II*, 1936
Unmanned balloon flights
(*c*) Lepape and Colange, 1935
(*d*) E. Regener, 1936
(*e*) Glueckauf and Paneth, 1946

It appears from this table that there is no significant change in either the *He* content or the O_2 content below 20 km. The *He* surplus observed between 21 and 25 km averages 3.3 per cent, an enrichment which should be found at the top of a column of still air only 250 m high. The biggest O_2 deficit, at about 28 km, corresponds to a column of still air only 1100 m high. It is therefore apparent that at the heights reached by sounding balloons there is sufficient turbulence to reduce the changes in the *He* content to about $\frac{1}{40}$ of what one would expect from a gravitational equilibrium starting at the tropopause. The recent analysis by Chackett, Paneth, and Wilson [6] of three air samples collected by a V-2 rocket from a height of 50 to 70 km, gave variations of +0.3 to −4 per cent for *He*, variations of −0.3 to −0.7 per cent for *Ne*, and variations of −0.4 to +1.0 per cent for *A*. These results make it certain that no diffusive separation is maintained even at these great heights.

ISOTOPIC COMPOSITION OF THE ATMOSPHERIC GASES

Increased attention to the isotopic composition of the atmospheric gases is likely to throw light on a number of problems. The composition in most cases is very similar to that found in the same atomic species in other parts of the earth's crust[4] (see Table VIII).

Water Vapour. Because of differences in the vapour pressures, mainly of $^1H_2{}^{16}O$, $^1H_2{}^{18}O$, and $^1H^2H^{16}O$, the density of atmospheric water should be slightly less than that in the oceans from which it originates. This was confirmed by Riesenfeld and Chang [19] who found a deficit of 3.8 γ in the density of snow water, and of 2.5 γ for rain water ($1\ \gamma = 10^{-6}$ g ml^{-1}). These figures are approximately what would be expected from the known vapour pressures.

Oxygen. The differences in the composition of the oxygen in (liquid) water, gaseous oxygen, and carbon dioxide are much smaller, the densities being in the ratio 1:1.0000073:1.0000116 [23]. From this follows a slight enrichment of the ^{18}O isotope in the ratio 1:1.0033:1.0053. The difference of the oxygen density for atmospheric CO_2 and for that of carbonate rocks is negligible [9].

TABLE VIII. ISOTOPIC COMPOSITION OF THE MAIN ATMOSPHERIC GASES

Element	Atomic mass numbers (in parentheses) and percentages (in italics) of isotopic species
H in H_2O	(1) *99.98* (2) *0.02*
He	(3) *1.1* × 10^{-4} (4) *100*
C in CO_2	(12) *98.9* (13) *1.1* (14) *0.95* × 10^{-12}
N	(14) *99.62* (15) *0.38*
O	(16) *99.757* (17) *0.039* (18) *0.204*
Ne	(20) *90.00* (21) *0.27* (22) *9.73*
A	(36) *0.307* (38) *0.061* (40) *99.632*

Hydrogen. The difference in the isotopic composition between atmospheric H_2 and water vapour in air has not been determined, but if the two are in equilibrium (which is not necessarily the case), one would expect a considerably reduced deuterium-hydrogen ratio in the gaseous hydrogen, as $\frac{(D/H)_{\text{water vapour}}}{(D/H)_{\text{hydrogen}}} = 3.6$ (Suess [21]).

Helium.[5] Much greater differences are observed for the ^3He content of helium found in air, in rocks, and in oil wells (Aldrich and Nier [2], and Coon [8]), the ratios $^3He/^4He$ being 1.2×10^{-6}, 1.5×10^{-7}, and 3×10^{-8}, respectively. This clearly points to a different mode of origin of the two *He* species in the three cases. The ^3He in the atmosphere is suspected to arise from the reaction of nitrogen with neutrons derived from cosmic radiation. The ^3He in the lithosphere is presumably due to the action of neutrons on *Li* where the neutrons arise from known reactions of small atomic nuclei with the α particles of the natural radio-ele-

4. The questions of the origin and development of the atmosphere, though of interest to meteorologists, cannot be adequately dealt with in this paper. Attention is drawn to detailed articles by Chamberlin, Brown, and Kuiper [15, Chaps. 8, 9, 12], and by Wildt [24] where further references may be found.

5. (*Added in press*) See also the recent note by P. Harteck and V. Faultings on "The ^3He-Problem of the Atmosphere." *Nature*, 166:1109 (1950).

ments. The connection of ^3He in minerals with Li is borne out by the abnormally high $^3He/^4He$ ratio in spodumene, a lithium aluminium silicate.

Carbon. Another isotope which owes its existence to the neutrons from cosmic radiation is the radioactive ^{14}C, which is produced by $^{14}N + n \rightarrow {}^1H + {}^{14}C$.

As cosmic ray neutrons are produced mainly in the upper atmosphere, the ^{14}C starts its career as CO_2 and enters into all organic matter by the process of plant assimilation. After the death of the plant the radioactive carbon decays with a half-life of 5720 years, so that old coal and oil deposits are no longer radioactive. The proportion of radiocarbon was found to be 0.95×10^{-12} g of ^{14}C per g ^{12}C in living matter [3].

FUTURE DEVELOPMENTS

Oxygen and Carbon Dioxide. Our best values of the O_2 content or, for that matter, of the O_2 and CO_2 content of air date from 1912. Since then a number of improvements in the control of thermostats and in all manner of measuring devices have been made, which should render possible an increased accuracy. A redetermination is particularly desirable in order to get an idea of any long-term variations of the O_2 content of air.

There are, moreover, the unexplained O_2 values of Lockhart and Court in the Antarctic which should be either confirmed or refuted. This applies also to the high CO_2 values observed by Krogh in certain arctic regions, where the observed variations were very large, and further investigations are likely to bring to light some interesting phenomena responsible for such changes. Callendar's suggestion of a CO_2 increase during this century due to industrial CO_2 production requires that the CO_2 content of the Northern Hemisphere should be slightly larger than that of the Southern Hemisphere, a feature which should be subject to experimental verification by modern techniques.

The results of Buch and of Wattenberg make it fairly certain that there is a CO_2 circulation in the oceans involving uptake of CO_2 at high latitudes and its release at low latitudes. Some light on the time scale of this cycle might be thrown by a ^{14}C analysis of the CO_2 released from the sea at low latitudes, which, if the cycle exceeds 10^3 years, would result in a noticeable decrease of ^{14}C activity.

Methane. The CH_4 in the atmosphere, discovered only recently, still offers a few problems. Its percentage in air requires more accurate determination, and the question of its production requires further study.

Hydrogen. Our knowledge of the H_2 content of the air is so far inadequate.

Ozone. Large gaps still exist in our knowledge of the atmospheric O_3 near the ground. Its geographical distribution at ground level is quite unexplored; its dependence on weather conditions has only been touched on, and requires much more systematic investigation, particularly with reference to its vertical distribution. The increased occurrence of O_3 near cloud levels and its connection with thunder clouds also require more detailed investigation, including a study of its possible mode of generation under such conditions.

Upper Atmosphere. It now seems certain that the upper atmosphere has essentially the same composition as that found at the ground, at least up to heights of 70 km, though further confirmation of the rocket data is desirable and will no doubt be obtained in the near future. This uniformity means that turbulent mixing in the stratosphere is considerably greater than was formerly expected. An explanation for this turbulence has been suggested by Brewer [4] who assumes an air circulation involving a movement of air into the stratosphere at the equator followed by slow poleward movement in the stratosphere, accompanied by a slow sinking movement in the temperate and polar regions. As this hypothesis requires the abandonment of the idea of a stratosphere which is in radiative equilibrium, many new and interesting problems arise which require experimental confirmation.

REFERENCES

1. Adel, A., "Concerning the Abundance of Atmospheric Carbon Monoxide." *Phys. Rev.*, 75:1766–1767 (1949).
2. Aldrich, L. T., and Nier, A. C., "The Occurrence of ^{3}He in Natural Sources of Helium." *Phys. Rev.*, 74:1590–1594 (1948).
3. Anderson, E. C., and others, "Natural Radiocarbon from Cosmic Radiation." *Phys. Rev.*, 72:931–936 (1947).
4. Brewer, A. W., "Evidence for a World Circulation Provided by the Measurements of Helium and Water Vapour Distribution in the Stratosphere." *Quart. J. R. meteor. Soc.*, 75:351–363 (1949).
5. Callendar, G. S., "Variations of the Amount of Carbon Dioxide in Different Air Currents." *Quart. J. R. meteor. Soc.*, 66:395–400 (1940).
6. Chackett, K. F., Paneth, F. A., and Wilson, E. J., "Chemical Analysis of Stratosphere Samples from 50 to 70 Km. Height." *J. atmos. terr. Phys.*, 1:49–55 (1950).
7. Chalonge, D., Götz, F. W. P., et Vassy, E., "Mesures simultanées de la teneur en ozone des basses couches de l'atmosphère à Jungfraujoch et à Lauterbrunnen." *C. R. Acad. Sci., Paris*, 198:1442 (1934).
8. Coon, J. H., "Isotopic Abundance of ^{3}He." *Phys. Rev.*, 75:1355–1357 (1949).
9. Dole, M., and Slobod, R. L., "Isotopic Composition of Oxygen in Carbonate Rocks and Iron Oxide Ores." *J. Amer. chem. Soc.*, 62:471–479 (1940).
10. Glueckauf, E., "The Ozone Content of Surface Air and Its Relation to Some Meteorological Conditions." *Quart. J. R. meteor. Soc.*, 70:13–19 (1944).
11. —— "A Micro-analysis of the Helium and Neon Contents of Air." *Proc. roy. Soc.*, (A) 185:98–119 (1946).
12. —— and Paneth, F. A., "The Helium Content of Atmospheric Air." *Proc. roy. Soc.*, (A) 185:89–98 (1946).
13. Götz, F. W. P., Schein, M., und Stoll, B., "Messungen des bodennahen Ozons in Zürich." *Beitr. Geophys.*, 45:237–242 (1935).
14. Krogh, A., "Abnormal Carbon Dioxide Percentage in the Air in Greenland. . . ." *Medd. Grønland*, 26:407–435 (1904).
15. Kuiper, G. P., ed., *Atmospheres of the Earth and Planets*. Chicago, University of Chicago Press, 1949. (See Chaps. 8, 9, 10, and 12.)

16. Paneth, F. A., "The Chemical Composition of the Atmosphere." *Quart. J. R. meteor. Soc.*, 63:433–438 (1937).
17. Regener, E., "Ozonschicht und atmosphärische Turbulenz." *Meteor. Z.*, 60:235–269 (1943).
18. Regnault, M. V., "Recherches sur la composition de l'air atmosphérique." *Ann. Chim.* (*Phys.*), 3e sér., 36:385–405 (1852).
19. Riesenfeld, E. H., und Chang, T. L., "Die Verteilung der schweren Wasser Isotope auf der Erde." *Naturwissenschaften*, 24:616–618 (1936).
20. Slobod, R. J., and Krogh, M. E., "Nitrous Oxide as a Constituent of the Atmosphere." *J. Amer. chem. Soc.*, 72: 1175–1177 (1950).
21. Suess, H., "Isotopen Austauch Gleichgewichte" in *FIAT Rev. of German Sci., 1939–1946*, Vol. 30, *Physical Chemistry*, pp. 19–24. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948.
22. Usher, F. L., and Rao, B. S., "The Determination of Ozone and Oxides of Nitrogen in the Atmosphere." *J. chem. Soc.*, 111:799–809 (1917).
23. Vinogradow, A. P., and Teis, R. V., "Isotopic Composition of Oxygen of Different Origins." *C. R.* (*Doklady*) *Acad. Sci. URSS*, 33:490–493 (1941).
24. Wildt, R., "The Geochemistry of the Atmosphere and the Constitution of the Terrestrial Planets." *Rev. mod. Phys.*, 14:151–159 (1942).

RADIATION

Solar Radiant Energy and Its Modification by the Earth and Its Atmosphere *by Sigmund Fritz* 13

Long-Wave Radiation *by Fritz Möller* 34

Actinometric Measurements *by Anders Ångström* 50

SOLAR RADIANT ENERGY AND ITS MODIFICATION BY THE EARTH AND ITS ATMOSPHERE

By SIGMUND FRITZ

U. S. Weather Bureau, Washington, D. C.

The sun is the principal source of the energy which, by devious means, becomes the internal, potential, and kinetic energy of the atmosphere. The solar irradiation of a unit horizontal surface at the outer limits of the earth's atmosphere can be evaluated, at least in relative units, from astronomical and trigonometrical considerations [61]; thus on a relative scale, the diurnal and seasonal variations of this solar irradiation above the atmosphere are known. On the average, variations similar to these occur also at the earth's surface, and the associated diurnal and seasonal changes in atmospheric temperature are commonplace knowledge [52].

There are, however, additional changes in effective solar irradiation of the planet Earth which are superposed on the trigonometrical variations. These are of two kinds. The first is due to the change in the quality and quantity of energy which leaves the sun. The second is caused by changes in the reflectivity of the atmosphere (including clouds) and of the earth's surface; the solar energy which is immediately reflected to space cannot be meteorologically effective.

In contrast to the regular, astronomically induced changes in meteorological parameters (notably diurnal and seasonal atmospheric temperature changes), the large-scale meteorological consequences of these irregular changes in solar irradiation are far from obvious, if, indeed, any such meteorological effects can be shown to be induced at all by them. For example, changes of the first kind (*i.e.*, in solar output) have been invoked as possible causes of abnormal heating in the ozone layer with subsequent pressure changes at the earth's surface [44]; changes of the second kind (*i.e.*, in reflection by the earth or clouds) are important, for instance, in local turbulence of the air near the ground, but are rarely used to explain widespread meteorological phenomena. These as well as other solar-induced meteorological phenomena are discussed elsewhere in this Compendium. In this article we shall, for the most part, examine the solar energy itself and shall mention its meteorological effects only incidentally.

SOLAR RADIATION OUTSIDE THE EARTH'S ATMOSPHERE

The Sun. The sun, located about 93,000,000 miles from the earth, is a large, hot, gaseous mass. When viewed through a smoked glass it appears as a smooth circular disk which is called the *photosphere*, but when examined with the aid of more refined techniques, the photospheric surface appears highly granulated and is surrounded by a gaseous envelope which is commonly divided into three layers for descriptive purposes. Of these the one just outside the photosphere is the relatively thin *reversing layer*, so called because the spectral lines ordinarily seen as dark absorption lines in the photospheric spectrum appear as bright emission lines when examined in the reversing layer; still farther from the photosphere is the *chromosphere*; and beyond that is the *corona* [4]. The sun's "surface" and its surrounding atmosphere are by no means static. The presence of short-lived photospheric grains, dark sunspots, bright areas (faculae and flocculi), and erupting prominences indicate that the entire observable sun is in a state of considerable turmoil. This turmoil is associated with variations in the quantity and spectral intensity distribution of the solar energy which subsequently irradiates the outer limits of the earth's atmosphere.

Average Spectral Distribution of Sunlight. To calculate the solar energy available for meteorological processes, it is desirable to measure the amount of solar energy I_0 (the so-called *solar constant*) which reaches the outer atmosphere of the earth. To determine the interaction between our atmosphere and the sun's energy, the distribution of spectral intensity $I_{0\lambda}$ of the extraterrestrial solar energy is also required. For both of these quantities we are largely indebted to the Astrophysical Observatory of the Smithsonian Institution, whose determinations of I_0 and $I_{0\lambda}$ are monuments to the work of the Observatory.

Until recently, the sun's energy had not been directly observed below wave lengths of about 2900 A because of the absorption of energy of shorter wave lengths by the envelope of ozone which surrounds the earth up to about 50 km. For wave lengths greater than 2900 A, selective scattering and absorption, mainly by air, dust, and water vapor, modify the solar spectrum; these troublesome modifications must be eliminated from measurements made at the ground under cloudless skies to obtain the extraterrestrial spectrum. In the region from about 0.29 μ to 2.5 μ numerous measurements and extrapolations have been made. The Smithsonian Institution [3] is the main source of spectral information for the region above 3400 A, but its latest summary of data was compiled in 1923 [2]. Recently, V-2 rocket observations have extended the measured spectrum down to 2200 A [47].

Visible and Near Infrared Radiation. Both the total amount and the spectral distribution of the radiant energy emitted by a black body are determined by its temperature. It is therefore convenient to describe the radiant energy from a source in terms of the black-body temperature which would most nearly produce the observed energy. The sun *does not* radiate as a black body. Despite this fact, the average observed energy curve from about 0.45 μ to about 2.0 μ closely approxi-

mates the spectral-energy distribution of a black body (Fig. 1). It is therefore not unusual to speak of the black-body temperature of the sun as 6000K, although this numerical value may vary by several hundred degrees depending upon the method used to calculate it [4]. For example, the total energy emitted by a black body per unit area per unit time is given by σT^4, where T is its temperature in degrees absolute and σ is the Stefan-Boltzmann constant. The solar constant is commonly taken as 1.94 ly min^{-1} (where ly = langley = g cal cm^{-2}), and considering the distance from the earth to the sun, we can calculate the total energy emitted by the sun. Then, by introducing the sun's

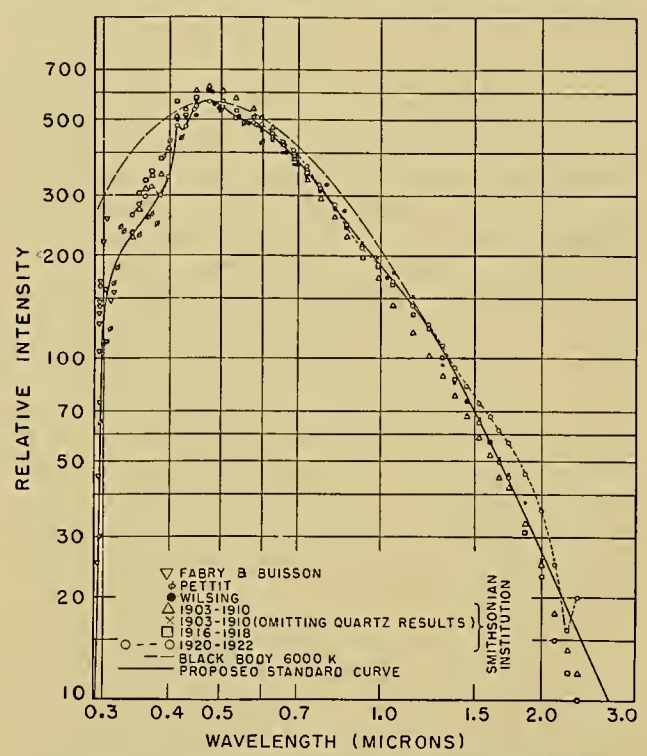


FIG. 1.—Spectral intensity distribution of solar radiation outside the earth's atmosphere. (*After Moon* [63].)

area, we calculate the energy radiated per square centimeter of solar surface. Placing that energy equal to σT^4, we get T = 5770K.

Another temperature estimate results from Wien's displacement law for black bodies: $\lambda_{max}\ T$ = const, where λ_{max} is the wave length of maximum intensity. Abetti [4] takes λ_{max} = 4740 A and finds T = 6080K.

It should be emphasized that at wave lengths outside the region 0.45 μ to 2.0 μ solar energy departs markedly from that of a black body at 6000K. Moon [63] has combined the results of several authors and obtained Fig. 1, which shows the relative intensity $I_{0\lambda}$ of the energy as a function of wave length.

Ultraviolet Radiation. The energy in the ultraviolet spectrum is well below that of a black body at 6000K; this has been established by several investigators [35, 37, 72]. Hulburt's curve [47] in Fig. 2 contains the Naval Research Laboratory rocket measurement at 55 km in which no ozone could be detected and was one of a series of spectra observed at levels up to 88 km. Other measurements from rockets up to 155 km [20] show results similar to those at 55 km and indicate that the energy is less than that of a black body at 6000K.

A recent curve by Götz and Schönmann [35], based on five observations in the region from 3300 A to 5000 A, lies far below (by a factor of 2 or more at 3400 A) Moon's curve of Fig. 1; this is also true of Hess's data [55, p. 324]. Perhaps these lower values are due to the fact that the Smithsonian observations were troubled by scattered light at short wave lengths. If the rocket data (Fig. 2) had been joined to the data of Götz and Schönmann, the resultant spectral curve would ob-

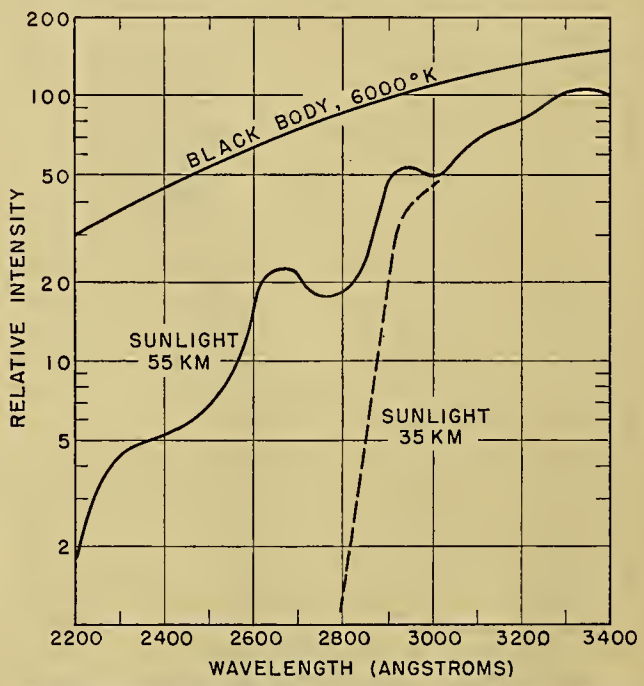


FIG. 2.—Spectral intensity distribution of ultraviolet solar radiation measured from a rocket. (*After Hulburt* [47].)

viously have been much farther below the 6000K curve. Future measurements will decide the true state of affairs.

Infrared Radiation. The near infrared has been observed many times by the Smithsonian Institution. Abbot and others [2] considered their 1920–22 curve as the best estimate. Their data exceed the black-body curve (Fig. 1) in the region near 2 μ, but Moon, also accepting the earlier data, adopted the 6000K blackbody curve as the true one beyond 1.25 μ. Beyond 2.5 μ, Adel [5] has observed the spectrum up to about 24 μ. He finds that the solar temperature is about 7000K in the far infrared.

Short-Wave Radio Radiation. Measurements up to 24 μ have been made by optical means. Recently the observations have been extended by radio detection instruments into wave lengths of the order of a few centimeters to a few meters [39, 62]. These observations indicate black-body temperatures of the order of 10^6K

or more. It should be pointed out, however, that the black-body energy is so small at those wave lengths—even during disturbed sun conditions—that radiation from a black body at 10^6K can contribute only a very small amount of thermal energy by comparison with the principal solar energy [39].

The Unmeasured Radiation. We turn now to an important region of unmeasured radiation, namely, the ultraviolet energy below 2200 A.

From a consideration of the ionized states of certain elements observed in the sun and from the consideration of the state of ionization of the earth's ionosphere, several investigators have concluded that in the region below 1000 A the sun radiates energy corresponding to temperatures above 6000K. Greenstein [37] mentions the uncertainty regarding the need for assuming such excess ultraviolet radiation and concludes that at least for $\lambda > 1215$ A the temperature corresponding to solar energy is probably less than 6000K and is near 5000K. For $\lambda \leqq 900$ A he discusses a suggestion by Kiepenheuer and Waldmeir that corona temperatures of 10^6K enhance the photospheric radiation by a factor of 2 at 900 A and by 2×10^6 at 600 A. For $\lambda > 900$ A no appreciable radiation is contributed by the corona. Superposed on the *average* radiations, measured and unmeasured, are the radiations from the "cool," dark sunspots (temperature about 4800K [4]), and the increased radiation from the bright areas, that is, from faculae and flocculi.

In addition to the electromagnetic radiation described above, the sun also emits particles which are responsible for such effects as the aurorae and some types of magnetic and ionospheric storms.

Summary. It appears then that in the optically measured spectrum, from about 0.45 μ to 24 μ the sun radiates as a black body whose temperature is close to 6000K, perhaps increasing to 7000K towards the higher wave lengths. For shorter wave lengths, at least down to 0.22 μ, the measured radiation corresponds to a considerably lower temperature of about 4000K to 5000K. For 1000 A $< \lambda <$ 2200 A this lower temperature may also apply [37]. But for $\lambda <$ 1000 A the hot corona may dominate, resulting in much higher effective black-body radiation. In the comparatively long wave lengths from 1 cm to 30 m, high temperatures (10^6K) are again indicated. In the region from 24 μ to 1 cm, no measurements seem to have been made.

Fluctuations of Emitted Radiation. The bright and dark markings and other manifestations of changes on the sun can be expected to produce spectral emissions which differ from the average solar emission. These spectral variations are indeed observed directly or indirectly in nearly all parts of the solar spectrum.

Far Ultraviolet. That there are marked changes in the far ultraviolet emission from the sun is evident from measurements of the reflection of radio waves by the ionospheric layers in the earth's upper atmosphere. These layers are regions where solar ultraviolet energy ($\lambda <$ 1000 A) has ionized the atmospheric gases, with the result that they have the property of reflecting radio waves of certain frequencies. The approximate heights [38, 62] and pressures of those layers are given in Table I.

Table I. Approximate Heights and Pressures of Ionized Layers

Layer	Height (km)	Approximate pressure (mb)
F_2	300	5×10^{-7}
F_1	200	1×10^{-5}
E	100	5×10^{-3}
D	80 to 50	5×10^{-2} to 1

The highest frequency radio signal which can be reflected is called the *critical frequency* f_o for the reflecting layer, and it can be shown that under certain assumptions [64, p. 59]

$$f_o^2 \propto N_i, \tag{1}$$

and that

$$N_i^2 \propto I_{0\lambda}, \tag{2}$$

where N_i is the maximum ion density of an ionized layer. Hence on combining equations (1) and (2),

$$I_{0\lambda} \propto f_o^4. \tag{3}$$

The critical frequency f_{oF_2} of the F_2-layer fluctuates considerably from day to day. When the measured daily value of f_{oF_2} at noon is plotted against the daily sunspot number, no significant correlation can be found [69]; on a monthly basis a small correlation may be present [13]. However, if a twelve-month running average of noon f_{oF_2} is plotted against the twelve-month running averages of sunspot numbers, a linear relation results with very little scatter of the points [64], and f_{oF_2} increases by a factor of about 2 from sunspot minimum to sunspot maximum. Similar results appear for the F_1- and E-layers; but for these latter layers, f_o increases by a factor of about 1.2 from sunspot minimum to sunspot maximum [9]. The D-region variation is apparently similar to the E and F_1 variation [64].

If equation (3) is applied to the smoothed data over the sunspot cycle, $I_{0\lambda}$ (which may be in a different wavelength region for each layer) increases by a factor of approximately 2 for the E- and F_1-regions [9, 64]. But since f_{oF_2} increases by a factor of 2, $I_{0\lambda}$ would increase sixteenfold in the F_2-region [64]. However, other relations have been mentioned for the F_2-layer. Mitra [62] indicates that instead of equation (2), $N_i \propto I_{0\lambda}$ applies here; this leads to $I_{0\lambda} \propto f_o^2$. Allen [9] prefers $I_{0\lambda} \propto f_o$. Mitra's relation leads to a fourfold increase in $I_{0\lambda}$, while Allen's relation leads to a twofold increase. We may conclude therefore that $I_{0\lambda}$, which produces the F_2-layer, surely increases from sunspot minimum to sunspot maximum and that the amount of the increase is at least twofold.

What is the significance of ionospheric heating for tropospheric meteorology? Since very little direct relation exists between the critical frequencies and the unsmoothed sunspot data, some effect (possibly solar)

other than sunspots must produce these ionospheric variations; only on the *average* is the ultraviolet intensity correlated with sunspots. We should expect therefore that the meteorology of the troposphere would not be correlated with sunspots on a daily basis if ionospheric variations were used as a criterion. What about longer periods? The wave lengths involved are of the order of 1000 A or less, so that the amount of energy is small and the pressure (about 10^{-3} mb or less) and density of the absorbing layers are so small that it seems unlikely that heating in the ionosphere by the increased ultraviolet energy can directly affect the meteorology of the lower atmosphere; at least that is the view of one school of thought [71, p. 504]. Therefore for longer periods also, if *only* the ionosphere were involved in variable ultraviolet absorption, no tropospheric effect would be noticed.

At present, one cannot specify the height at which anomalous heating in the upper atmosphere can affect the troposphere through dynamic processes. However, if in addition to an increase in the radiation which heats the ionosphere, increases also occur in the ultraviolet energy which can penetrate to lower layers, for example, to the top of the ozone region (40–50 km) where the pressure is of the order of 1 to 3 mb, then dynamic pressure changes caused by additional heating are more likely [44; 71, p. 504]. And indeed such ultraviolet energy increases *may* occur, but it is not certain at present that significant increases occur at wave lengths which can heat the ozonosphere.

There are several solar-induced effects in the ionosphere, as revealed by magnetic and radio data, and from the ozone-heating viewpoint at least one of these ionospheric phenomena deserves further mention. That phenomenon is the radio fade-out or sudden ionospheric disturbance (SID). SID's are caused when radio waves transmitted upward from the ground are strongly absorbed in the D-layer, so that they cannot be reflected back to the ground by the higher ionospheric layers. The increased absorption of the radio waves is caused by a sudden increase in the ionization of the D-layer; the increased D-layer ionization is caused by a sudden large increase in the solar ultraviolet energy which reaches and is absorbed by the layer. In short, SID's are caused by sudden increases in ultraviolet solar radiation and occur simultaneously with the appearance of visible bright solar flares on the sun (chromospheric eruptions).

Moreover, it is found that during SID's the F-layers are practically unaffected and the E-layer is only slightly affected. Hence the upper ionospheric regions are apparently transparent to the ionizing radiations in this case, while the lower D-region absorbs them strongly. The duration of SID's is of the order of a few minutes to a few hours, and their intensity is of course variable.

Here then is a phenomenon which produces short-lived intense ionization, and hence heating in the vicinity of 60 km. Since the upper layers are unaffected, the radiation must be of wave lengths such that the air above 80 km is transparent. Wulf and Deming [79] offer an interesting explanation of this ionization. Ozone absorbs very strongly in the region 2300 A to 2800 A, but this spectral region is transmitted readily by the atmosphere above 60 km. They suggest that partial absorption at the top of the ozone layer ionizes the ozone there. Increased solar emission at those wave lengths may therefore be responsible for the increased D-region ionization and hence for SID's. It should be pointed out however that the recent V-2 rocket measurements [47] could not detect any ozone above 55 km. Probably an amount of ozone smaller than could be detected by the rocket exists above 55 km and this is sufficient to produce the observed ionization.[1]

According to some writers, the increased emission during solar flares is contributed largely by specific elements, hydrogen and calcium, for example [25], although emissions from other elements have been measured. Hydrogen does not radiate in the wave lengths 2300–2800 A. Calcium and some of the other elements, on the other hand, do emit monochromatic radiation there, so that some increase in $I_{0\lambda}$ occurs during solar flares in this region—how much of an increase is still unknown.

Mitra [62] discusses some other SID explanations and prefers 973 A as the wave length of the ionizing radiation. Heating by such radiation in a narrow spectral band may be small and should not extend very far into the ozonosphere. But if, as Wulf and Deming suggest, increased radiant energy in the broad band 2300–2800 A is emitted, then the resulting increased heating, not only at and above 60 km but also lower in the ozonosphere, may have important implications for tropospheric meteorology [44]. It would therefore be highly desirable to measure the distribution of spectral intensity in this region at frequent intervals. Hulburt [47] reports that the solar spectrum near 2200 A was detected at 34 km from the V-2 rocket, and Brasefield [16] has described some temperature measurements from an unmanned balloon up to 140,000 ft (42 km). If such balloons could be equipped for constant-level flights at 40 km or higher and designed to carry spectral measuring equipment, it would be possible to determine the solar spectrum near 2200 A and for $\lambda > 2700$ A at frequent intervals during days of both disturbed and undisturbed solar conditions. Such measurements may determine whether current ideas regarding solar control of weather through heating of ozone have any basis.

Near Ultraviolet. In a series of optical measurements during the years 1924–32, Pettit [68] determined the intensity at 3200 A relative to the intensity at 5000 A and extrapolated in the usual manner to the "top" of the atmosphere. Presumably the intensity at 5000 A changed rather little, so that the variation in the ratio reflects the time variation in the intensity at 3200 A. His averaged data show a rather good agreement with the smoothed sunspot numbers, low sunspot numbers

1. At the January 1950 meeting of the American Meteorological Society in St. Louis, Missouri, R. Tousey of the Naval Research Laboratory reported a rocket measurement of small amounts of ozone up to 70 km.

corresponding to low ultraviolet radiation, as in the ionosphere, but at times the correlation was negative (June 1928 to June 1929). Since the intensity at sunspot maximum was about 1.5 times that at sunspot minimum, he felt that in general the range of the observed variations at 3200 A was too great and that an atmospheric effect might have been partly responsible for the range.

In support of Pettit's doubts, Bernheimer [15] found an annual trend in Pettit's data, and indicated that the observed variations may have been due to changes in atmospheric transmission, rather than to increased solar emission. As we shall see later, this difficulty of extrapolating through the earth's atmosphere is also a continual worry in determinations of variations in the solar constant.

Visible and Infrared Radiation. The fluctuations in ultraviolet radiation seem to be related to hot, bright areas, for example, faculae and flocculi [9]. These areas, which can be seen visually, cover rather small portions of the sun. As the temperature of a body increases, the intensity of the emitted radiation increases at all wave lengths, but the wave length of maximum intensity shifts towards shorter wave lengths. Hence, since the maximum intensity occurs at a wave length of about 4700 A in the average solar spectrum (Fig. 1), it might be expected that, as the sun's temperature increases, the relative increase of intensity at $\lambda < 4700$ A will be greater than in the longer wave length visible or infrared regions. Such is indeed the case; the intensity of the radiation in the visible and infrared changes by small amounts in response to the dark and bright markings on the sun.

Abbot [3, Vol. 6, p. 165] found that, on days with high solar constant, $I_{0\lambda}$ at 0.35 μ increased by about 5 per cent over its value on days with low solar constant. At 1.7 μ, $I_{0\lambda}$ *decreased* by about 1 per cent under the same circumstances. It should be noted, however, that solar-constant variations are not well correlated with sunspots [8].

Radio Waves. As stated earlier, radio-wave emission from the sun (from a few centimeters to a few meters in wave length) indicates temperatures of 10^6K; under disturbed conditions, radiation corresponding to 10^9K can be observed [60, p. 328]. However, the amounts of these energies are very small.

Summary. There is good evidence from radio measurements that large fluctuations in solar radiation occur at $\lambda < 1000$ A and at $\lambda = 10$ to 10^3 cm. The variations seem to appear in daily measurements and also appear systematically in the sunspot cycle. In the meteorologically important spectral region of 2000–2800 A there is a suggestion that at least short-lived large fluctuations may occur (SID) [79]; direct measurements of these fluctuations to determine their magnitude would be very desirable. At somewhat longer wave lengths, for example 3200 A [68], measurements indicate solar-controlled fluctuations, but doubts have been raised about the reality of their magnitude. That variations in the visible spectrum from parts of the sun occur can be seen from the bright and dark areas on the sun. However, these areas are small and the variations in the visible and near-infrared radiation represent only a small percentage of the average radiation from the entire sun.

The Solar Constant. So far we have discussed the *relative* spectral distribution of intensity in solar radiation. To describe the radiation it is also necessary to specify the *amount* of radiation on an absolute basis.

The "solar constant" is a measure of the total amount of heat which reaches the outer atmosphere of the earth. Specifically, it is often expressed as the amount of energy which, in one minute, reaches a square centimeter of plane surface placed perpendicular to the sun's rays outside our atmosphere when the earth is at its mean distance from the sun.

Methods of Measurement. To clarify the following discussion let us review the basis for the fundamental (or "long") method of the Smithsonian Institution for measuring the solar constant [3, Vol. 6, p. 30]. The intensity I_λ of parallel monochromatic energy transmitted through the earth's cloudless atmosphere is given by

$$I_\lambda = I_{0\lambda} e^{-k_\lambda m} = I_{0\lambda} a_\lambda{}^m, \tag{4}$$

or

$$\ln I_\lambda = \ln I_{0\lambda} + m \ln a_\lambda,$$

where k_λ is the extinction coefficient, a_λ is the atmospheric transmission with the sun in the zenith, and m is the optical air mass or the path length of the parallel light through the atmosphere measured in terms of the zenith path as unity.[2]

Except for large zenith angles Z, the value of m is given by sec Z. Hence m can be readily determined; I_λ is measured in relative units. If a_λ remains constant, then by equation (4) a graphical plot of ln I_λ against m will yield a straight line whose slope is ln a_λ and whose intercept for $m = 0$ (outside the atmosphere) is ln $I_{0\lambda}$; I_λ is measured nearly simultaneously for the spectral region $0.34\,\mu < \lambda < 2.5\,\mu$. This is repeated for several values of m, so that by the graphical method just mentioned $I_{0\lambda}$ can be evaluated at several wave lengths in the region. From the measurement and graphical extrapolation, a plot of I_λ vs. λ, and another plot of $I_{0\lambda}$ vs. λ can be made in *relative* units for the region 0.34 μ to 2.5 μ. We desire to find the areas $\Sigma I_\lambda \Delta\lambda$ and $\Sigma I_{0\lambda}\,\Delta\lambda$, respectively, under these graphs in *absolute* units. Therefore, by means of a pyrheliometer, a nonspectral measurement in absolute energy units is made of the total radiation I; I includes not only the energy which reaches the observer for λ between 0.34 μ

2. The value of m is ordinarily specified as unity for zenith path length at sea level. Values of m at sea level are given by sec Z for values of Z (sun's zenith distance) up to 70°; for larger Z, Bemporad's formula [56] is commonly used. To compute the air mass m_p for elevated stations where the pressure is p, the sea-level value of m is multiplied by p/p_0 where p_0 is the sea-level pressure. When I_λ is measured at one station, however, it is not necessary to correct m for pressure to find $I_{0\lambda}$ from equation (4).

and 2.5 μ, but also the energy for $0.29\ \mu < \lambda < 0.34\ \mu$ and $\lambda > 2.5\ \mu$ (since energy below 0.29 μ does not reach the ground). Hence, if to the area $\Sigma I_\lambda \Delta\lambda$ under the spectral curve is added the energy ϵ in the wave lengths which reach the ground but are not measured spectrally (namely $0.29\ \mu < \lambda < 0.34\ \mu$ and $\lambda > 2.5\ \mu$) then the area under the curve, $\epsilon + \Sigma I_\lambda \Delta\lambda$, will be proportional to the corresponding pyrheliometric measurement I.

The area I_{0a} under the curve of $I_{0\lambda}$ versus λ (for $0.34\ \mu < \lambda < 2.5\ \mu$) can now be converted to energy units, since

$$\frac{I}{I_{0a}} = \frac{\Sigma I_\lambda \Delta\lambda + \epsilon}{\Sigma I_{0\lambda} \Delta\lambda}. \tag{5}$$

It remains to add to I_{0a} the energy ϵ_0 for $\lambda < 0.34\ \mu$ and $\lambda > 2.5\ \mu$ in order to obtain I_0, the solar constant.

This "long" method implies that a_λ is constant during the period of measurements (2–3 hr), and also that I, the pyrheliometrically measured radiation, is known in absolute units (langleys per minute). In practice the corrections for all unmeasured spectral regions are applied together.

Equation (4) requires that the atmospheric transmission a_λ be constant during a series of spectral measurements which are performed during a period of a few hours (from air mass 5.0 to 1.5). To the extent that a_λ is not constant, errors will be introduced in $I_{0\lambda}$ and hence in the solar constant. Partly for that reason, the Smithsonian Institution devised a "short" method for measuring the solar constant. In this method, a_λ is determined by a measurement of the brightness of the sky in the vicinity of the sun and from an empirical relation between the brightness and the transmission coefficient a_λ.

The Numerical Value of the Solar Constant. Let us consider I. The standard of pyrheliometry adopted by the Smithsonian Institution in 1913, when applied to solar-constant measurements, leads to 1.94 ly min^{-1} as the average value for I_0. This is the value most often used at present. The Smithsonian standard was known to be about 3.5 per cent higher than the Ångström standard, but in 1932 and on subsequent occasions Abbot and Aldrich redetermined their standard and found that their 1913 standard was too high by 2.3 per cent [12], thus decreasing the disagreement with the Ångström scale. The new Smithsonian standard scale reduces the value of the average solar constant from 1.94 to 1.90 ly min^{-1}.

Before accepting this new value we should examine the corrections applied to the solar constant for the unmeasured spectral radiation [3, Vol. 5, p. 103]. In practice a correction is made for the unmeasured radiation in the interval $0.34\ \mu > \lambda > 0.27\ \mu$, and for this spectral region about 3.4 per cent of the total measured radiation is added. Apparently no energy is added for radiation below 0.27 μ. A composite curve of the 55-km rocket measurement and Smithsonian surface measurements indicates that the region from 0.34 μ to 0.27 μ comprises 2.9 per cent of the area between 0.34 μ and 2.5 μ and that about 0.5 per cent of the total radiation is included between 0.22 μ and 0.27 μ [56]; hence the energy ordinarily added for $\lambda < 0.34\ \mu$ is about right. We have assumed here that the 55-km rocket observation represents $I_{0\lambda}$ for $\lambda < 0.34\ \mu$.

In the infrared, 2 per cent is added for radiation beyond 2.5 μ during I_0 determinations. For a black body at 6000K this should amount to about 3.1 per cent. Hence if the energy beyond 2.5 μ is that of a black body at 6000K or more [5, 63], it would appear that a somewhat greater amount than 3.1 per cent is the necessary correction.

The area $\Sigma I_\lambda \Delta\lambda$ is adjusted to agree with I by estimating the unmeasured spectral energy. In this process of adjustment, errors in the estimated correction are offset somewhat by the adjustment mechanism itself. This adjustment process, however, does not account for errors in ϵ_0 in the spectral regions which cannot be observed at the ground. It is difficult therefore to say exactly what effect the substitution of new corrections would have on I_0. As a first approximation we might assume that the ultraviolet correction is about right and that the infrared correction is too low by 1 or 2 per cent, so that the computed solar constant can tentatively be given by $1.90 < I_0 < 1.94$, the range making some allowance for the uncertainty of the corrections. On the basis of the Smithsonian measurements there does not seem to be much justification for a solar constant of more than 2.0 ly min^{-1}.[3]

Variation of the Solar Constant. We turn now to a consideration of the following questions: (1) Does the "solar constant" vary with time? and (2) Do the measurements of I_0 indicate the actual variation? The controversy regarding these questions has been raging for a long time, and a definite answer to the second question cannot yet be given. The state of the controversy is shown in a criticism by Paranjpe [67] and a reply by Abbot [1]; Waldmeir [77] has summarized some of the earlier arguments. Two main points have been at issue: (1) Do the measured values at two or more widely separated stations vary in the same way with time? and (2) Do the I_0 measurements or the transmission coefficients a_λ vary seasonally? If so, one would expect that the earth's atmosphere is introducing the variations and that they are not true solar changes.

Abbot has pointed out on several occasions that the data from the various stations do agree for monthly means, but that daily values show appreciable departures. Paranjpe [67] states, however, that the data from the stations undergo statistical adjustment in such a way as to *make* the data between stations comparable. From this view, of course, interstation correspondence would have no significance. Abbot [1] states definitely

3. Karandikar [50] assumed a value of more than 2.0 ly min^{-1}. At the Ionospheric Physics Conference held at State College, Pennsylvania, in July, 1950, Dr. M. O'Day announced that a measurement from a rocket indicated a value of more than 2.0 ly min^{-1}. This measurement has apparently not yet been completely checked. See *Discussion* of the paper "Physical Characteristics of the Upper Atmosphere" by T. R. Burnight in "Proceedings of the Conference on Ionospheric Physics (July 1950)." *Geophys. Res. Pap.*, Air Force Cambridge Research Laboratories, Cambridge, Mass. (1951) (in press).

that this is not the case, but that the data are independent. It should be pointed out, however, that questions regarding the accuracy of data from all solar-constant stations except Montezuma have been raised from time to time by Abbot and his colleagues, especially regarding the data prior to 1920 [8].

With regard to the seasonal variation of solar constants, Abbot has made comparisons of differences between solar-constant measurements at Southern Hemisphere and Northern Hemisphere stations on a monthly basis and finds no seasonal variation in the differences. The season being opposite in the two hemispheres, this would indicate lack of seasonal variation in I_0. Although he found no twelve-month period in I_0 variations, Abbot has found fourteen other different periods in the solar-constant variations. Paranjpe questioned the existence of these periods, but Sterne [73], while supporting Abbot's claim as to the statistical significance of some of his periods, found a strongly significant period of twelve months, suggesting a possible terrestrial effect.

We see therefore that there has been considerable controversy regarding the reality of the variations shown by the solar-constant measurements, and numerous additional arguments, pro and con, could be unearthed. But the present state of affairs can be summed up as follows: The energy emitted by the sun does change from time to time. This is indicated by the changes which can be seen in the sun's surface and by the changes revealed by the ionospheric and radio measurements. However, the percentage change in the total energy output is small, amounting at most to 1 or 2 per cent [3], and whether the solar-constant measurements as observed from the earth's surface are sufficiently accurate to reveal the time and/or magnitude of such variations is still a debatable question.

This controversy may be resolved in several ways. The surest way would be to make the measurements from a satellite stationed outside the earth's atmosphere. If rockets could be adequately equipped, frequent measurements from them would also be satisfactory. Perhaps balloons may serve for this purpose. But if these methods are economically or experimentally remote, the establishment of a few additional surface stations might help. In science, important results ordinarily are not finally accepted until they have been corroborated by independent observers. Here too, if the questions raised as to the amount and time variations of the solar constant are to be answered, any additional stations should be operated by independent observers, as was long ago suggested by Marvin [58].

However, there is a more important parameter to measure than the variation of I_0; that parameter is the earth's albedo. From the meteorological viewpoint, the predominant interest is not in I_0 variations as such, but in the amount of energy absorbed by the earth and its atmosphere, and in the variations of this amount. As will be shown later, the solar energy reflected by the planet Earth varies so much that the energy absorbed may vary by ± 15 per cent or more from the mean absorbed energy. This is to be compared with a possible change of ± 1 per cent in the solar constant.

Extraterrestrial Solar Energy on a Horizontal Surface. Of fundamental importance for meteorology is the energy Q which reaches a horizontal area at the earth's surface. For purposes of comparison and to permit certain computations of Q, Milankovitch [61] computed Q_E, the extraterrestrial value of Q, from the relation

$$Q_E = \int_{t_1}^{t_2} \frac{I_0}{\rho^2} \cos Z \, dt, \tag{6}$$

where t_1 is the time of sunrise, t_2 the time of sunset, Z the sun's zenith distance, and ρ the radius vector of the earth. If we take $I_0 = 1.94$ ly min^{-1}, Q_E, in langleys per day, is given in Fig. 3.

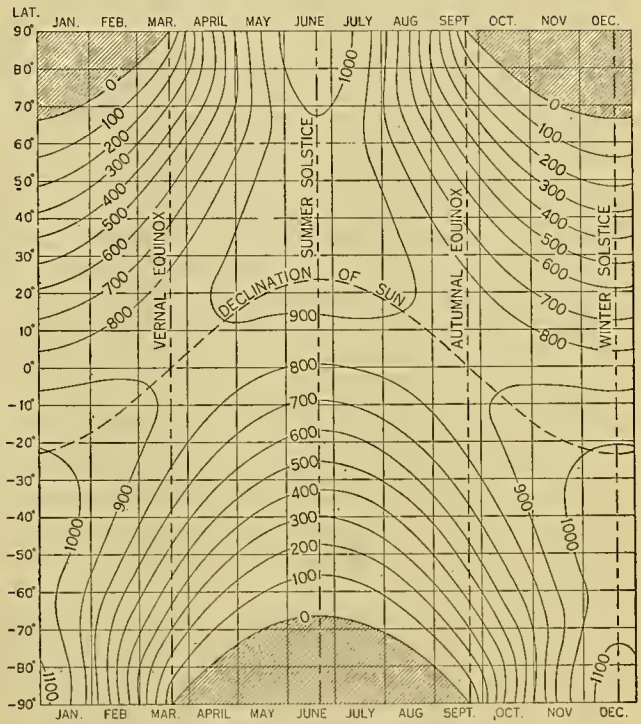


FIG. 3.—Solar radiation on a horizontal surface outside the earth's atmosphere (ly day^{-1}). (*After List* [56].)

TERRESTRIAL EFFECTS ON SOLAR RADIATION

Having considered the amount and the spectral intensity of extraterrestrial solar energy, we turn now to the effect which the earth and its atmosphere have upon the incident radiation. In general the atmospheric elements absorb and scatter part of the incident solar energy.

Absorption. At the outset it should be noted that $\lambda < 2900$ A (approximately) is not observed at the ground; nearly all energy of $\lambda < 2900$ A is absorbed and a small part is scattered back to space by the gases of the atmosphere.

The Ionosphere and Ozonosphere. Energy of $\lambda < 1000$ A is highly absorbed by O, O_2, or N_2 [62]; such energy is responsible for the ionization and heating of the ionospheric regions. For energy in the region from 1300 A to 3500 A, Craig [21] has made a thorough study

of the available absorption-coefficient measurements for O_2 and O_3; these absorption coefficients appear elsewhere in this Compendium.[4]

In equation (4), k_λ is the *extinction* coefficient; it includes the effect of both absorption and scattering. We can write

$$k_\lambda = \alpha_\lambda + s_\lambda, \tag{7}$$

where α_λ is the absorption coefficient and s_λ is the scattering coefficient. In places where s_λ is small, such as at high elevations in the atmosphere, the transmission can be written approximately

$$I_\lambda = I_{0\lambda}\, 10^{-\alpha'_\lambda x}, \tag{8}$$

where α'_λ is the decimal absorption coefficient, and x is the path length through the absorbing gas. The amount of absorbed energy is

$$I_a = \Sigma(I_{0\lambda} - I_\lambda)\Delta\lambda = \Sigma I_{0\lambda}(1 - 10^{-\alpha'_\lambda x})\Delta\lambda. \tag{9}$$

In the case of ozone, computations of I_a have usually been based on the distribution of $I_{0\lambda}$ in a black body at 6000K and give $I_a \approx 0.06\ I_0$. However, as indicated above, $I_{0\lambda}$ in the ultraviolet is considerably less than the value for a black body at 6000K. Therefore, I_a is correspondingly smaller. If one accepts the 55-km V-2 rocket and Smithsonian spectral measurements for $I_{0\lambda}$, then $I_a \approx 0.02\ I_0$ [30]. Should further measurements verify still lower values of $I_{0\lambda}$ [35, 55], I_a for ozone will have to be reduced even more.

Obviously, the assumption of 6000K black-body intensities will also yield values for the absorption at specific altitudes in the ozonosphere which are too large. Thus computations such as Karandikar's yield values that are too high for ultraviolet absorption [50]; Gowan [36] shows the effect of the lower absorption on the temperature of the air.

In addition to its absorption in the ultraviolet, ozone absorbs small amounts of solar energy in other spectral regions. One of these, the Chappuis band, extends through the visible region from 4400 A to 7600 A and has a peak at 6000 A. The absorption coefficient is, however, so much smaller than that in the ultraviolet that, despite the much higher solar intensity, the total absorption in the Chappuis band is about $0.007I_0$. In the infrared, absorptions are centered at 4.75 μ, 9.6 μ, and 14.1 μ, but $I_{0\lambda}$ is so small at those wave lengths that almost no energy is absorbed [50].

Spectral Absorption by Water Vapor. Among the atmospheric gases, water vapor absorbs the largest amount of solar energy, and for the most part our knowledge regarding these water-vapor absorptions is due to Fowle [26, 27]. For $\lambda > 0.9\,\mu$, energy is absorbed by water vapor with varying intensity in wide bands. Figure 4 shows the *positions* of these bands[5] up to about 2.1 μ and includes two bands (B and A) for oxygen absorption; Fig. 5 shows the *fractional transmission* of energy in the band widths indicated by broken arrow-headed lines at the bottom of Fig. 4. In addition to the bands shown in Fig. 4, several weak lines appear below 0.7 μ, and very strong absorption bands appear beyond 2 μ and particularly near 6 μ.

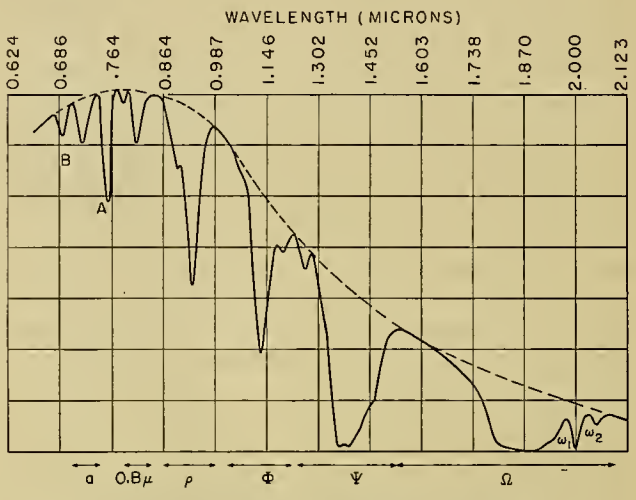


FIG. 4.—Location of absorption bands of oxygen (A and B) and of water vapor. (*After Fowle* [26].)

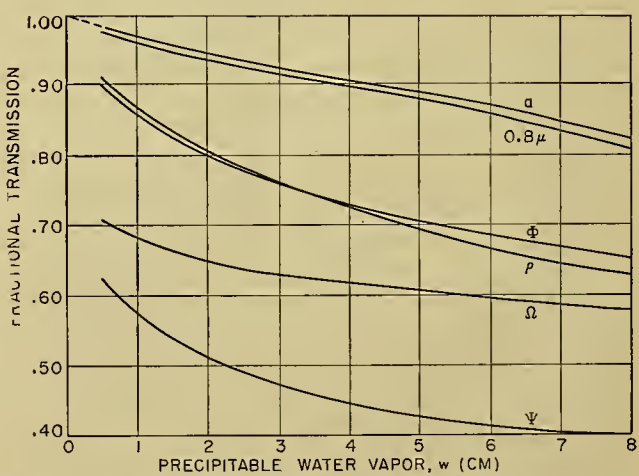


FIG. 5.—Fractional transmission of solar radiation by water-vapor absorption bands as a function of precipitable water vapor. (*After Fowle* [26].)

Fowle [27] measured the absorption in the regions 2–9 μ and some of his results are given in Table II.

Fowle's data were measured at atmospheric pressure. It has long been recognized that the fractional absorption by a water-vapor band depends on the total pressure of the air; and laboratory pressure measurements on two water-vapor absorption bands have recently been made, although the two sets of measurements lead to somewhat different results. Drummeter and Strong [24] examined the maximum and minimum absorption points in the 1.85-μ band and found that absorption at the maximum points (1.82 μ and 1.88 μ) in-

4. Consult "Radiative Temperature Changes in the Ozone Layer" by R. A. Craig, pp. 292-302.

5. The band labelled Ω really includes three bands, namely, Ω plus two small bands ω_1 and ω_2.

creased linearly with $p^{1/2}$. For the minimum points (1.87 μ and 1.90 μ) the same pressure relation existed up to p = 400 mm Hg; for higher pressure the absorption increased more slowly as $p^{1/2}$ increased. For each of the entire bands at 1.35 μ and 1.85 μ, Chapman, Howard, and Miller [19] find that the fractional transmission increases as a nonlinear function of $(p + p_w)^{1/4}$ where p_w is the partial pressure of the water vapor. They do not graph their functional absorption relation explicitly as a function of pressure.

Table II. Per Cent Absorption of Radiation by Water Vapor
(*After Fowle* [27])

Wave-length interval (μ)	Precipitable water vapor	
	0.008 cm	0.082 cm
	%	%
1.3 –1.75	6.1	18
1.75–2.2	13.6	29
2.2 –3.2	23.6	41
3.2 –4.0	21.7	37
4.0 –4.9	32.5	50
4.9 –5.4	18	42
5.4 –5.9	47	85
5.9 –6.4	64	97
6.4 –7.0	68	97
7.0 –8.0	25	62

The amount of solar energy beyond 3 μ is of course very small, and hence the absorption of that energy cannot greatly affect the temperature of the atmosphere. Karandikar [50] shows the amount of absorption in langleys per second in the various wave-length bands as a function of precipitable water vapor w for values of w from 10^{-4} cm to 1 cm and for bands in the region from 0.9 μ to 8 μ; he also gives the total amount of energy absorbed by water vapor in the stratosphere and shows that above 40 km the absorption is practically zero, but that at 30 km it approaches absorption by ozone (especially since the ozone absorption given is probably too high).

It should be pointed out that the data of Fig. 5 may not be used with Beer's law in the customary absorption computations, since the bands are too wide to be considered as monochromatic; the empirical data of Fig. 5 must be utilized. However, for many purposes the use of Beer's law will lead to practical results [63].

A few transparent regions exist in the far infrared. The region near 10 μ is relatively transparent with regard to water-vapor absorption, and although very little solar energy is available in this region, it has been used to measure atmospheric ozone which has an absorption band near 9.6 μ. Adel has found that the 17–24 μ region is also relatively transparent.

Total Water-Vapor Absorption. In order to determine the temperature change in the atmosphere produced by absorption of radiation by water vapor, a summation of absorption over all wave lengths is required, and such summations have been computed by several authors. For example, using Fowle's absorption data, Kimball [52] computed the total absorption as shown in curve (16), Fig. 6. The curve shows the fraction of I_0 absorbed by water vapor as a function of the water vapor in the sun's spectral path, that is, as a function of mw. By using data similar to these absorptions, Tanck [74] computed temperature rises of about 0.1–0.7 centigrade degrees per day at Hamburg depending on the season and the height (up to 6 km). The order of magnitude of the absorptions given by Tanck's equation agrees with airplane measurements [29].

Clouds. Thanks to Fowle and to some recent studies [19, 24], the status of water-vapor absorption is fairly well known; however, the absorption of solar energy by clouds is comparatively unknown. A few theoretical estimates have been published [45], but very few measurements have been made.

One method of measuring the amount of energy absorbed is to measure simultaneously, with pyrheliometers, the energy leaving and entering the cloud layer both at its upper and lower surfaces; the difference between the energy which enters the cloud and that which leaves it is the amount absorbed. Neiburger [66], using one blimp on which to mount two pyrheliometers, one facing upward and the other downward, made many vertical traverses through stratus clouds. When the blimp was below the clouds, he estimated the upward- and downward-flowing radiation at the top of the cloud. Because he lacked a second blimp, simultaneous measurements could not be made both below and above the cloud, and since the albedos of clouds vary appreciably over short distances and times, errors may have been introduced because of the lack of simultaneity. At any rate, Neiburger's measurements, which were probably the first absorption measurements made, indicate that in the mean about 5 to 9 per cent of the incident radiation is absorbed in stratus clouds, and that there may be large variations from the mean.

To measure the absorption in other types of clouds, especially over extensive cloud decks, the United States Weather Bureau, through the cooperation of the Air Force and the Office of Naval Research, has made measurements using B-29 airplanes as platforms for the two pyrheliometers. On a few occasions two airplanes, each equipped with two pyrheliometers, have been flown, one vertically above the other, with the cloud deck between them. When only one airplane was available, the plane, carrying two pyrheliometers, was flown above the cloud deck and above a pyrheliometer which was located on the ground. In the absence of ground snow cover, a good estimate can be made of the albedo of the ground; or if the ceiling is not too low, the albedo of the ground can be measured by flying the airplane below the clouds. Preliminary results from these measurements indicate that the absorption by these deep widespread systems averages about 20 per cent of the solar radiation incident on the cloud.[6] These measurements are still few in number and should be verified by additional determinations.

6. Reported by T. H. MacDonald at the January 1950 meeting of the American Meteorological Society in St. Louis, Missouri.

This amount of absorption is much higher than the maximum of 6 per cent which Hewson's theoretical computations indicate [45]. If it is assumed that the measurements are correct, the discrepancy can be attributed to several factors. Clouds are complex anisotropic physical entities, and the theory must make numerous assumptions to handle even isotropic clouds of *liquid* water drops. In particular, the theory does not include absorption by water vapor in clouds. A cloud whose liquid-water content is 0.5 g m^{-3} could easily contain 5 g m^{-3} of water vapor, and 10 g m^{-3} or more are quite possible. Hence with the radiation undergoing numerous reflections by the liquid water drops, the path length of a ray through the water vapor could easily be 10 to 20 times that of its path through liquid water. Furthermore, the fractional absorption of water vapor is by no means negligible by comparison with that of liquid water [6]. The absorption by the water vapor might therefore be comparable with that by liquid drops. Such an effect would bring computations such as Hewson's more in line with the measurements.

Another factor is absorption by the ice and snow particles which exist in cirrus and altostratus clouds. The mechanism of such absorption is unknown and has not been included in theoretical discussions. In view of these and other theoretical complexities, measurements for many types of clouds are required to establish firmly the magnitude of absorption by clouds. Apparently, relatively diffuse thick clouds (such as altostratus) will reflect a smaller amount of energy than less diffuse thick clouds, such as stratus or stratocumulus clouds of the same or even smaller vertical thickness [18, 32]. Mecke [59] has pointed out that, for infinitely thick clouds, high reflection will be associated with low absorption and vice versa. These ideas are in qualitative agreement with the relatively low absorption in thick stratus with its high albedo [66] and with the relatively high absorption in thick altostratus of extensive cloud systems with its low albedo.

The Earth's Surface. The energy which reaches the surface of the earth is either absorbed or reflected. The albedo of the earth's surface will be discussed later; it is the order of magnitude of the absorbed energy which should be emphasized here. In middle latitudes, at least, during cloudless conditions about 80 per cent of the energy Q_E incident in a day at the outer atmosphere reaches the ground. Except for snow-covered areas, an albedo of 10 per cent may be assumed for purposes of rough evaluation. Hence under these conditions, about 72 per cent of Q_E is absorbed by the earth's surface. This is very much larger than the absorption of 2 per cent by ozone, or of about 8 per cent by water vapor, or even of the 20 (?) per cent by extensive cloud systems. Of course, in overcast areas, the energy which is absorbed by the ground is smaller than 70 per cent and may be equal to or less than the energy absorbed in the cloud. But with average cloudiness, the ground absorption approaches about 50 per cent of the extraterrestrial radiation.

Therefore, although absorption by ozone may cause large temperature changes at 40 km because of the low atmospheric density at that height, by far the largest amount of energy which potentially becomes available for atmospheric processes is absorbed by the earth's surface itself.

Miscellaneous Absorptions. Several gases absorb minor amounts of solar energy. Oxygen, in addition to the important absorptions in the ionosphere and ozonosphere, has some minor absorptions in the near infrared (Fig. 4). Nitrogen compounds and CO_2 absorb small amounts. Carbon dioxide is, of course, of great importance in long-wave terrestrial radiation, but plays a minor role in solar radiation absorption [50]. The presence of methane has recently been announced by several authors [54]. Its role in the heat balance of the atmosphere has not yet been considered.

Scattering. In the absence of clouds, energy is depleted from the direct solar beam through absorption and scattering by air molecules, water vapor, and dust. Where absorption is negligibly small the scattering coefficient may be expressed as

$$s_\lambda = s_{a\lambda} + s_{w\lambda} + s_{d\lambda}, \tag{10}$$

where $s_{a\lambda}$, $s_{w\lambda}$, and $s_{d\lambda}$ are the scattering coefficients of pure dry air, water vapor, and dust, respectively.

Spectral Molecular Scattering. When the isotropic particles which cause scattering of energy are very small by comparison with the wave length of the light ($< 0.1\ \lambda$), the theory developed by Rayleigh [34] shows that the scattering coefficient depends on λ^{-4} or, if the mass of air in the vertical at sea level is taken as unity,

$$s_{a\lambda} = \frac{32\pi^3(n_\lambda - 1)^2 H\lambda^{-4}}{3N_a}, \tag{11}$$

where n_λ is the index of refraction of air for light of wave length λ, N_a is the number of molecules per cubic centimeter of air, and H is the height of the homogeneous atmosphere.

If scattering by air molecules alone is considered, the monochromatic energy which reaches sea level as the original parallel beam is given by

$$I_\lambda = I_{0\lambda} \exp(-s_{a\lambda} \sec Z). \tag{12}$$

Values of $s_{a\lambda}$ and of the transmission $a_{a\lambda} = I_\lambda / I_{0\lambda}$ for the Rayleigh scattering law are given by List [56]. For air molecules which are not spherical it might be necessary to multiply those values of $s_{a\lambda}$ by 1.04 [76].

Total Molecular Scattering. We are often interested not in the spectral transmissions but in the total transmission I/I_0. If equation (12) is integrated, we see that

$$\begin{aligned} I/I_0 &= (1/I_0) \int_0^\infty I_\lambda \, d\lambda \\ &= (1/I_0) \int_0^\infty I_{0\lambda} \exp(-s_{a\lambda} m) \, d\lambda. \end{aligned} \tag{13}$$

The data for $I_{0\lambda}$ are available from Smithsonian Institution measurements (Fig. 1), and several authors have computed I/I_0 from equation (13). Figure 6 shows such a computation by Kimball [52]. Curve (1), for $w = 0$,

is in excellent agreement with Linke's results [55, p. 248].

Spectral Scattering by Water Vapor. The atmosphere is, however, never pure nor dry; both dust and water vapor are ever present in varying degree. Fowle [28] examined the transmission of the atmosphere at wave lengths where water vapor does not absorb. At those wave lengths, if dust is neglected,

$$I_\lambda = I_{0\lambda} \exp(-s_{a\lambda} m) \exp(-s_{w\lambda} wm) = I_{0\lambda} (a_{a\lambda} a_{w\lambda}^w)^m = I_{0\lambda} a_\lambda^m. \quad (14)$$

Here the transmission coefficient through one dry atmosphere at vertical incidence is given by $a_{a\lambda} = \exp(-s_{a\lambda})$ and through one centimeter of precipitable water vapor by $a_{w\lambda} = \exp(-s_{w\lambda})$. From equation (14)

$$a_\lambda = a_{a\lambda} a_{w\lambda}^w, \quad (15)$$

or

$$\ln a_\lambda = \ln a_{a\lambda} + w \ln a_{w\lambda}. \quad (16)$$

A plot of $\ln a_\lambda$ against w should yield a straight line whose slope is $\ln a_{w\lambda}$ and whose intercept at $w = 0$ is $\ln a_{a\lambda}$. Hence, $a_{w\lambda}$ and $a_{a\lambda}$ can be determined.

As might be expected, for a given λ the observed corresponding values of $\ln a_\lambda$ and w do not fall exactly on a straight line. In order to remove random fluctuations Fowle compiled average values, but it is not clear whether he averaged values of a_λ or of $\ln a_\lambda$. At any rate, Fowle plotted his average $\ln a_\lambda$ against w. The points still scatter quite a bit, but the "best" straight line was drawn through the data, and the corresponding $a_{a\lambda}$ and $a_{w\lambda}$ were determined. His values of $a_{w\lambda}$ are given by List [56].

Fowle [28] assumed that the λ^{-4} law (equation (11)) applied also for scattering by water vapor and found that his transmission coefficients $a_{w\lambda}$ were such that the scattering is greater than might be expected from the number of water-vapor molecules. He concluded from this that the water vapor existed as aggregates of water-vapor molecules (see also [76]). Moon, however, using Fowle's data, plotted the logarithm of the average $a_{w\lambda}$ against $\ln \lambda$, and found that

$$a_{w\lambda} \propto \lambda^{-2}. \quad (17)$$

Any relation between $a_{w\lambda}$ and λ (such as equation (17)) should apply for a single set of measurements of $a_{w\lambda}$. It seems necessary therefore to plot $\ln \lambda$ against $\ln a_{w\lambda}$ and not against $\ln \overline{a_{w\lambda}}$. Differences in their procedure perhaps account for the difference in the wavelength laws obtained by Moon and by Fowle.

It should be pointed out that Ångström [11] questioned the validity of assuming that the water vapor was the actual scattering agent. If dust and water-vapor advection usually occur together so that an increase in one accompanies an increase in the other, it may be the dust which is actually performing the scattering. However, Fowle evaluated the effect of dust in his data, and found it to be only about one-half of 1 per cent for a_λ and 2 per cent for $a_{w\lambda}$ [28]. As these conflicting views indicate, the wave-length dependency of water-vapor scattering is not very well known.

Total Water-Vapor Scattering. Here again the total transmission is often desired in lieu of the spectral transmission. With the aid of Fowle's data for $a_{a\lambda}$ and $a_{w\lambda}$, Kimball [52] computed the fractional transmission of solar energy at normal incidence through a dustless atmosphere for various values of w. The computations, which include scattering but exclude absorption, are shown in Fig. 6 by the dashed curves (1) through (8) for values of w up to 6.0 cm. By adding the depletion due to water-vapor absorption, Kimball computed the transmissions shown in curves (9) through (15); the latter curves include the effect of both scattering and absorption.

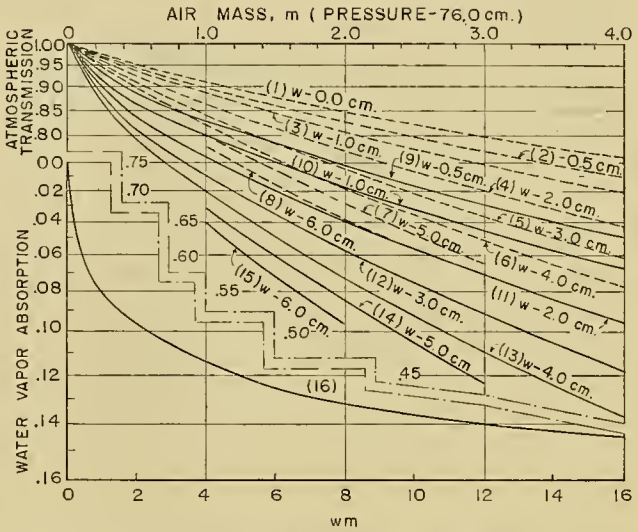


FIG. 6.—Fractional transmission of solar energy through the earth's atmosphere. Curves (1)–(8), scattering only; curves (9)–(15), scattering plus absorption by precipitable water vapor (w) amounts shown on curves; curve (16), fractional absorption by water vapor as a function of wm. (*After Kimball* [52].)

Dust. There remains the scattering from the direct solar beam by dust. Ångström [11] has derived a law for dust extinction,[7]

$$s_{d\lambda} = \beta \lambda^{-\gamma}, \quad (18)$$

where β is a constant representing the number of scattering particles, and γ another constant representing the size of the scattering particles. Ångström describes graphical methods for obtaining β either from a total radiation measurement or by filter measurement. By analyzing the spectral measurements of the Smithsonian Institution, he found $\gamma = 1.3$ on the average. From laboratory measurements of the relation between γ and the size of scattering particles, he showed that the average size of the scattering particles was 1 μ. Ives and his co-workers [48], however, have found from actual particle measurements near the ground in cities that the most frequent diameter size is near 0.5 μ. Their

7. Ångström's extinction coefficient, here designated by $s_{d\lambda}$, includes scattering by water vapor, but not absorption.

findings are valid only for particles larger than 0.2 μ in diameter because their microscope could not detect smaller particles. Recently Crozier and Seely [22], making measurements from airplanes, found the greatest frequency of particle diameter in the free air to be 3–6 μ; their instruments could not catch the smaller particles efficiently. Linke and von dem Borne [55] showed that γ varied with the altitude of the observing station and from place to place. In general, for higher altitude and less smoky areas, γ was larger than elsewhere, and hence the average particle size was smaller.

All these are average values, and both β and γ vary with time even at one observing station, for the number, size, and shape of the scattering particles vary continually. Also, at any time the whole spectrum of particle sizes ranging from below 0.2 μ to 1 μ and larger are to be expected.

This complex scattering by particles which are larger than molecules is a serious problem in spectral measurements. In particular, dust is troublesome in ozone measurements which ordinarily utilize light at 3000–4000 A where differential spectral scattering usually becomes an important problem. Ramanathan and Karandikar [70] found that peculiar ozone effects can easily be produced by improper assumptions regarding dust. They also found, as might be expected, that γ is not really constant and that in India it lies on the average between 0 and 1.3 in the region 3000–4450 A.

Another measure of the dust parameter, the "turbidity factor," was designed by Linke. He defines the turbidity factor τ from

$$I_\lambda = I_{0\lambda} \exp(-k_{a\lambda}\tau_\lambda m), \tag{19}$$

where

$$k_{a\lambda}\tau_\lambda = k_{a\lambda} + k_{w\lambda}w + k_{d\lambda}d,$$

$k_{a\lambda}$, $k_{w\lambda}$, and $k_{d\lambda}$ are the extinction coefficients of air, water vapor, and dust, respectively, and d represents the "quantity" of dust [55, p. 266]. He uses an "appropriate" average value and integrates equation (19), so that

$$I = I_0 \exp(-k_a\tau m), \tag{20}$$

where k_a and τ now represent mean values over wave length. Such averages cannot really be obtained, so that the discussion which follows represents an approximation.

Let I'_m designate the intensity of light transmitted through a pure dry atmosphere; then $e^{-k_a m} = I'_m/I_0$. Equation (20) can now be written

$$\tau = \frac{\ln I_0 - \ln I}{\ln I_0 - \ln I'_m}. \tag{21}$$

By means of equation (21), τ can readily be computed from a single measurement of I, for I_0 is a constant (generally taken as 1.94) and I'_m can be obtained from graphs such as Fig. 6.

In equation (20), τ can be interpreted as the number of pure, dry air masses which would produce the extinction observed in the moist, dusty atmosphere. It was found, however, [42, 55] that τ was a function of m and thus was not a reliable measure of turbidity. To overcome this defect, Linke [55] decided to relate τ, not to a dry air mass, but to a dustless air mass containing 1 cm of precipitable water vapor. Instead of I'_m in equation (21), we therefore introduce $I'_{m,w=1}$. The new turbidity factor becomes

$$\Theta = \Phi_m \ln (I_0/I) \tag{22}$$

where

$$\Phi_m = \frac{1}{\ln I_0 - \ln I'_{m,w=1}}.$$

Linke's values of Φ_m are given in Table III.

Table III. Values of Φ_m in Equation (22)
(*After Linke* [55])

m	0.5	1	1.5	2	3	4	6	8	10
Φ_m	23.35	13.57	9.60	7.97	6.04	5.02	3.80	3.10	2.63

With the aid of a single pyrheliometric measurement, Θ can be computed. According to Linke, it will vary relatively little with air mass; any observed diurnal variation of Θ will be a real variation in turbidity. Turbidity factors for portions of the spectrum have also been designed and can be measured with filters [55].

Neither Ångström nor Linke attempted to separate scattering by water vapor from scattering by dust, and indeed, as Ångström pointed out, it may not be valid to do so. However, Kimball [52] assumed that scattering by dust could be separated. He estimated w from either surface vapor pressure or radiosonde data. He measured I and computed the transmission $a = I/I_0$. From Fig. 6, he found $a_{m,w}$, the transmission through a dustless atmosphere containing w cm of precipitable water vapor. Then $a_{m,w} - a = d_d$, the fraction of I_0 which is depleted by the dust. The value of d_d, of course, varies with m. Klein [53] gives some values of d_d which vary from 0 to 0.09 for $m = 1$, and from 0 to 0.13 for $m = 2$.

Wexler [78] and Haurwitz [42] have discussed the turbidity factors of Ångström and Linke and conclude that neither is wholly satisfactory. These factors are, however, among the best simple methods available at present for determining the atmospheric turbidity quantitatively. Although expressions of the total turbidity such as Θ may have some specialized uses, for spectral measurements such as those involved in ozone and solar-constant measurements we shall have to know more about the way dust affects light of various wave lengths. This will probably involve determination of the size and number of dust particles above an observer, which in turn may be helpful in studies of condensation nuclei.

Scattering by Liquid Water Droplets. Scattering of light by spherical particles which are not very small in comparison with the wave length is given by the well-known Stratton-Houghton curve. Recently Houghton and Chalker [46] have extended the earlier computation; their results appear in Fig. 7. The extinction by spherical water particles (refractive index 1.33) is re-

lated to $\pi r^2 N K_s$, or K_s times their geometrical cross section, so that

$$I_\lambda = I_{0\lambda} \exp(-\Sigma \pi r^2 N K_s), \qquad (23)$$

where r is the radius of the drops and N is their number. Figure 7 shows K_s as a function of the parameter $X = 2\pi r/\lambda$. Equation (23) holds for nonabsorbing particles and is valid at wave lengths where water absorption is negligibly small; the summation is required over drop radii when the drop sizes are not uniform. For particles where $n_\lambda \neq 1.33$, K_s will differ from the values of Fig. 7; van de Hulst [76] shows some variations. From Fig. 7 it is important to note that for some ratios of particle size to wave length the scattering does not always increase with decreasing wave length.

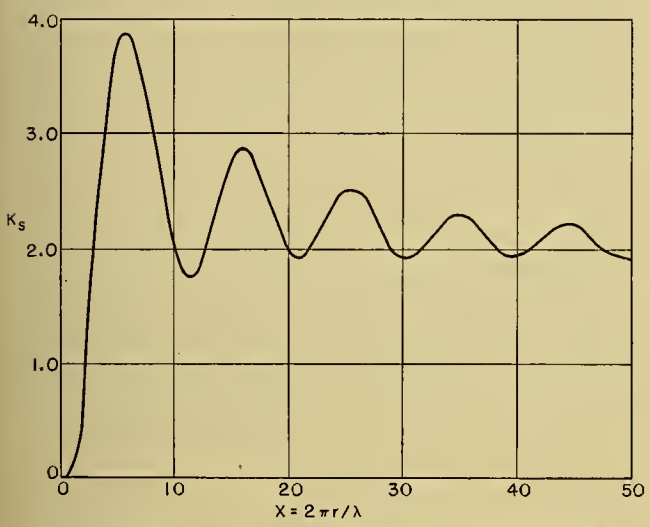


FIG. 7.—Scattering-area coefficient K_s for liquid-water drops in nonabsorbed spectral regions as a function of λ and of drop radius r. (*After Houghton and Chalker* [46].)

Multiple Scattering. In the case of pure Rayleigh scattering (if it ever can be said to exist in the atmosphere) the scattering is symmetrical about the particle so that, for example, the forward- and backward-scattered energy are equal. But as the particles become larger the scattering increases in the forward direction [34; 55, p. 161]. Calculations of the effect of such particles, especially for multiple scattering, become rather complex; however they have been undertaken by several authors [76].

The energy which is scattered from the original solar beam will in general be scattered more than once on its way down to the earth's surface or else out to space. From an empirical viewpoint, regarding the contribution of the cloudless sky radiation to the total radiation on a horizontal surface, Kimball [52] states that of the radiation scattered from the direct solar beam, half will be scattered down and half up. In a pure Rayleigh atmosphere this would be the case. But for the actual conditions of the atmosphere, it serves only as a useful approximation for average daily values of cloudless-sky radiation after the total scattering from the direct beam has been evaluated.

For instantaneous values, Kimball found the ratio of the direct sunlight component on a horizontal surface I_h to the total solar and sky radiation on a horizontal surface Q to be a function of the zenith distance of the sun (Table IV).

TABLE IV. VALUES OF I_h/Q AS A FUNCTION OF SUN'S ZENITH DISTANCE Z
(*After Kimball* [51])

Z (deg)	30.0	48.3	60.0	66.5	70.7	73.6	75.7	77.4	78.7	79.8
I_h/Q	0.84	0.84	0.80	0.78	0.76	0.72	0.69	0.67	0.65	0.63

Similar values have also been found by other observers [55, p. 356]. If we designate the diffuse sky radiation by D, then $D = Q - I_h$. As might be expected, the ratio D/Q increases where the atmospheric turbidity is high [55]. Measurements illustrating this point (*e.g.* illumination measurements) have often been made with instruments for relatively narrow spectral bands. It would be desirable to measure the ratio for nonspectral radiation with varying atmospheric transmissions or turbidity factors.

Albedo. By the albedo of a body we mean the fraction of the incident energy which is reflected by the body. Thus the albedo A of the planet Earth is the fraction of the energy incident at the "outer limits" of the atmosphere which is returned to space; the albedo of a cloud "surface" is the fraction of the energy incident upon the cloud which is reflected by the cloud.

The terrestrial entities which reflect solar energy are (1) the cloudless atmosphere, (2) clouds, and (3) the earth's surface. The sum of these three reflections is the total energy reflected by the Earth. Expressed as a fraction of the extraterrestrial solar energy intercepted by the Earth, this sum determines A.

The Albedo of the Cloudless Atmosphere. 1. *Pure Dry Air.* We have seen that within a few per cent the fractional spectral depletion of solar energy by molecular scattering is represented by the Rayleigh scattering law, and the summation over wave length is readily given as a function of air mass from curve (1) of Fig. 6. The earth as seen from the sun can be considered as made up of narrow concentric circular rings centered about the subsolar point. To an observer in a particular ring the sun is at a specified zenith distance, and hence in each ring the optical air mass m is known. From m and the area of the rings, together with the above-mentioned relation between scattering and air mass, the amount of energy scattered by the entire atmosphere can readily be calculated and found to be about 15 per cent of the incident energy [30]. This was calculated using Fowle's data; Fig. 6 gives a slightly smaller value. However, this energy is scattered in all directions. To determine the fraction scattered upward (away from the earth's surface) it is necessary to assume something about the angular scattering by each particle. Scattering by small particles, such as molecules, is symmetrical about the particle [34]. Hence as much energy is scattered up as is scattered down. Even for

multiple scattering this rule is closely obeyed. Consequently from 7 to 8 per cent of the incident solar energy would be scattered back by the pure, dry, cloudless atmosphere.

2. *Water Vapor.* The scattering by water vapor and questions concerning it were discussed earlier. If one accepts $1 - a_w$ as the depletion by water vapor due to scattering, then scattering for the total spectrum can be computed as a function of air mass, as was done by Fowle, and also by Kimball (Fig. 6). From this functional relation, the scattering by water vapor over the whole earth can be computed. But for this, of course, data on the distribution of water vapor with latitude are necessary. From a rough estimate of w, obtained from the distribution of surface vapor pressure, and after weighting the areas of the earth involved, it is found that 10 per cent of the incoming solar energy is reflected by water vapor [30].

This, again, represents the energy scattered in all directions, but the portion which is scattered back to space is not so readily ascertained as it is for air molecules. Fowle believed that the scattering particles were aggregates of water-vapor molecules; Ångström suggested that dust might be the scattering agent. In either case, it is recognized that if the size of the particles approaches the wave length, the forward scattering exceeds the backward scattering. At any rate, two extremes can be postulated: (1) Either half the scattered energy (5 per cent) is directed upward; or (2) none of the scattered energy is directed upward.

3. *Dust.* Concerning dust our knowledge is most meagre. Not only is the transmission coefficient for dust in doubt, but the distribution of dust over the world is inadequately known. Klein [53] gives a compilation of some values which serve as a guide, but even these "measured" values may be inaccurate in view of the inadequate understanding of dust depletion. From Klein's data one may estimate a depletion of about 5 per cent of the initial radiation on a world-wide basis. In view of the probable greater forward scattering and some absorption, less than 2.5 per cent is scattered back. An estimate of 1 per cent might be about right.

4. *Total Atmospheric Reflection.* A summation of the reflection by atmospheric elements gives a reflection lying between 8 and 13 per cent, depending on how much is allowed for upward scattering by water vapor and dust. There are not very many direct measurements which yield the reflection by the atmosphere. In the lower layers of the atmosphere, some airplane measurements [29] indicate that about 5 per cent of the energy incident at 10,000 ft is scattered upward by the air below 10,000 ft. These measurements were made with a solar zenith distance Z of 53° on a rather smoky day. At greater heights and zenith distances greater reflection presumably would have been measured. Teele [75] measured the visible energy scattered back at 72,000 ft. He interpreted his measurements as indicating that 6 per cent of the incident energy is reflected by the air when Z is 60°.

These measurements and calculations are valid for the cloudless atmosphere only. For the average condition of 0.54 cloudiness, calculation indicates that the cloudless portions of the atmosphere reflect about 6 or 9 per cent of the incident solar energy, depending on the assumptions made, when about 2.7 per cent is included for the reflection by the air above the clouds.

Clouds. Most of the solar energy reflected to space is reflected by clouds. However, the albedo of clouds is so variable, even for one type of cloud, and our information about the spatial distribution of cloud types is so poor that it is impossible, from individual cloud-albedo measurements, to specify an average value over time and space for the total reflection by clouds. Hence, in order to estimate the average albedo of clouds, it is necessary to measure the albedo of the whole earth first and then to subtract the contributions by the atmosphere and by the ground. The remainder is the albedo of clouds.

Table V. Albedo of Various Cloud Types

Source	Cloud type	Albedo (per cent)
Luckiesh [57] (visible light)	Very dense clouds of extensive area and great depth	78
	Dense clouds, quite opaque	55–62
	Dense clouds, nearly opaque	44
	Thin clouds	36–40
Fritz [32] (total radiation, extensive systems; cloud types below measured clouds not specified)	Stratocumulus, overcast	56–81
	Altostratus, occasional breaks	17–36
	Altostratus, overcast	39–59
	Cirrostratus and altostratus	49–64
	Cirrostratus, overcast	44–50
Aldrich [7]	Stratus, 600–1000 ft thick	78

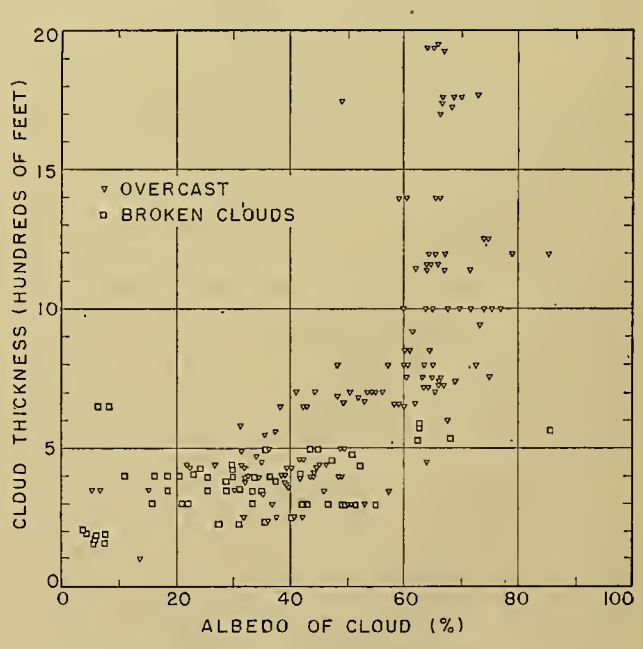


Fig. 8.—Observed albedo of stratus clouds. (*After Neiburger* [66].)

Such a computation [30], based on lunar measurements (to be described later), indicates that the average albedo of clouds is about 50–55 per cent of the energy incident on the clouds. This value may be compared with several series of measurements given in Table V and in Fig. 8.

It will be noted that 80 per cent seems to be a very high value, and that the 78 per cent which Aldrich [7] measured over a single stratus cloud is not a suitable average. According to Bullrich [18], for average clouds of infinite thickness the albedos by cloud types will be as follows: stratocumulus, 0.78; stratus, 0.74; and altostratus, 0.46. A low albedo (39 to 59 per cent) for the altostratus type of cloud can be seen in the measurements in Table V.

Albedo of the Earth's Surface. The albedo of the ground has a very wide range of values also. In general, the lowest albedos occur over water. Forests or other configurations which can trap solar energy act nearly as black bodies and therefore likewise have low albedos. On the other hand, snow-covered terrain has a very high albedo; values up to 90 per cent have been measured over fresh snow. As the snow becomes older or turns to ice, its albedo decreases markedly, and it absorbs more energy. Data for several types of terrain are given in Table VI. It will be noted that general terms such as "grass" give only an approximate estimate of the albedo, so that for investigation of such quantities as local turbulence near the ground, the albedo at the time and near the place in question will have to be measured.

Table VI. Albedo of Various Surfaces
(*After List* [56])

Surface	Albedo (per cent)
Desert	24–28
Fields, various types	3–25
Forest, green	3–10
Grass, various conditions	14–37
Ground, bare	7–20
Mold, black	8–14
Sand, dry	18
Sand, wet	9
Snow or ice	46–86
Water (direct sun* only) Z (degrees)	
0	2.0
20	2.1
40	2.5
50	3.4
60	6.0
70	13.4
80	34.8
85	58.4
90	100.0

* For reflection of sun *plus* sky radiation Ångström [10] gives:

Z (deg)	43.0	46.9	70.5	77.9	84.5
A (per cent)	3.9	5.7	13.8	30.0	46.5

The albedo of large water surfaces depends very largely on the angle of incidence of the light, or the sun's zenith distance Z. Under cloudless conditions, the direct solar radiation follows Fresnel's law of reflection very closely even when the wind is 10 mph [10]. For small values of Z (sun overhead) the reflectivity is very low, and it is not until the angle reaches about 65° that the albedo for solar and sky radiation becomes as much as 10 per cent. Under overcast skies the albedo of the sea surface is about 10 per cent, but changes a little even then with solar altitude and cloud thickness [65]. When the wind velocity is large enough to cause whitecaps, the albedo of the water increases, and Brooks [17, p. 460] gives 31 per cent as the albedo when the water surface is rough. Ångström states, however, that in ordinary geophysical problems the data of Table VI are applicable.

The low albedo of water and, in general, of land means that the contributions of the earth's surface to the total albedo of the planet Earth must be small. As a rule, areas which are snow-covered receive little sunlight; the same may be said for water areas which have a high albedo (low solar altitudes). The net result is that, when the average cloudiness is considered, the earth's surface contributes less than 4 per cent of the incident energy to the total albedo of the earth; this contribution is probably between 2 and 3 per cent [30].

The Albedo of the Planet Earth. The sum of the albedos of the various entities is the albedo A of the planet Earth. As stated earlier, since there is a variability of the albedo of clouds, and since clouds contribute most to the albedo of the planet Earth, A can be found only by measurements from or on bodies outside the earth.

Danjon [23] has made such measurements by viewing the moon with a suitable photometer. The moon is illuminated by two sources of light. The light side of the moon is illuminated directly by the sun; the dark side, by the earth. The light from the earth is *sunlight* which has been reflected by the earth to the dark side of the moon. Consequently, the brightness on the dark side of the moon, as compared with the brightness on the light side, is a measure of the sunlight reflected by the earth, that is, it is a measure of A. Of course, such measurements involve many difficulties, in addition to observational ones. For example, the spectral distribution of the earth's light is different from that of sunlight and doubtless changes with the albedo itself. Hence, any selective reflectivity by the moon would introduce errors. Another problem is the stellar magnitude of the moon. When his paper was already in press, Danjon learned of a new value of the moon's brightness which would somewhat alter his calculations. However, the value which Danjon originally used for the moon's brightness is still quoted in some astronomy textbooks, so his calculations may be correct. Danjon's average value of the earth's albedo for visible light (about 0.4 μ to 0.74 μ) is 39 per cent. But visible light comprises only about half, or even a little less, of the total extraterrestrial solar energy. Hence, to determine the total albedo, adjustments to his measured values must be made for the ultraviolet and infrared portions of the solar energy. Calculation of the corrections gives an albedo of about 28 per cent for the infrared, about 50 per cent for the ultraviolet, and 35 per cent for the total sunlight. The reflection in some portions of the ultraviolet where ozone does not absorb energy, as at 0.36 μ for example, may be considerably higher than 50 per

cent, which is the calculated reflection for the whole ultraviolet radiation ($\lambda < 0.4\ \mu$).

Several other estimates of A have been computed by assuming some value for the average reflection by clouds and/or by studying the average transmission to the earth's surface and estimating the amount of energy absorbed by the atmosphere and clouds. Notable among these calculations of A are the 43 per cent of Aldrich [7] and the 41.5 per cent of Baur and Philipps [14].

In addition to the average value of the earth's albedo, Danjon found large fluctuations of the albedo from season to season and among his individual measurements. Some average values of the visual albedo varied from 30 per cent in August to 50 per cent in October. This large variation is not too surprising. Nor does absence of a relation either to snow cover or cloud amount cast serious doubt on this albedo variability. The amount of solar energy incident on the snow-covered areas is small so that they contribute little to the total albedo. As for cloud amount, even if this latter quantity were known on a world-wide basis, it would not uniquely determine the contribution of clouds to the earth's albedo because the albedos of individual cloud types themselves are extremely variable.

The meteorological significance of this large albedo variation is obvious. If there is any possibility, as several authors believe, that variations in the solar constant of about 1–2 per cent could influence the state of the weather, then the much larger changes in the reflected energy and hence in the absorbed energy must be significant indeed. Only the absorbed energy can affect the atmosphere. Perhaps short-period changes in the solar constant cause changes in the earth's albedo, but such a causal relationship is not obvious a priori and would have to be established before it could be accepted. At any rate, whatever the cause of its change, measurements of the earth's albedo would indicate the *large* changes in the solar energy which is absorbed and potentially becomes available for meteorological processes.

The best way to measure the albedo, as with other solar-variation effects, is to mount instruments on a satellite stationed outside the earth. In the absence of such a satellite, techniques similar to Danjon's must suffice and his results will have to be verified. Danjon's technique would require a little elaboration. For example, from measurements in France, reflections from the Pacific Ocean area would not influence the moon's dark-side illumination. Hence, measurements from widely separated regions would be needed to determine the albedo of the whole earth. Also, as stated previously, an estimate of the ultraviolet and infrared albedo would also be required. However, the main changes in albedo from day to day would be given by visual albedos alone if the corrections should prove difficult to obtain.

Solar Energy at the Earth's Surface. We have considered the depletion of solar energy through scattering and absorption by the various constituents of the atmosphere. The remainder of the direct radiation reaches the ground, and of course, a considerable portion of the scattered radiation also arrives at the ground as radiation from either clear or cloudy skies.

Cloudless Sky. Figure 6 shows the fractional transmission at normal incidence to the sun as a function of air mass and of water-vapor content for a moist, dustless atmosphere. Multiplication by the solar constant will give the transmission in absolute units. For many purposes it will be of greater interest to determine the cloudless-sky energy Q_0 which reaches the horizontal

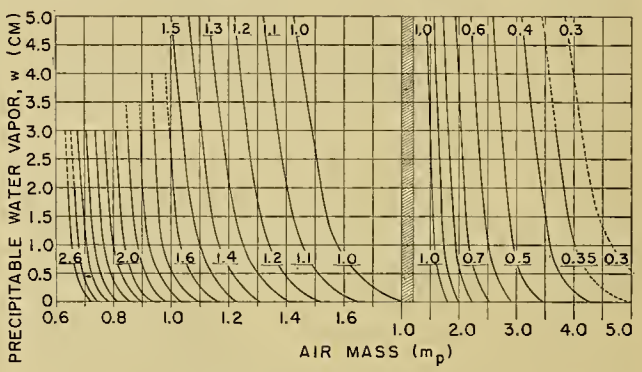


FIG. 9.—Intensity of solar radiation on a horizontal surface (ly min^{-1}) as a function of air mass m_p and precipitable water vapor w for a cloudless, dustless atmosphere. For elevated stations where pressure is p in millibars, multiply radiation by $p/1013$. Dashed lines represent extrapolated values. (*After Fritz* [31].)

ground. Figure 9 shows the radiation on a horizontal surface during dust-free conditions, as a function of optical air mass and precipitable water-vapor content. The computations are based on Fig. 6 and a combination of equations from Klein [31, 53]; it is assumed that half the scattered radiation reaches the ground. Much of the scattered radiation does reach the ground, but obviously this final result is only an approximation which agrees fairly well with observations and with other calculations. Summations over 24 hours (with the aid of Fig. 9) furnish daily totals which, when corrected by comparison with observations at several locations, show the geographic and seasonal variation of cloudless-day radiation over large areas [31]. Such computations indicate that in the United States about 80 per cent of the incident extraterrestrial energy reaches the ground during cloudless days.

Some variations from average values are naturally to be expected. In cities particularly, one would expect marked average decreases of the order of about 20 per cent [40]. For snow-covered terrain, increased radiation will be measured because the strongly reflected radiation will be partially scattered back to the ground by the atmosphere.

Transmission through Clouds. Clouds, however, will generally introduce the largest variations of the solar energy which reaches the ground. Haurwitz [43] has given the *average* solar energy transmitted to a horizontal surface through various types of clouds at Blue Hill, Massachusetts, and also the percentage of cloudless-sky radiation transmitted. His data are shown in Fig. 10 and in Table VII. Large variations about these

average values are of course to be expected for individual cases.

The effect of snow in increasing the solar radiation will be particularly pronounced for overcast conditions because the energy which is reflected by the snow will be strongly reflected towards the ground by the base of the cloud. This is shown for visible radiation in Fig. 11 [49]. For wider application the data of Table VII and

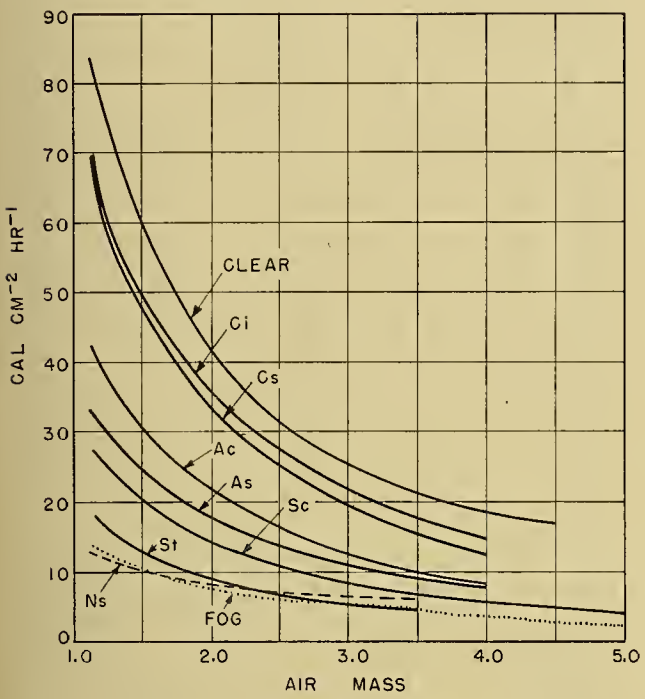


Fig. 10.—Solar radiation on a horizontal surface through the types of clouds indicated. (*After Haurwitz* [43].)

Table VII. Ratio of Solar Radiation with Overcast Skies to Solar Radiation for Cloudless Skies
(in per cent)
(*After Haurwitz* [43])

Air mass	*Ci*	*Cs*	*Ac*	*As*	*Sc*	*St*	*Ns*	Fog
1.1	85	84	52	41	35	25	15	17
1.5	84	81	51	41	34	25	17	17
2.0	84	78	50	41	34	25	19	17
2.5	83	74	49	41	33	25	21	18
3.0	82	71	47	41	32	24	25	18
3.5	81	68	46	41	31	24		18
4.0	80	65	45	41	31			18
4.5					30			19
5.0					29			19

Fig. 10 should be separated for conditions of snow-covered and snow-free ground. Also similar data should be computed for areas in which the climate is different from that of Blue Hill.

Average Conditions. We turn now to the estimation of the average solar radiation Q which reaches the earth's horizontal surface on the average for all days. It is commonly assumed that

$$Q = Q_0 (a + bS) \tag{24}$$

will give average values of Q; here a and b are constants such that $a + b = 1$, S is the percentage of possible sunshine, and Q_0 is the value of Q when $S = 100$ (clear skies). This was really already assumed implicitly by Kimball in 1919 [51]. From examination of the radiation on individual days, average values of a have been found for $S = 0$ and/or for $c = 10$ (overcast) by many

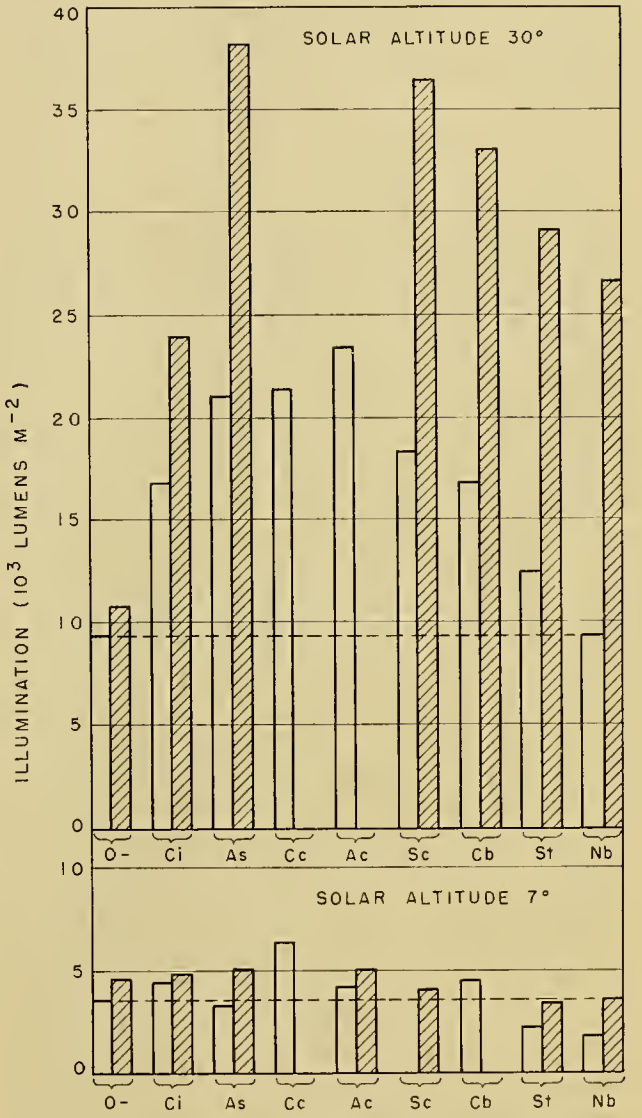


Fig. 11.—Diffuse visible skylight for different types of clouds (10/10 sky cover) above ground with snow cover (hatched bars) and without snow cover (clear bars) for solar altitudes of 7° and 30°. Dashed horizontal lines apply for clear skies. (*After Kalitin* [49].)

observers. Kimball found $a = 0.22$ for Washington, D. C., and obtained

$$Q/Q_0 = (0.22 + 0.78S). \tag{25}$$

Ångström found $a = 0.235$ for Stockholm; for annual values at Blue Hill, Haurwitz found $a = 0.22$ or 0.30, depending on the assumptions; Mosby obtained $a = 0.54$ for the Arctic [42]. Recently several other authors have found values of a which vary considerably. Never-

theless, when Q for average cloudiness conditions is desired, equation (25) is commonly used.

In equation (24), a is usually determined from measurements of Q for individual days on which $S = 0$ and $S = 100$. Very few places have an average of $S = 0$ or $S = 100$ for a period of a month, and it is not certain that equation (24) will yield average values of Q for periods which are made up of a mixture of clear, partly cloudy, and overcast days. A study of all suitable data in the United States [33] in which only monthly mean values of Q and Q_0 were used gave the equation

$$Q = Q_0 (0.35 + 0.61S), \tag{26}$$

with a correlation coefficient of $+0.88$ between Q/Q_0 and S. Here Q_0 is the cloudless-day radiation. An examination of the data (unpublished) revealed that in the western part of the country a was higher than in the eastern part.

That a should vary is to be expected from Haurwitz's data (Fig. 10). The frequency of cloud types is not often available, but it can be seen that areas with greater frequencies of high clouds will have higher values of a than areas with high frequency of low clouds. Hence, the application of one equation such as equation (25) or (26) everywhere should be expected to lead to some inaccuracies even when the *amount* of cloudiness or of sunshine is known. Fortunately, for a considerable range of S, equation (25) or equation (26) will give very similar results for Q/Q_0. Only in very cloudy or relatively clear areas will there be a difference in the result. For lack of more detailed information, Kimball applied equation (25) to the best cloud-amount data available to him and computed maps showing the distribution of radiation over the ocean area [52]. However, if the frequency of cloud types in some ocean areas is different from the average for the eastern United States, results different from those of Kimball should be expected.

The computations from any of these formulas can approximate only long-term average values. At any one place and time large fluctuations from the normal value occur. For example, at Washington, the total radiation on a horizontal surface for a particular February, say, may vary by as much as $\pm$ 20 per cent from the average for all Februaries and bear very little relation to S for that particular February. Ångström [11] notes that in the short interval 1923–28 the *annual* total of radiation at Stockholm varied by $\pm$ 10 per cent from the mean for the six-year period. There is ample evidence therefore of large variations in the radiation that reaches the ground at particular points on the earth's surface. If the variations which Danjon found in the earth's albedo are verified, large fluctuations also occur in the total radiation that reaches the *entire* earth's surface in a period of a month or even of a year. Except for regions where clouds are rare, areas of below or above normal radiation (or surface heating) may occur anywhere, being influenced predominantly by the prevalence of the amount and type of clouds.

GENERAL SPECULATION AND CONCLUSION

From the thoughts implied or expressed earlier, we can extract a few for amplification and emphasis.

Solar Radiation at the Ground. The solar energy Q absorbed by a unit area of the earth's surface is by far the largest portion of the absorbed energy. It is customary when discussing the average general circulation to consider that the greatest heating occurs at the equator where the air rises and that a whole chain of events takes place as a consequence. This equatorial heating is supposed to represent an *average* condition. Actually, even as an average condition, it is not a true picture for the summer hemisphere. In the Northern Hemisphere summer, average Q is distributed rather uniformly with latitude, so that there is no pronounced maximum of radiation near any one latitude, and the maximum, such as it is, probably occurs near lat 40°N rather than at lower latitudes. The maximum surface temperatures occur in the middle of the large cloudless land masses, not necessarily at the equator.

Let us speculate somewhat on the possible significance of the fairly uniform surface heating in summer in middle latitudes. In the annual cycle of temperature it is well known that in inland areas the maximum of the average surface air temperature lags the maximum solar radiation by about four to eight weeks. Since advection in summer is generally small, the heat balance at the ground *in situ* may be an important factor in determining the future average air temperature over a period of time such as a week or a month. In effect the ground acts as a heat reservoir; it gets hot to a considerable depth under the influence of the solar radiation which it receives, and it gives up its heat gradually to the atmosphere.

The lag between the average surface air temperature and the solar radiation at the ground occurs in the mean or normal situation. What happens in any particular summertime week or month? As indicated earlier, the amount of solar radiation which reaches the ground in a period of a week or month may differ markedly from the normal value of solar radiation. Suppose the solar radiation received during a particular June is much higher than normal. Perhaps we can assume that the ground would get hotter (over and above its normal rate of heating) than it had been in the previous May and that the effect of the hotter ground would be to modify the air temperature towards higher temperature in the following July.

Actually, we should of course examine the heat balance at the ground and not the incoming solar energy alone. The solar energy is regularly measured in many places, and in the United States the network of such measurements is now nationwide. But the heat losses at the ground are not regularly measured as universally. In the face of such an observational deficiency, is it justifiable as a first approximation to assume that the variation of incoming solar energy in summer predominates over other effects, such as variations in outgoing radiation? If this assumption is justified, and summertime advection is not a predominant effect, an ano-

malous solar energy Q might influence the air temperature in inland areas with a lag in the same way as the normal surface air temperature lags the normal solar radiation. If such an effect could be found, the forecasting value of the observed solar radiation would be obvious.

Other studies involving the observed anomalous radiation may be profitable. The average annual cycle of atmospheric circulation is of course related to, if not entirely caused by, the change in the gradient of solar heating of the earth's surface. Here again perhaps study of the observed anomalous radiation distribution for a week or a month may reveal it to be a factor in the *cause* of large-scale weather changes.

The Albedo of the Earth. To a certain extent the previous remarks concerning the significance of the distribution of surface heating apply also to the heating of the entire planet. The planet-wide heating cannot be observed at present from surface observations. However, by Danjon's technique the earth's albedo can be observed, although corroborating independent observations are highly desirable. Furthermore, by making observations from several longitudes it may even be possible to estimate the longitudinal distribution of the albedo and therefore of the heating. The variations of these albedos from their average values may be related to the subsequent exchange of heat and of masses of the atmosphere between latitudes and between longitudes.

Upper-Atmospheric Heating. The uncertainties regarding the extent of heating in the ozone layer were discussed earlier and a quasi-practical experiment to determine the fluctuation of radiation during SID's was proposed. It is, of course, possible to attempt correlations of numerous solar parameters with meteorological parameters. The number, size, shape, and polarity of sun spots, as well as faculae, flocculi, prominences, and coronal lines are but a few of the parameters which have been or could be used. However, it seems to the author that if any relation does exist between solar variability and tropospheric meteorology over a period of a few days or longer, those solar parameters which are *known* to produce effects somewhere in the upper atmosphere would offer the greatest promise, and it would be better to seek correlations not with the solar variation itself but rather with the known terrestrial effect which has been produced. Prominent among such solar-induced terrestrial effects are SID's and magnetic variations.

The suggestion then is that we measure the variation of solar ultraviolet radiation which reaches the top of the ozone layer, and the variation of temperature there. Until such measurements are made we cannot be sure that solar variation is heating the ozone layer significantly above its normal temperature. After the magnitude of ozone heating has been established, correlations of tropospheric effects with solar parameters which produce the ozone heating would obviously be indicated. In the meantime, the SID seems to be a parameter which is likely to be related to abnormal ozone heating; as such it *may* be related to surface pressure changes.

Absorption by Clouds. Since some of the measurements suggest large absorption of solar energy by clouds, it would be desirable to investigate the absorption further. This can be done from airplanes, as described earlier; but it may also be feasible to investigate the absorption from the ground. This might be accomplished by spectroscopic measurements on clouds from $\lambda = 0.5\ \mu$ to $\lambda = 2.5\ \mu$. On the basis of the data of Fig. 1 and our knowledge of the spectral distribution of clear-sky light [55], we can approximate the spectral distribution of the solar energy which irradiates the tops of clouds. Furthermore, Fig. 7 shows the extinction of parallel light by droplets in the *absence of absorption*, and shows that if the droplets are large ($2\pi r/\lambda > 50$), K_s becomes nearly independent of λ for the spectral region between $0.5\ \mu$ and $2.5\ \mu$. This ought to be true also for diffuse radiation such as that produced by clouds. At $0.5\ \mu$ the absorption coefficient of liquid water is in reality very small, so that the extinction by clouds at $0.5\ \mu$ is caused wholly by scattering. Consequently $I_{0.5}$, the measure of spectral intensity at $0.5\ \mu$, can serve as a standard against which to compare the spectral intensities at longer wave lengths such as $2.0\ \mu$.

From Fig. 7, together with some reasonable assumption about the droplet sizes in the cloud and the relative spectral irradiation of the cloud top, we can calculate $I_{2.0}$ relative to $I_{0.5}$ on the assumption that no absorption is present. The difference between the calculated $I_{2.0}$ and the observed $I_{2.0}$ might serve as a first approximation to the absorption. Greater refinement can probably be obtained from Mecke's theoretical discussions [59].

Conclusion. We have discussed the state of our knowledge in the field of solar radiation and speculated about some of the deficiencies in that knowledge and in its application to meteorology. On the question of the solar energy potentially available for "use," the suggestions (1) that the albedo (35 per cent) of the earth is smaller than the value (43 per cent) which has often been accepted, and (2) that the absorption by clouds is higher than formerly assumed may require a re-evaluation of the disposition of the radiation received by the earth in the mean. That the long-term average radiative balance controls the average weather pattern is, of course, not subject to question. The excess or deficit of absorbed solar radiation by comparison with the normal absorbed radiation should also significantly influence the weather elements averaged over relatively short periods. Whether a week, a month, a season, or longer is the required "relatively short period" remains to be investigated.

REFERENCES

An attempt has been made to limit the number of references. Many of those listed here contain additional useful ones. A relatively large number of references are contained in several of the items listed below; these have been marked with an asterisk.

1. Abbot, C. G., "The Variations of the Solar Constant and Their Relation to Weather. Reply to Paranjpe and Brunt." *Quart. J. R. meteor. Soc.*, 65:215–236 (1939).

2. —— Fowle, F. E., and Aldrich, L. B., "The Distribu-

tion of Energy in the Spectra of the Sun and Stars." *Smithson. misc. Coll.*, Vol. 74, No. 7, 30 pp. (1923).

3. —— and Hoover, W. H., *Ann. astrophys. Obs. Smithson. Instn.*, 6 Vols. (1900–1942).
4. Abetti, G., *The Sun*, trans. by A. Zimmerman and F. Bourghouts. London, C. Lockwood and Son, 1938. New York, Van Nostrand, 1938.
5. Adel, A., "Selected Topics in Infrared Spectroscopy of the Solar System." (In [54], pp. 269–283.)
6. Albrecht, F., "Eine einheitliche Darstellung des Absorptions-spektrums von wasserdampfhaltiger Luft und flüssigem Wasser." *Z. Meteor.*, 1:194–196 (1947).
7. Aldrich, L. B., "The Reflecting Power of Clouds." *Smithson. misc. Coll.*, Vol. 69, No. 10, 9 pp. (1919).
8. —— "The Solar Constant and Sunspot Numbers." *Smithson. misc. Coll.*, Vol. 104, No. 12 (1945).
9. Allen, C. W., "Variation of the Sun's Ultra-Violet Radiation as Revealed by Ionospheric and Geomagnetic Observations." *Terr. Magn. atmos. Elect.*, 51:1–18 (1946).
10. Ångström, A., "On the Albedo of Various Surfaces of Ground." *Geogr. Ann., Stockh.*, 7:323–342 (1925).
11. —— "On the Atmospheric Transmission of Sun Radiation. II." *Geagr. Ann., Stockh.*, 12:130–159 (1930). See also "On the Atmospheric Transmission of Sun Radiation and on Dust in the Atmosphere." *Ibid.*, 11:156–166 (1929).
12. —— "Survey of the Activities of the Radiation Commissions of the IMO and of the IUGG." *Tellus*, Vol. 1, No. 3, pp. 65–71 (1949).
13. Appleton, E. V., and Naismith, R., "Normal and Abnormal Region E Ionization." *Proc. phys. Soc. Lond.*, 52:402–415 (1940).
14. Baur, F., und Philipps, H., "Der Wärmehaushalt der Lufthülle der Nordhalbkugel in Januar und Juli und zur Zeit der Äquinoktien und Solstitien." *Beitr. Geophys.*, 42:160–207 (1934).
15. Bernheimer, W. E., "On the Solar Radiation in the Ultraviolet Region of the Spectrum" in Quatrième rapport, *Relations entre les phénomènes solaires et terrestres*. Florence, Conseil Intern. Union Sci., 1936.
16. Brasefield, C. J., "Exploring the Ozonosphere." *Sci. Mon.*, 68:395–399 (1949).
17. Brooks, C. F., "Oceanography and Meteorology" in *Physics of the Earth. V—Oceanography*. Bull. 85, pp. 457–519, Washington, Nat. Res. Council of the Nat. Acad. Sci., 1932.
18. Bullrich, K., "Lichtdurchlässigkeit in Wolken." *Z. Meteor.*, 2:321–325 (1948).
19. Chapman, R. M., Howard, J. N., and Miller, V. A., *Atmospheric Transmission of Infra-Red*. Rep. No. 18, Ohio State Univ. Res. Found., Contract W 44-099-eng-400, Engineer Center, Ft. Belvoir, Va., 1949.
20. Clearman, H. E., "The Ultraviolet Solar Spectrum." (In [54], pp. 125–134.)

*21. Craig, R. A., "The Observations and Photochemistry of Atmospheric Ozone and Their Meteorological Significance." *Meteor. Monogr.*, Vol. 1, No. 2 (1950).

22. Crozier, W. D., and Seely, B. K., *Airborne Particles Collected on a Transcontinental Airplane Flight*. New Mexico School of Mines Tech. Rep. No. 1-NR, ONR Proj. No. NR-082-013, 1949.
23. Danjon, A., "Nouvelles recherches sur la photométrie de la lumière cendrée et l'albedo de la terre." *Ann. Obs. Stras. bourg*, 3:139–180 (1936).
24. Drummeter, L. F., and Strong, J., *Infrared Absorption of Water Vapor at 1.8 Microns*. Johns Hopkins Univ. Rep., ONR Contract N5-ori-166, T.O. V, 1949.
25. Ellison, M. A., "Characteristic Properties of Chromospheric Flares." *Mon. Not. R. astr. Soc.*, 109:1–27 (1949).
26. Fowle, F. E., "The Transparency of Aqueous Vapor." *Astrophys. J.*, 42:394–411 (1915).
27. —— "Water-Vapor Transparency to Low-Temperature Radiation." *Smithson. misc. Coll.*, Vol. 68, No. 8 (1917).
28. —— "The Atmospheric Scattering of Light." *Smithson. misc. Coll.*, Vol. 69. No. 3, 12 pp. (1918).
29. Fritz, S., "The Albedo of the Ground and Atmosphere." *Bull. Amer. meteor. Soc.*, 29:303–312 (1948).
30. —— "The Albedo of the Planet Earth and of Clouds." *J. Meteor.*, 6:277–282 (1949).
31. —— "Solar Radiation during Cloudless Days." *Heat. & Ventilating*, 46:69–74 (1949).
32. —— "Measurements of the Albedo of Clouds." *Bull. Amer. meteor. Soc.*, 31:25–27 (1950).
33. —— and MacDonald, T. H., "Average Solar Radiation in the United States." *Heat. & Ventilating*, 46:61–64 (1949).

*34. Gaertner, H., *The Transmission of Infrared in Cloudy Atmosphere*. NAVORD Rep. 429, 67 pp. Washington, D. C., Gov't Printing Office, 1947.

35. Götz, F. W. P., und Schönmann, E., "Die spektrale Energieverteilung von Himmels- und Sonnenstrahlung." *Helv. phys. Acta*, 21:151–168 (1948).
36. Gowan, E. H., "Ozonosphere Temperatures under Radiation Equilibrium." *Proc. roy. Soc.*, (A) 190:219–226 (1947).
37. Greenstein, J. L., "The Upper Atmosphere Studied from Rockets." (In [54], pp. 112–122.)
38. Grimminger, G., "Analysis of Temperature, Pressure, and Density of the Atmosphere Extending to Extreme Altitudes." *Project Rand*. Santa Monica, Calif., The Rand Corp., 1948.
39. Haeff, A. V., "On the Origin of Solar Radio Noise." *Phys. Rev.*, 75:1546–1551 (1949).
40. Hand, I. F., "Atmospheric Contamination over Boston, Massachusetts." *Bull. Amer. meteor. Soc.*, 30:252–254 (1949).
41. ——"Weekly Mean Values of Daily Total Solar and Sky Radiation." *U. S. Wea. Bur. Tech. Pap.* No. 11, Washington, D. C. (1949). See also "Review of U. S. Weather Bureau Solar Radiation Investigations." *Mon. Wea. Rev. Wash.*, 65:415–441 (1937); and "A Summary of Total Solar and Sky Radiation Measurements in the United States." *Ibid.*, 69:95–125 (1941).
42. Haurwitz, B., "Daytime Radiation at Blue Hill Observatory in 1933 with Application to Turbidity in American Air Masses." *Harv. meteor. Stud.*, No. 1 (1934).
43. —— "Insolation in Relation to Cloud Type." *J. Meteor.*, 5:110–113 (1948).
44. —— "Solar Activity, the Ozone Layer and the Lower Atmosphere" in "Centennial Symposia, Dec. 1946." *Harv. Obs. Monogr.*, No. 7, pp. 353–368 (1948).
45. Hewson, E. W., "The Reflection, Absorption and Transmission of Solar Radiation by Fog and Cloud." *Quart. J. R. meteor. Soc.*, 69:47–62 (1943). See also "Discussion," *ibid.*, pp. 227–234.
46. Houghton, H. G., and Chalker, W. R., "The Scattering Cross Section of Water Drops in Air for Visible Light." *J. opt. Soc. Amer.*, 39:955–957 (1949).
47. Hulburt, E. O., "The Upper Atmosphere of the Earth." *J. opt. Soc. Amer.*, 37:405–415 (1947).
48. Ives, J. E., and Co-workers, "Atmospheric Pollution of American Cities for the Years 1931 to 1933." *Publ. Hlth. Bull. Wash.*, No. 224 (1936).
49. Kalitin, N. N., "Einfluss der Bewölkung auf die Hellig-

keit der Erdoberfläche durch diffuses Licht der Atmosphäre." *Strahlentherapie*, 39:717–728 (1931).

50. Karandikar, R. V., "Radiation Balance of the Lower Stratosphere. Part I—Height Distribution of Solar Energy Absorption in the Atmosphere." *Proc. Ind. Acad. Sci.*, Sect. A, 23:70–96 (1946).
51. Kimball, H. H., "Variations in the Total and Luminous Solar Radiation with Geographical Position in the United States." *Mon. Wea. Rev. Wash.*, 47:769–793 (1919).
52. —— "Solar Radiation and Its Role" in *Physics of the Earth. III—Meteorology*. Nat. Res. Coun., Nat. Acad. Sci., Washington, D. C., 1931. (See pp. 35–66)
53. Klein, W. H., "Calculation of Solar Radiation and the Solar Heat Load on Man." *J. Meteor.*, 5:119–129 (1948).
*54. Kuiper, G. P., ed., *The Atmospheres of the Earth and Planets*. Chicago, University of Chicago Press, 1949.
*55. Linke, F., *Handbuch der Geophysik*, Bd. 8, Lief. 1 u. 2. Berlin, Gebr. Borntraeger, 1942, 1943.
*56. List, R. J., ed., *Smithsonian Metcorological Tables*, 6th ed. Washington, Smithson. Instn. (in press).
57. Luckiesh, M., "Aerial Photometry." *Astrophys. J.*, 49:108–130 (1919).
58. Marvin, C. F., "On the Question of Day-to-Day Fluctuations in the Derived Values of the Solar Constant." *Mon. Wea. Rev. Wash.*, 53:285–303 (1925).
59. Mecke, R., "Die Gesetze der Lichtausbreitung in optisch trüben Medien und das Sichtweiteproblem." *Meteor. Z.*, 61:195–199 (1944).
60. Menzel, D. H., "Earth-Sun Relationships" in "Centennial Symposia, Dec. 1946." *Harv. Obs. Monogr.*, No. 7, pp. 319–329 (1948).
61. Milankovitch, M., *Phénomènes Thermiques*. Paris, Gauthier-Villars et Cie, 1920. (See also "Mathematische Klimalehre und astronomische Theorie der Klimaschwankungen," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Band I, Teil A. Berlin, Gebr. Borntraeger, 1930.)
*62. Mitra, S. K., *The Upper Atmosphere*. Calcutta, The Royal Asiatic Society of Bengal, 1948.
63. Moon, P., "Proposed Standard Solar-Radiation Curves for Engineering Use." *J. Franklin Inst.*, 230:583–617 (1940).
64. National Bureau of Standards, *Ionospheric Radio Propagation*. Nat. Bur. Stand. Circ. 462, Washington, D. C., 1948.
65. Neiburger, M., "The Reflection of Diffuse Radiation by the Sea Surface." *Trans. Amer. geophys. Un.*, 29:647–652 (1948).
66. —— "Reflection, Absorption and Transmission of Insolation by Stratus Clouds." *J. Meteor.*, 6:98–104 (1949).
67. Paranjpe, M. M., "The Variations of the Solar Constant and Their Relation to Weather." *Quart. J. R. meteor. Soc.*, 64:459–476 (1938).
68. Pettit, E., "Measurements of Ultraviolet Solar Radiation." *Astrophys. J.*, 75:185–221 (1932).
69. Phillips, M. L., "The Ionosphere as a Measure of Solar Activity." *Terr. Magn. atmos. Elect.*, 52:321–332 (1947).
70. Ramanathan, K. R., and Karandikar, R. V., "Effect of Dust and Haze on Measurements of Atmospheric Ozone Made with Dobson's Spectrophotometer." *Quart. J. R. meteor. Soc.*, 75:257–267 (1949).
71. Rossby, C.-G., "The Scientific Basis of Modern Meteorology" in *Handbook of Meteorology*, F. A. Berry, E. Bollay, and N. R. Beers, ed., pp. 502–529. New York, McGraw, 1945.
72. Stair, R., "Measurement of Ozone over the Organ Mountains, New Mexico." *J. Res. nat. Bur. Stand.*, 40:9–19 (1948).
73. Sterne, T. E., "On Periodicities in Measures of the Solar Constant." *Proc. nat. Acad. Sci., Wash.*, 25:559–564 (1939).
74. Tanck, H.-J., "Die tägliche Erwärmung der Atmosphäre infolge der Absorption der direkten Sonnenstrahlung durch den atmosphärischen Wasserdampf." *Ann. Hydrogr., Berl.*, 68:47–64 (1940).
75. Teele, R. P., "Insolation, Earth and Sky Brightness" in *Nat. Geogr. Soc.-U. S. Army Air Corps Stratosphere Flight of 1935 in the Balloon "Explorer II."* Nat. Geogr. Soc., Washington, D. C., 1936. (See pp. 133–138)
*76. van de Hulst, H. C., "Scattering in the Atmospheres of the Earth and the Planets." (In [54], pp. 49–111.)
77. Waldmeir, M., *Ergebnisse und Probleme der Sonnenforschung*. Leipzig, Becker & Erler, 1941.
78. Wexler, H., "A Comparison of the Linke and Ångström Measures of Atmospheric Turbidity and Their Application to North American Air Masses." *Trans. Amer. geophys. Un.*, 14th Annual Meeting, pp. 91–99 (1933).
79. Wulf, O. R., and Deming, L. S., "On the Production of Ionospheric Regions E and F and the Lower-Altitude Ionization Causing Radio Fade-outs." *Terr. Magn. atmos. Elect.*, 43:283–298 (1938).

LONG-WAVE RADIATION*

By FRITZ MÖLLER

Gutenberg University at Mainz

Introduction

Long-wave radiation occupies a peculiar position in the science of meteorology in that its effects in the free atmosphere are known only through theoretical calculations and not through measurements. These calculations are, nevertheless, based on experiments in the laboratory and in the atmosphere. The atmospheric radiation is very seldom measured from balloons [3] or from aircraft [17], but quite frequently on the ground where the radiation from above is observed. It appears that part of the measurements from airplanes are in error; and the measurements on the ground have seldom found an evaluation that went beyond the formulation of empirical equations to approximate their average values. The theoretical deductions concerning the radiation in the atmosphere are much more extensive, and it would be most desirable to support them by careful measurements, especially since the experimental tools are available.

Initially, much was expected from the investigations of long-wave radiation in the free atmosphere. It was hoped that they would furnish an explanation for the latitudinal differences in the temperature and altitude of the tropopause, an interpretation of the variations of these values from day to day, and a physical elucidation of the origin of the inversions in the free atmosphere. In all these problems the solutions at times seemed to be near at hand, but then they receded. It appears to the author that today the emphasis of research is directed more toward the investigation of the radiative balance and heat budget in the atmosphere.

In the following, the techniques of measurement are omitted and little will be said concerning the mathematical-physical basis of the calculations, since two detailed treatises which are still up to date are available [14, 28]. They cover these subjects very thoroughly. On the other hand, we shall discuss in great detail the number of instances where long-wave radiation is involved in the problems of general meteorology.

Long-wave radiation is a heat radiation; its energy is derived from the kinetic energy of the molecules. The radiators of this energy are those atmospheric gases that have absorption bands in the temperature radiation range of 4 to approximately 100 μ, that is, water vapor, carbon dioxide, and ozone. The effect of the water vapor is restricted almost exclusively to the troposphere, in which carbon dioxide and ozone are of little significance. The latter two gases become of great importance in the stratosphere between 15 and 35 km. In addition, heat is radiated by the ground and the clouds.

* Translated from the original German.

The Absorption Laws

From the beginning, the theoretical calculations were set up in such general terms that it became possible to investigate the radiation properties of any atmosphere with any temperature distribution and any possible arrangement of radiating and absorbing media. This was done by the development of special radiation diagrams or radiation charts. This approach was justified by two facts: (1) The intensity of the radiation which is emitted by an element for a given wave length is proportional to the black-body radiation and is thus a function only of the absolute temperature; (2) this radiation is proportional to the mass of the radiating medium, which means, in the troposphere, to the amount of water vapor. It is of particular importance that the proportionality constant, that is, the absorption coefficient k, is really a constant in its first approximation and not a function of any other quantities. However, in reality such a dependence of k on the air pressure and on the temperature does exist and gives rise to particular difficulties in advanced investigations.

It should immediately be pointed out that the construction and application of radiation diagrams become impossible if there exist in the atmosphere two different media whose masses in a given volume element vary independently of each other from case to case, but whose emissive power at the same wave length is of the same order of magnitude so that the effect of the one medium cannot be neglected as compared to that of the other. Such conditions prevail in the stratosphere, for instance in the case of ozone and carbon dioxide. It is possible to consider two media in the same space element of the atmosphere only when one of the radiators is a gray radiator, that is, when its absorption coefficient k is the same for all wave lengths [31].

Parallel Radiation. If an absorbing medium m is penetrated by a beam of parallel rays, absorption takes place according to an exponential law. The emergent radiation I_λ is

$$I_\lambda = I_{0\lambda} e^{-k_\lambda m}, \tag{1}$$

where the subscript λ indicates monochromatic radiation, and I_0 is the incident radiation. The absorbed portion is

$$A = (I_0 - I)/I_0 = 1 - e^{-k_\lambda m}. \tag{2}$$

The length of the path which the radiation follows in penetrating m does not appear in this equation. This means that the absorption is only a function of the penetrated mass m, or that the absorption coefficient k_λ is independent of the density of m throughout the penetrated cylinder. If I_0 is black-body radiation at a temperature T, whose value is given by Planck's law,

$$I_0 = \mathfrak{E}(\lambda, T)\, d\omega,$$

(where $d\omega$ is the solid angle), and if the mass m has the same temperature T, then m radiates, according to Kirchhoff's law, the same amount of energy as it absorbs; that is, it emits

$$E = \mathfrak{E}(\lambda, T)\, d\omega \cdot A = \mathfrak{E}(\lambda, T)\, d\omega\, (1 - e^{-k_\lambda m}). \quad (3)$$

A mass element dm of temperature T therefore sends through the absorbing mass m the radiation

$$dE = \mathfrak{E}(\lambda, T)\, d\omega \cdot dA = \mathfrak{E}(\lambda, T)\, d\omega \cdot e^{-k_\lambda m} k_\lambda dm. \quad (4)$$

If the mass element has a different temperature T_1, the radiation is simply

$$\mathfrak{E}(\lambda, T_1)\, d\omega \cdot e^{-k_\lambda m} k_\lambda dm \quad (5)$$

in which it is assumed that k_λ is not a function of the temperature of the mass m.

Diffuse Radiation. The simple basic law represented by equations (3) and (4) holds essentially for all complications that appear in the atmosphere. First of all we must take into consideration the fact that the radiation is not parallel. Let us consider the radiation of a layer of air of infinite horizontal extent which is part of a uniformly stratified air mass and which is to contain the radiating mass dm over an area of one square centimeter. Then the radiation emitted from this mass will arrive on a receiving surface at all angles of incidence $0° \leq \varphi \leq 90°$. The integration of equation (4) over φ can be reduced to known functions and the radiation of the layer element is

$$dS = \pi \mathfrak{E}\, (\lambda, T)\, 2H_2(k_\lambda m)\, k_\lambda dm. \quad (6)$$

The radiation of a layer of finite thickness with a temperature T is then

$$S = \pi \mathfrak{E}\, (\lambda, T)\, [1 - 2H_3(k_\lambda m)]. \quad (7)$$

The functions H_2 and H_3 (called Ei_2 and Ei_3 by Elsasser) are tabulated [27], so that they can be used for exact calculations. The expression $A^d = [1 - 2H_3\ (k_\lambda m)]$ can be considered as an absorption function of diffuse radiation, and its differential is accordingly

$$dA^d = 2H_2(k_\lambda m)\, k_\lambda dm.$$

Within a large range the following approximations can be made:

$$2H_2(x) \approx 1.66e^{-1.66x} \text{ and } 2H_2(x) \approx e^{-1.66x},$$

which means that the laws for diffuse radiation between atmospheric layers are replaced by the laws for parallel radiation in which, however, the absorption coefficient is multiplied by $\frac{5}{3}$.

Radiation from a Spectral Line. A further modification of the simple absorption laws is required by the physical processes that take place when gas masses radiate. The absorption bands of the multimolecular gases are not continuous, but are resolved into numerous closely spaced absorption lines. The absorption coefficient is extremely high in the center of each of these lines, while it is smaller by about two orders of magnitude between two lines. Thus, even if we consider only a very small spectral band $\Delta\lambda$ which contains only a single line, we must take into consideration a variation of k_λ at a ratio of 1:100. This is most easily done by determining once and for all the absorption function of a line. The form of a spectral line, that is, the law of the decrease of the absorption coefficient from the center towards the edges, is known as the dispersion form:

$$k_\nu = \frac{k_0 \delta^2/4}{(\nu - \nu_0)^2 + \delta^2/4}. \quad (8)$$

(Instead of λ, the frequency $\nu = 1/\lambda$ cm^{-1} is used; δ is also given in cm^{-1}.) The significant values in this law are (1) the absorption coefficient in the center k_0, (2) the half-value width δ of the line, and (3) the distance between two adjacent lines $\Delta\nu$. The ratio $\delta/\Delta\nu$ determines to what fraction of k_0 the absorption coefficient k decreases between two lines. Neither δ nor $\Delta\nu$ has the same value in one band, let alone in different bands, of the same spectrum. The distance between lines $\Delta\nu$ varies irregularly, because of the overlapping of differing laws for the various spectral lines. Only very few values have been determined for δ, because the measurements are extremely difficult to make. Therefore, it is usually assumed that δ, as well as $\Delta\nu$, is constant for the entire spectrum. Although this assumption is only an expedient, there is no possibility of a more exact evaluation at the present time.

If we integrate the absorption laws for parallel or diffuse radiation over such a dispersion form of a spectral line, we arrive at new laws, that is, new absorption functions, which can be expressed by

$$L^d(k_0 m) = \frac{1}{\Delta\nu} \int_{-\Delta\nu/2}^{+\Delta\nu/2} \left[1 - 2H_3\left(\frac{k_0 m \delta^2/4}{(\nu - \nu_0)^2 + \delta^2/4}\right)\right] d\nu \quad (9)$$

for the radiating layer m, and by

$$L(\mathrm{k_0 m}) = \frac{1}{\Delta\nu} \int_{-\Delta\nu/2}^{+\Delta\nu/2} \left[1 - \exp\left(-\frac{k_0 m \delta^2/4}{(\nu - \nu_0)^2 + \delta^2/4}\right)\right] d\nu \quad (10)$$

for the radiating column. As before, the radiation of a layer element is the first derivative of this function with respect to m, that is, $\frac{\partial L\,(k_0 m)}{\partial m}\, dm$ for parallel or diffuse radiation. It is of no consequence whether this integral can be solved analytically or numerically. If it can be tabulated, it can be used for any further calculations. The two new equations, (9) and (10), contain as a parameter the ratio between the half-value width and the distance between lines, $\alpha = \Delta\nu/\delta$.

An approximate solution for L may be obtained, if we assume $\delta^2/4$ in the denominator to be negligible compared to $(\nu - \nu_0)^2$. In that case

$$L\,(k_0 m) \approx \sqrt{k_0 \pi m/\alpha^2}.$$

This indicates that the radiation of a gas layer of finite thickness m is proportional to the square root of m if the absorption occurs in individual lines. If, on the other hand, we were dealing with continuous absorption, we would arrive at an exponential function. Thus,

simply by plotting experimental absorption values against the square roots of m, or by the usual representation of ln $(1 - A)$, we can determine whether we are dealing with a continuous distribution of the absorption coefficient, or with a resolution into individual lines. It is important to realize that this method is applicable even if the apparatus is not sufficiently sensitive to resolve the individual lines. Strong [48] applied this method to entire bands whose separate lines may have very different values for k_0.[1]

Radiation Diagrams

The knowledge of the absorption of a spectral line gives us one basis for the calculation of atmospheric heat radiation. The second basis is the distribution of the values for k_0 over the various wave lengths. For water vapor, which is most effective in the troposphere, the available measurements have been tabulated by Elsasser [14] and Möller [34]. Callendar and Cwilong give analogous figures. The absorption shows very great differences. In the rotation spectrum, k_0 is of the order of 10^4 cm^2 g^{-1}; in the rotation-oscillation spectrum at 6 μ it is about 10^3 cm^2 g^{-1}; whereas it decreases in the "window" of the water vapor to 1 cm^2 g^{-1} or below. In spite of these large differences we can never neglect one spectral range as compared to another, because the intensive radiation in the range of large values of k_0 is readily absorbed even by very thin layers, whereas the low radiation intensities at small values of k_0 are scarcely absorbed. However, thick and distant layers can participate in the emission of this radiation. Therefore we need the cumulative effect of all wave lengths for comparison with measurements. Mügge and Möller [37] were the first to use a graphical method in which the integration over all wave lengths is computed in advance and represented in diagrammatic form. Elsasser [14] developed a similar diagram, which is the same in principle, but which differs somewhat in arrangement.

According to the foregoing discussion, the radiation at wave length λ of a thin layer of gas of temperature T is given by

$$dS_\lambda = \pi \mathfrak{E}(\lambda, T) \frac{\partial L^d(k_0 m)}{\partial m} dm, \qquad (11)$$

where k_0 varies from line to line. Even larger spectral ranges that comprise a number of lines can be combined as long as k_0 varies less than k_ν in the range of one line, that is, less than 100:1. Then the total radiation of a layer element dm is obtained by summation over all wave lengths,

$$dS = \pi dm \sum_\lambda \mathfrak{E}(\lambda, T) \frac{\partial L^d(k_0 m)}{\partial m} \Delta\lambda, \qquad (12)$$

and the radiation of a layer m of finite thickness upon a surface element situated in the one boundary surface of the layer m becomes

$$S = \pi \int_0^m dm \sum_\lambda \mathfrak{E}(\lambda, T) \frac{\partial L^d(k_0 m)}{\partial m} \Delta\lambda. \qquad (13)$$

This is the radiation of an atmospheric layer upon a unit area, for instance the radiation of the entire atmosphere upon the unit area of the sensitive surface of a measuring instrument placed on the ground. The integration over m is difficult at first, because the temperature T is in general not constant but a function of m, that is, a function of the radiating mass situated between an altitude above ground and the surface of the earth. If we consider an isothermal atmosphere of temperature T_0, then

$$S_{T_0} = \pi \int_0^m dm \sum_\lambda \mathfrak{E}(\lambda, T_0) \frac{\partial L^d(k_0 m)}{\partial m} \Delta\lambda = X_{T_0}(m) \qquad (13a)$$

will be a function X of m. If we now plot the absorption

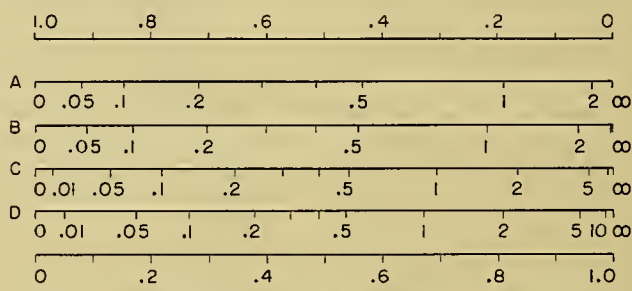


FIG. 1.—Absorption functions. The linear scale at the top gives the transmitted radiation. The linear scale at the bottom gives the absorptive or emissive power. Numbers on the function scales from A to D are the absorbing mass m.
(A) $1 - e^{-km}$ for $k = 1.66$. Parallel radiation.
(B) $1 - 2\,H_3(m)$. Diffuse radiation.
(C) $L^d_{\alpha=5.5}(k_0 m)$ for $k_0 = 6.5$. Diffuse radiation of a spectral line with $\alpha = 5.5$.
(D) $L^d_{\alpha=12}(k_0 m)$ for $k_0 = 20$. Diffuse radiation of a spectral line with $\alpha = 12$.
The values of k and k_0 are chosen so that for the same mass m the absorption will be 0.5 in each case.

function X as the abscissa and provide it with a scale of m, as in Fig. 1, we can read at the scale division m, the radiation intensity emitted by the isothermal layer of temperature T_0. However, X also indicates the amount absorbed by this layer when an infinite surface of temperature T_0 transmits black-body radiation through this layer. If k_0 does not vanish for any wave length, an infinitely thick layer ($m = \infty$) will have total absorption. Accordingly, the radiation emitted by an infinitely thick layer of gas is equal to the black-body radiation: the point $m = \infty$ of the scale lies at $X = \sigma T_0^4$.

It can be seen immediately that the radiation of a layer element dm of temperature T_0 is also given by the differential of X:

$$dS = \frac{\partial}{\partial m} X_{T_0}(m)\, dm.$$

In order to find the radiation of a layer element of a temperature other than T_0, a new evaluation of the

1. (Note added July, 1950.) Callendar [11] used an empirical absorption law $L(w) = w/(w + w_0)$ in which w_0 is a constant characterizing the degree of absorption; this law is easily applicable in theoretical investigations and covers the observations well. Using all experimental data of water vapor, Cwilong [12] recently deduced an empirical absorption function, but he gives numerical values only and not an analytic expression. The values of the function are available for narrow frequency intervals of the whole long-wave water-vapor spectrum.

summation over λ in (12) or (13) with T as a parameter becomes necessary. Let the ratio of the radiation of the two layer elements be y; then

$$y(T, T_0, m) = \frac{\sum_\lambda \mathfrak{E}(\lambda, T) \frac{\partial}{\partial m} L^d(k_0 m)\Delta\lambda}{\sum_\lambda \mathfrak{E}(\lambda, T_0) \frac{\partial}{\partial m} L^d(k_0 m)\Delta\lambda}. \quad (14)$$

We are now able to determine the radiation of the layer element dm of temperature T from the product

$$dS_T = y\,(T, T_0, m)\frac{\partial}{\partial m} X_{T_0}(m)\, dm. \quad (15)$$

From (15) follows the key equation of the radiation diagram

$$S = \int_0^m y(T, T_0, m)\, dX_{T_0}(m), \quad (16)$$

where T may now be any function of m, that may be given by observation or assumption. If we now plot y on the ordinate against X on the abscissa, we obtain a graph with curves for every value of T, in which $y = 1$ signifies that $T = T_0$, and $T < T_0$ gives values of y less than 1 (Fig. 2*a*). The radiation of an isothermal

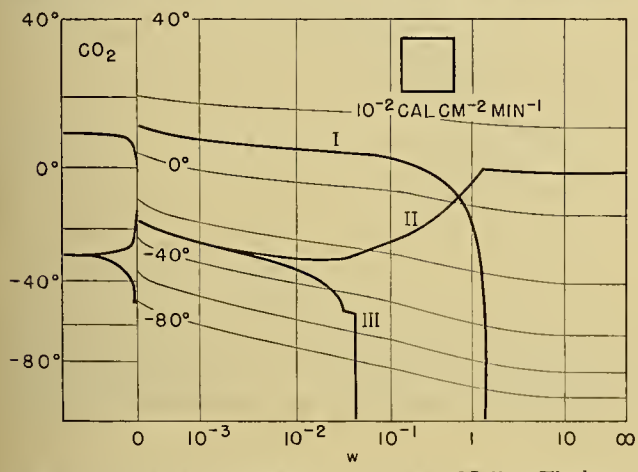


Fig. 2*a*.—Radiation diagram according to Möller. The heavy lines refer to downcoming radiation of the atmosphere at the ground (I), the radiation at 7 km received from below (II), and the radiation at 7 km received from above (III).

atmospheric layer is then given by the area bounded laterally by parallels to the y-axis through $m = 0$ and $m = m$, by the X-axis at the bottom, and the line T at the top. The radiation of a nonisotherma atmospheric layer in which T is a function of m can be found by plotting the temperature distribution $T(m)$ in the network of curves for m and T, and integrating. This is the basic principle of all radiation diagrams. Aside from the use of different absorption values, k_0, Elsasser's chart is an authalic transformation of this principle, in which the isotherms are made rectilinear, and, as a result, the lines of equal m-values become curved. The curvature is hardly noticeable, because the lines for $m = 0$ and $m = \infty$, which, as the boundaries of the graph, remain straight lines, intersect at the point $T = 0$. Thus the diagram assumes triangular or trapezoidal shape (Fig. 2*b*). Furthermore, abscissa and ordinate are interchanged by comparison with Möller's diagram.

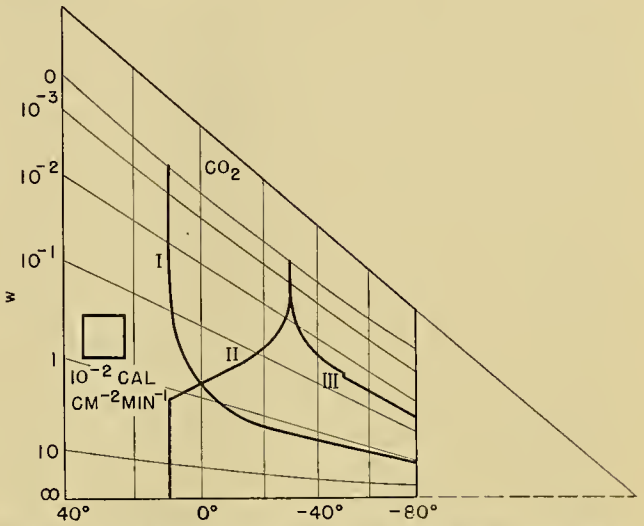


Fig. 2*b*.—Radiation chart according to Elsasser. The heavy lines correspond to those in Fig. 2*a*.

Even though the principle of the two radiation diagrams is the same, there are certain numerical differences. These are best illustrated by a comparison of the functions X_{40C} and X_{-80C} in the two charts. The values $T = +40C$ and $-80C$ are the highest and lowest temperatures shown. Table I shows that there

Table I. Radiation of an Isothermal Layer with a Water-Vapor Content w in Per Cent of σT^4 as Given by Elsasser (E) and Möller (M)

Temperature		w (g cm^{-2})						
		0 (CO_2 only)	10^{-3}	10^{-2}	10^{-1}	1	10	∞
		%	%	%	%	%	%	%
+40C	E	17.0	26.8	41.1	57.8	75.9	91.9	100.0
	M	13.3	24.4	41.5	58.3	72.7	89.0	100.0
−80C	E	16.6	38.1	58.0	76.1	88.8	96.6	100.0
	M	13.1	32.7	56.4	75.0	85.2	93.3	100.0

is good agreement with differences of less than 1 per cent in the middle range of amounts of water vapor between 3×10^{-3} and 3×10^{-1} g cm^{-2}. In the range of smaller or larger amounts of water vapor Möller's values are about 2 per cent lower than Elsasser's. (The first edition of Elsasser's chart, as well as the earlier edition of the chart by Mügge and Möller both showed radiation values 10 to 15 per cent lower in the middle range of water-vapor amounts. The agreement of the revisions made by both authors independently during the war is rather remarkable.)

There is more of a difference in the evaluation by the two authors of the radiation of carbon dioxide. For the amount of CO_2 normally present in the atmosphere, there is almost complete absorption in the extraordinarily intense band around 14.9 μ even by only very thin layers. The weak extreme boundaries of the band extend to 12.5 μ and 17.5 μ, respectively. Elsasser now

assumes that in the region from 13.1 μ to 16.9 μ the CO_2 always absorbs so strongly that we can assume total absorption, and that therefore the radiation emitted in that range by an atmospheric layer is equal to the black-body radiation at a temperature existing at the surface of such a layer. Möller [35] assumes a width of only 3 μ for this range, that is, from 13.5 to 16.5 μ, and estimates that within these wave lengths the absorption by water vapor is equal to that by CO_2 only when the specific humidity is 100 g kg^{-1} or more. Outside of these boundaries, however, the absorption by H_2O is greater than that by CO_2. Accordingly, the CO_2 absorption at 273K is 18.4 per cent according to Elsasser and 14.6 per cent according to Möller. In addition, Möller gives a scale in his chart which permits determination of the radiation effect of CO_2 for large temperature variations at low altitude (close to the ground) or for very low CO_2 content of the air (stratosphere).

Another point of comparison consists of the numerical values assigned to the absorption coefficient k_0 by Elsasser and by Möller. In the 6-μ band Möller's values are somewhat higher. The same applies to the absorption increase from 10 μ to the rotation band, while in the core of this band the values are lower. Elsasser first calculated his graph with the coefficients of his table. Later, however, he corrected it according to measurements of total absorption and concluded from these measurements that the k_0 values around 6 μ were originally too low, while those for 50 μ were too high. This makes the agreement of the k_0 values even better than a comparison of the numerical values would indicate.

Lately, an important objection has been raised against both radiation diagrams. The absorption functions for a spectral line, $L^d(k_0 m)$, which are used by both authors, contain the assumption that $\alpha = \Delta\nu/\delta = 5.6$, wherein δ was set at 0.5 cm^{-1} and the distance between lines, $\Delta\nu$, was assumed to be 2.8 cm^{-1} as the mean of the range investigated by Randall and his collaborators. According to recent measurements by Adel [1] on two lines near 16 and 18.6 μ, it was found that $\delta = 0.23$ cm^{-1}; from this it follows that $\alpha = 12$. The absorption function for a line L^d is also shown in Fig. 1 for $\alpha = 12$. The author suspects that the application of the new function will not lead to any important differences from the previous radiation charts, since the differences of k_0 at 10 μ, as compared to those at 6 μ and 50 μ, are so large as to render all refinements negligible.

The Downcoming Radiation of the Atmosphere

The simplest possible application of the radiation charts arises in the calculation of the downcoming radiation R of the atmosphere and of the effective nocturnal radiation $E = \sigma T_0^4 - R$ of the ground. Only a few direct comparisons of the measured and the calculated values are available.

Wexler [50] compared measurements made in Alaska and North America under winter conditions with radiation values calculated from Elsasser's diagrams and found that on the average the calculated outgoing radiation values were about 0.035 cal cm^{-2} min^{-1} higher than the observed values. This deviation, for which Wexler has no explanation, must probably be ascribed to the use of the earlier edition of Elsasser's diagram which, in the range of water-vapor content in question, furnishes a value for downcoming radiation approximately 10 per cent lower than the later edition of this chart.

F. A. Brooks [7] and Robinson [43] carried out comparative calculations for some cases of their numerous observations, but used them essentially to compute a radiation diagram on an empirical basis (see p. 40). However, in part of these observations in the free atmosphere only the total water-vapor content was used. So far, in most cases, no aerological measurements were made concurrently with the radiation measurements. Therefore only the variables measured on the ground such as pressure, temperature, vapor pressure, and cloudiness were used to organize the measurements and to develop interpolation formulas. The best known are those by Ångström and Brunt which give the ratio $R/\sigma T_0^4$ as a function of the vapor pressure at ground level only where T_0 is the air temperature at the point of observation. Numerous authors have derived the constants of this formula from their measurements, but the scattering of the constants given by the different authors is as great as the scattering of the individual measurements around the curves plotted by each author [28]. Only recently Bolz and Falckenberg [6] gave constants for Ångström's formula which result in values for the downcoming atmospheric radiation which are 7 per cent higher than the constants so far assumed as best values (Fig. 3).

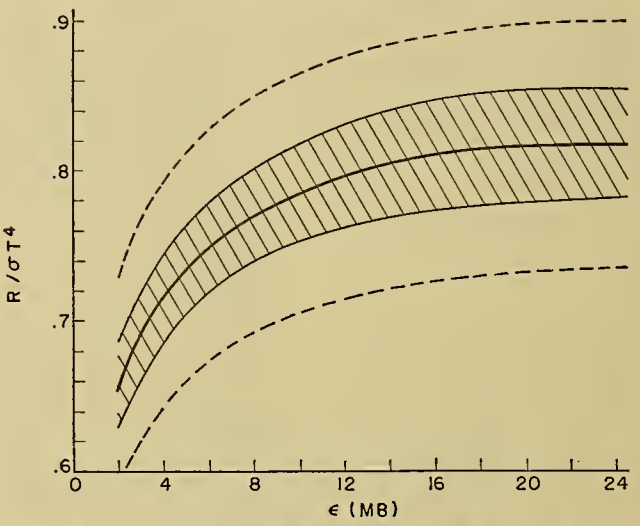


Fig. 3.—Scattering of the relative atmospheric radiation $R/\sigma T^4$ according to Bolz and Falckenberg [6]. Seventy per cent of all measured values lie in the hatched region, and ninety-eight per cent of all values lie within the region bounded by the dashed line.

If we assume a relative humidity that does not vary with height and a normal lapse rate of 6C km^{-1}, we obtain the values for R (at $T_0 = 283$K) given in Table II, according to Möller. These figures are higher than

comparative values [28] that can be calculated according to

$$R_A = \sigma T_0^4 (0.79 - 0.174 \times 10^{-0.041 e_0}) \quad \text{(Ångström)}$$

$$R_B = \sigma T_0^4 (0.48 + 0.60 \sqrt{e_0}). \quad \text{(Brunt)}$$

Three influences may cause the discrepancies and also the large scattering of individual measurements around the interpolation formulas. They are (1) the consideration of temperature, which is all too inaccurate, (2) the neglect of ground inversions, and (3) the additional effect of absorbers other than H_2O and CO_2. These three influences will now be considered in turn.

1. The formulas of Angström and Brunt give the radiation of the atmosphere as proportional to T_0^4. This law is taken into account in Möller's diagram by the fact that the area under a given isotherm T, which represents the radiation of an isothermal atmosphere with $w = \infty$ and CO_2-content $= \infty$, is equal to the black-body radiation σT^4. An isothermal atmosphere of 0C with the limited vapor content 3 g cm^{-2} has a radiative power $R = 0.826\ \sigma T_0^4$; a layer of equal content at +40C gives $R/\sigma T_0^4 = 0.804$ and at −40C, $R/\sigma T_0^4 = 0.866$. The variation of this factor, frequently

Table II. Calculated Downcoming Radiation R and Effective Terrestrial Radiation E
(For T_0 = +10C, ground-level vapor pressure e_0, and total water-vapor content W)

e_0 (mb)	1.23	3.7	6.1	8.6	17.3	20.0
W (g cm^{-2})	0.224	0.67	1.12	1.56	2.24	3.64
R (cal cm^{-2} min^{-1})	0.334	0.369	0.385	0.396	0.408	0.425
E (cal cm^{-2} min^{-1})	0.196	0.161	0.145	0.134	0.122	0.105
$E/\sigma T_0^4$ (per cent)	37	30	27	25	23	20

called emissivity, is caused by the changes of the ordinates of the T-lines in the diagram with w. The variation may also be expressed by the assumption that the radiation of such a limited vapor mass increases more slowly with T than with T^4. Accordingly, it appears that we can set

$$R/\sigma T_0^4 = C\,[1 - \gamma(T_0 - 273)],$$

in which T_0 = observed temperature near the earth's surface, and

$$C = (R/\sigma T_0^4)_{T_0 = 273}$$

is a function of the water content W of the atmosphere.

W g cm^{-2}	0.2	0.5	1	2	3	4
$\gamma\ 10^{-4}$ deg^{-1}	−2.0	1.0	4.0	7.0	8.5	9.5

It can be seen that measurements at high atmospheric temperatures, when they are taken as representative for 0C, yield too low a downcoming radiation, except when the vapor content of the atmosphere is exceptionally low. Brunt prefers presentation by an exponential law and finds the radiation as proportional to $T^{3.5}$ without indication of the vapor content. In the little-known formula of Robitzsch [45] for atmospheric radiation, R varies as σT_0^3.

2. On nearly all cloudless nights during which radiation measurements are made there exists a ground inversion, the magnitude and temperature of which are usually unknown. However, the layers next to the ground also furnish a considerable portion of radiation from above. (The values in parentheses in the following statement are percentages of the black-body radiation.) According to an estimate by Möller [28] for a normal atmosphere, 37 (28) per cent of the radiation proceeds from the layer 0–10 m, 71 (53) per cent from 0–100 m, and 88 (66) per cent from 0–500 m. Accordingly the effect of a ground inversion is great; an attempt at estimating this effect is shown in Table III. The downcoming atmospheric radiation R and the effective terrestrial radiation E are given for an atmosphere of T_0 = 273K and W = 1.12 g cm^{-2}.

This table indicates that a ground inversion can reduce the effective terrestrial radiation to $\frac{4}{5}$ its normal value, and under the extreme conditions found by Mosby during the polar night, to as low as $\frac{1}{2}$. Therefore, any comparison between calculation and observation becomes impossible if the vertical temperature

Table III. Downcoming Atmospheric Radiation R and Effective Terrestrial Radiation E in Relation to Vertical Temperature Distribution

Inversion thickness (m)	Temperature increase (deg C)	R (cal cm^{-2} min^{-1})	E (cal cm^{-2} min^{-1})
100	9.4	.360	.099
200	8.8	.356	.103
300	8.2	.353	.106
400	7.6	.351	.108
500	7.0	.350	.109
1000	0.0	.344	.115
Lapse rate	6C km^{-1}	.334	.125
	adiab.	.333	.126

distribution, especially that part close to the ground, is not known. Frequently, the temperature immediately contiguous to the ground will increase with altitude even more rapidly than assumed in the examples in Table III, and can thus cause an even larger positive deviation of the atmospheric radiation. Generally there will be no inversion over mountain stations, although so-called mountain inversions do occur. The diurnal variation of the temperature gradient may also explain the diurnal variation of the atmospheric radiation and its dependence on the air mass as demonstrated by Falckenberg [16].

3. Robinson [43] has carried out very careful evaluations of the measurements at Kew which included soundings of the free atmosphere. He found that on only few of the clear nights could the radiation values easily be fitted on a smooth curve that represented the relationship between radiation and the vapor content of the atmosphere. For other nights R rose to 0.03 cal, that is, more than 10 per cent higher. Such supplementary radiation appeared at times within an hour. It is impossible to seek an explanation in the variation of the content of CO_2 or O_3. It is rather more plausible to conceive a sudden development or advection of ground inversions or of very thin invisible cloud veils. Robinson suspects, rather, an additional radiation

caused by smoke or combustion gases from the chimneys of London. Quite similarly, Falckenberg [16] tried to explain the deviations he found for different air masses not by temperature gradients, but by the haze content which is not indicated by the water vapor. Volz [49] attempted to measure the emissivity of various substances in the laboratory.

Robitzsch [45] pointed out another as yet unexplained relationship. He established the formula

$$R = \sigma T_0^3 (0.135p + 6.0\epsilon_0)$$

which includes not only the vapor pressure ϵ_0 but also the air pressure p. This formula permits, in particular, the use in the interpolation equations of measurements on mountains or in the free atmosphere. Whether the term containing the air pressure can be attributed to an effect of the CO_2 content or to a diminution of ground inversion with decreasing values of p (anticyclonic and cyclonic conditions) would have to be determined by future research.

F. A. Brooks [7] and Robinson [43] developed radiation diagrams based upon observational data exclusively. The above-mentioned law for monochromatic radiation, according to which the radiation of an atmospheric layer is equal to that of a cylindrical column of air with $\frac{5}{3}$ times the water-vapor content, applies also to the "chromatic" radiation of the natural CO_2 plus H_2O atmosphere. Therefore, a radiation diagram for diffuse radiation can be used also for the investigation of linear or parallel radiation arriving from a definite zenith distance if we multiply the vapor scale by 1.66. The authors cited above proceeded conversely and developed from the measurements of a zonal sky radiation a graph that was then adapted to the use of hemispheric radiation.[2]

Atmospheric radiation finds an important application in the theoretical investigation of the nocturnal cooling process and in the prediction of frost. It has been shown by various authors that the nocturnal cooling cannot be traced essentially to a heat emission of the air by radiation (see the objections raised to this on page 45). The decisive factor is the heat loss from the ground by its effective radiation, and the distribution of this heat loss through conduction into the ground and through convection into the air. In the theoretical treatment of this problem the effective terrestrial radiation $E = \sigma T_0^4 - R$ was often assumed to be constant. However, σT_0^4 decreases with progressive cooling, and R decreases with the development of a ground inversion, but somewhat more slowly. Groen [21, 22] pointed out the necessity as well as the possibility of considering the change of E with T_0 by means of the radiation diagram in such a way that the equations for the nocturnal cooling continue soluble. His final equation represents an important step forward in the theory of nocturnal cooling, especially since he can include in his equation the different disturbing influences such as initial temperature distribution, condensation, and the influence of the wind on turbulent heat exchange.

The Absorption Coefficient as a Function of Pressure

So far we assumed the absorption coefficient k_ν to be independent of pressure and temperature. However, that is not exactly the case. True, the variation is so small that the downcoming atmospheric radiation at the ground is not materially changed, because 88 per cent of it originates in the lower 500 m of the atmosphere where the pressure differs little from that at ground level. At higher altitudes, however, the variation is more effective. The individual absorption line in a band spectrum increases its half-value width with increasing air pressure and temperature. Yet the "total intensity" of the line, that is, $\int_{\Delta_\nu} k_\nu d\nu$, is not changed. Only the shape of the line is altered. With increasing pressure it becomes wider and less intensive, with decreasing pressure it becomes narrower and more intensive. This means that, as the pressure decreases, the absorption increases even more in the narrow center of the line with a large value of k_ν, whereas k_ν becomes even smaller on the wings of the line [14]. As to the absorption by a line, it is found that, in thin layers, the radiation is absorbed almost completely in the core, whereas only in the case of thicker layers is the radiation also absorbed in the wings. With decreased pressure, the absorption in the core is increased and is very effective over the first part of the optical path length. However, subsequent absorption over the remainder of the path is diminished because of the small amount of radiant energy left to be absorbed. Since the initial segment can hardly be observed, the measurements indicate only that the mean absorption coefficient is reduced [46] in proportion to $\sqrt{p/p_0}$. In the absorption formulas k appears only in the product $k \cdot m$. Therefore, it is customary to apply the factor $\mu = \sqrt{p/p_0}$ not to k, but to the radiating mass m or to the densities of water vapor and CO_2. This correction is very important for all investigations into the radiation phenomena of the free atmosphere.

The intensity S of a line, $S = \int k_\nu d\nu$, remains unchanged with a change of the line width δ because of

2. (Note added July, 1950.) Robinson [44] recently published a detailed test of his diagram and of the Elsasser chart, for which he used numerous radiation measurements made at Kew. He found considerable differences which, to a great extent, are caused by the change in the emissivity of a vapor layer with temperature. According to his measurements, the emissivity increases with increasing temperature, whereas according to calculations by means of the Elsasser chart, it decreases by an amount half that of the measured increase (see numerical data on page 39). The prerequisites for the explanation of this discrepancy are as follows: (1) new experimental investigations are needed which would furnish the variation of absorption by vapor layers of finite thickness; (2) the theoretical computations must be checked; the change in radiation with temperature is derived (*a*) from the displacement according to Planck's radiation law, (*b*) from the change in the width δ of a spectral line with $\sqrt{T}$, and (*c*) from the change in the line intensity S with temperature. It appears that the last two influences have, to date, not been sufficiently considered.

the simultaneous change in k_0. Hence

$$S\Delta\nu = k_0\pi\delta.$$

If this expression is inserted into the square-root formula (page 35), we obtain

$$L = \sqrt{Sm\delta/\Delta\nu}.$$

The theory of line broadening by atomic collisions demands a simple proportionality of δ with p. Consequently, the absorption L must vary as $\sqrt{p}$. Schnaidt [46] emphasizes that the measurements by Falckenberg show a proportionality of the absorption *coefficient* k with $\sqrt{p}$. This means that, when an exponential law is used for *absorption*, $\sqrt{p}$ appears as a factor of m in the exponent. However, if the square-root law is used, the logical result is that $L \propto \sqrt{m\sqrt{p}}$ or $L \propto \sqrt[4]{p}$. Thus, there exists a contradiction between theory and observation. Nowadays the tendency is to place more confidence in theory than in measurements. Nevertheless, a repetition of the measurements would be desirable.

At the various wave lengths the transition from the absorption in the center to that on the flanks occurs with entirely differing thicknesses of vapor strata, because of the extraordinarily great differences of the absorption coefficients in the water-vapor spectrum. This leads to a complicated interspersion of absorption amplification and attenuation by the air pressure at the same level of the atmosphere. However, Möller [35] has shown by a rough calculation that, even on the assumption that $\delta \propto p$, the factor that must be applied to the vapor density of the atmosphere for use of the standard absorption equations (9), (13), and (15) has a value within the limits of p and $\sqrt{p}$ up to 100 mb (16 km). At still greater altitudes this factor increases again, because at the extremely low vapor content of the stratosphere only the center of the very strongest lines absorb. Möller proposes, instead of the factor $\mu = p/p_0$, a more complicated one, namely

$$\mu' = 0.985\ (p/p_0)^{0.8} + 0.015\ (p/p_0)^{-1}.$$

This factor has a minimum at around 100 mb and increases at higher levels. However, in the practical calculation of the cooling it is found that at these altitudes the radiation effect of water vapor becomes negligible owing to the greatly diminished vapor content, and that the radiation of other absorbers predominates. Therefore, the rigorous application of the correction factor μ' is unnecessary, as long as there are no better observations available for altitudes above 100 mb which would require more accurate calculations. Nevertheless, calculations to determine the effect of air pressure on changes of the shape of the lines have not been in vain; for, during the past few years, critics, on the basis of the necessary simplifications regarding this effect, questioned repeatedly the validity of calculations by means of the radiation diagrams [39].

Outgoing Atmospheric Radiation

By the use of the correction factor μ for the effect of the air pressure, the radiation diagrams become suitable for investigations of the free atmosphere. For a given level z_1, we can calculate the downward radiation R_1 from the atmospheric layers above it and, correspondingly, the upward radiation U_1 from the ground and from the atmospheric layers below this level. In the radiation diagram the ground is treated as an isothermal gas stratum having the temperature T_0 of the ground and an infinitely great content of water vapor and CO_2. The difference $E_1 = U_1 - R_1$ is then the net radiation which penetrates the reference level in an upward direction. The same calculation for another level z_2 furnishes E_2. The excess radiation $E_2 - E_1$, emitted by the air column $\Delta z = z_2 - z_1$ with air density ρ, causes a cooling

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho c_p}\frac{E_2 - E_1}{z_2 - z_1},$$

which becomes for the limiting case:

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho c_p}\frac{\partial E}{\partial z}. \tag{16}$$

Roberts' first studies [42] of the radiation flow E indicated that it increases with altitude. Normally this increase is very uniform with altitude in a cloudless atmosphere having a continuous vertical distribution of temperature and water vapor. However, the air density decreases with altitude so that the cooling rate which amounts to about 1C per day near the ground increases to two or three times that amount higher up in the troposphere. Only close to the ground and near the tropopause do special conditions prevail (see below).

It seems logical to interpret the cooling of the free atmosphere as a radiation into space. That, however, is not possible, for the shielding by the layers of water vapor above it is too great. It is, rather, a process similar to *heat conduction*. Basically, radiation, just as heat conduction, tends to equalize temperature differences. Therefore, we may also try to set up for radiation an equation that has a form similar to that for heat conduction, namely

$$\partial T/\partial t = K\ \partial^2 T/\partial w^2, \tag{17}$$

where K is a "virtual coefficient of conduction of the heat radiation," w is the mass of water vapor

$$w = \int \rho_w dz, \tag{18}$$

and ρ_w the density of the water vapor. This possibility was long ago developed theoretically by Falckenberg and Stoecker [18], and was later used as the basis of practical estimates by Brunt [9, 10]; here we shall use it only for an interpretation of the cooling. Substitution of equation (18) into (17) gives

$$\partial T/\partial t = K\gamma\rho_w^{-3}\,\partial\rho_w/\partial z, \qquad (\gamma = -\partial T/\partial z)$$

which is negative, because $\partial\rho_w/\partial z < 0$. In other words: The vapor masses at an equal geometrical distance above and below a given altitude are, to be sure, colder or warmer by the same temperature difference; but the

colder vapor masses are "nearer" than the warmer ones in terms of radiation, because the vapor density is lower above that altitude than below it. Therefore, more heat is emitted upward than is received from below. Thus, the cooling in the free atmosphere exists by virtue of the fact that T is not proportional to w.

This behavior of radiation, which is similar to heat conduction, is also clearly revealed by a break in the curve of vertical temperature distribution. There is an abrupt transition (schematically) from $\partial T/\partial z = -\gamma$ to $\partial T/\partial z = 0$ at the tropopause. Hence the vapor particle at this point receives radiant heat from the mass below, but cannot emit anything to the masses at equal temperature above. Therefore, in the absence of other influences, it should become warmer.

However, at higher altitudes the cooling does not depend only on a process similar to heat conduction, but in this case true emission occurs, that is, *heat is radiated to space*. Therefore, the amount of cooling is only partially due to the vertical temperature gradient and for the remainder to the mass of water vapor above and its screening effect on any heat radiation to space. The smaller this mass is, the greater the cooling. It was formerly assumed that the stratosphere had a high vapor content, or that the specific humidity remained constant with altitude.[3] This assumption leads to vapor contents that are too large and to contradictions between the magnitude of outgoing radiation and the actual temperature distribution. Today it is known from measurements by Regener [41] and by Dobson and others [13] that the relative humidity decreases sharply just above the tropopause. A more recent publication by Barrett and collaborators [5] also confirms these results. They found a decrease in humidity from about 10 per cent at the tropopause to about one per cent at 30-km altitude, but with a thin saturated layer interposed. Therefore, these altitudes are already close to the upper boundary of the "water-vapor sphere" and the radiation of the intensive absorption bands proceeds to space almost completely unscreened. Hence, the maximum of cooling lies at altitudes between 8 and 10 km. A calculation based on Elsasser's chart would shift this emitting layer to a somewhat lower level.

Probably, as a third factor, the *radiation by haze* at the tropopause must be considered. The troposphere is always filled with haze, whereas the stratosphere, contiguous to this hazy stratum, contains very dry and extremely clear air (as has been confirmed by numerous observations from aircraft). To be sure, the nature of this haze is not definitely known; but, whether minute droplets or solid particles constitute this haze, both are capable of emitting thermal radiation, even in the range from 9 to 12 μ, where the efficient "window" for outgoing radiation exists. Thus the haze boundary at the tropopause causes an additional cooling which may reach several degrees per day. Similar effects appear also at haze boundaries within the troposphere [31].

3. In the troposphere the decrease of the specific humidity with altitude indicates that vapor is lost and liquefied through cloud formation in the ascending currents of water vapor.

Cloud Radiation

The great effect of clouds on atmospheric radiation is also based on the fact that they radiate like black bodies in the wave-length range of heat radiation. Therefore, the upper cloud surfaces emit a very great quantity of heat in the range from 9 to 12 μ at every altitude of the atmosphere where they may occur; in these intervals there is almost no downcoming radiation from above. This produces an intensive heat loss, concentrated in a very thin layer. Naturally, this heat loss can become effective only if it is not counteracted by another process—as, conceivably, by an approximately equal absorption of radiant solar heat. However, 50 to 70 per cent of the solar radiation is reflected (Fritz [19]), and of the remainder only a very small portion is absorbed, whereas the greater portion traverses the clouds as a diffuse radiative flux and reaches the earth as scattered sky radiation (daylight). There is almost no absorption of solar radiation in the clouds and thus, the heat emission from cloud surfaces is not compensated by the solar radiation, *but acts unimpeded as a heat sink in the atmosphere*. Indeed, the higher such a cloud surface lies, the lower is the black-body radiation corresponding to its temperature. However, the atmospheric radiation which impinges on the cloud surface from above is diminished even more, for it decreases not only with temperature but also with the vapor content of the air situated above. Thus, the effective emission of the cloud increases with altitude.

A different process takes place at the lower boundary of the cloud. This surface receives from the underlying atmosphere, which is generally warmer, and from the ground, which is likewise warmer, a quantity of radiation that is greater than the black-body radiation emitted by the cloud in a downward direction. For this reason, the under surface of the cloud is heated by radiation from below. In this case there is also no compensation by other processes. The heating increases with increased altitude of the cloud, because of the increasing temperature difference between cloud and ground.

For a very thin cloud layer whose vertical extent must not exceed 100 m, both processes, the heat loss above and the heat gain below, can be considered together. Usually, the former is dominant, especially with low clouds such as stratus, and with middle clouds such as altostratus. If we assume that the heat budget of a thin cloud at an altitude of 5 km is distributed by turbulence and similar processes over a layer of air 1 km thick, we find a cooling of this mass of about 5C per day. A high cloud in the upper troposphere does not cool the air, because the radiation from below is relatively greater. In a tropical atmosphere, however, the conditions are quite different. The temperature difference between ground and cloud increase continuously up to about 18 km because of the normal decrease of temperature with height in the troposphere. Even if the assumption of a closed cloud cover is discarded and a scattered cloud cover of only $\frac{1}{10}$ is assumed, the heat balance of the cloud becomes positive at about 14 km. This means that even a thin cirrus cloud at this altitude

receives more heat than it emits. Thus it cannot exist as a cloud and must evaporate. This is probably the reason for the phenomenon observed in the tropics, namely, that the highest cirrus clouds are not found near the stratosphere, but at an altitude of about 14 km. Also the diurnal variation in the cirrus clouds (*i.e.*, dissolution toward noon, re-formation toward evening), which has been occasionally observed in the subtropical deserts, can probably be ascribed to the fact that the ground temperature is very high at noon [36].

A further consequence of the interaction between absorption of radiation from below and emission upward is the fact that a cloud layer must develop its own internal convection system. The absorption of heat in the lower portions will lead to an evaporation of the droplets, while the emission from the upper portions will lead to increased condensation and a descent of the heavier cloud. Thereby, the stratified cloud is resolved into individual convection cells, stratus is turned into stratocumulus and altostratus into altocumulus. This process may take place fairly rapidly. A stratified cloud ½ km thick, at an altitude of about 5 km, is converted and "destabilized" from the isothermal state to one with a temperature gradient of 0.5C per 100 m in approximately twenty minutes, whereas a similar cloud at a height of 2 km requires three quarters of an hour to complete the same change [32]. Keeping this in mind, it seems scarcely credible that an ordinary cloud layer can exist unchanged in the atmosphere for any length of time without being dissolved. If, in spite of the foregoing discussion, thin, closed, and stable altostratus cloud layers are observed, it becomes clear from the radiation calculations to what degree they must be sustained by a process which constantly re-forms them by new condensation. This process may be vertical austausch or upgliding, and its effectiveness must be considerable even in a cloud that appears to be stable and unchanging.

Synoptic Situations and Radiation of the Free Atmosphere

It is easy to survey schematically the radiation processes of an individual cloud. However, conditions become more complicated if it is desired to calculate approximately the effect of the clouds under different weather conditions or in different climatic regimes. Nevertheless, such calculations are necessary, since they offer the first possibilities for surveys. An attempt in this direction [36] is reproduced in Fig. 4. In this figure the average cloud conditions of a low-pressure area in middle latitudes are assumed, that is, 10 per cent of the area is clear, 20 per cent has clouds resting on the ground, 50 per cent has clouds with the lower level between 0 and 2 km, and 20 per cent has clouds with the lower level between 2 and 8 km. Correspondingly, in 40 per cent of the total low-pressure area, the upper cloud boundary is assumed to lie between 0 and 3 km, in 20 per cent between 3 and 8 km, and in 30 per cent between 8 and 10 km. In the average cooling curve shown in Fig. 4, the heat loss of the most important cloud levels at 2 and 9 km is distinctly noticeable, the lower by a cooling of almost 2C per day, the upper by more than 5C per day, for at this level the heat loss by the radiating upper cloud surface becomes more effective because of the reduced density of the air. The horizontal distribution in the low-pressure system cannot be distinguished, because the various factors are averaged over the entire area. However, the cooling rate must be about three times as great in the advance portion of the cyclone where the upgliding cloud screen lies as a closed, though loose and diffuse, cover at 8 km altitude. It is obvious that such a cooling of 15C per day will noticeably influence the weather development and the cloud dynamics. Further investigations of this kind should yield valuable insight into the thermodynamics of weather.

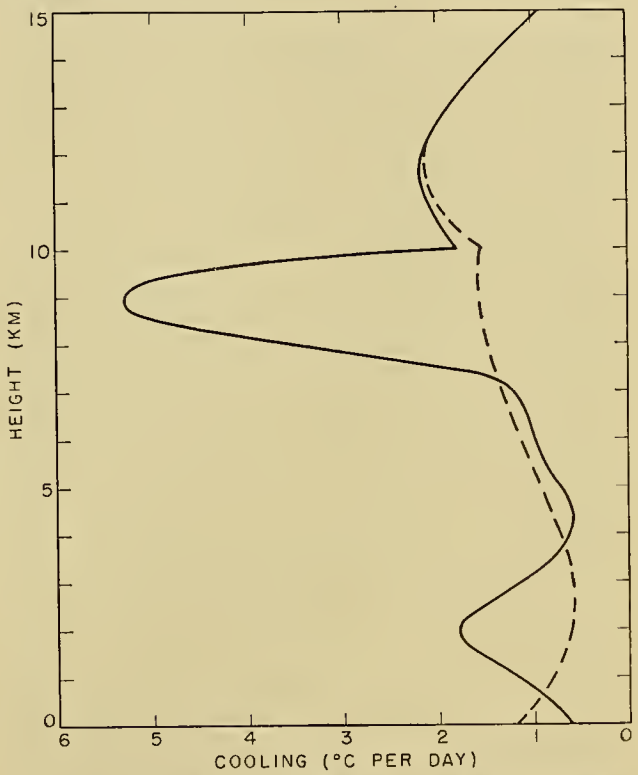


Fig. 4.—Cooling of the atmosphere by water-vapor radiation in degrees per day. The dashed line applies to a cloudless atmosphere and the solid line to an average distribution of clouds in a low-pressure area.

In 1935 the author tried to give a balance of all heating and cooling processes in the free atmosphere and their vertical distribution [29], in which not only the outgoing radiation but also the incoming solar radiation and the liberation of the latent heat of condensation were considered. The calculations for long-wave radiation will be revised here.

A normal temperature and humidity distribution for middle latitudes was assumed as a basis for the calculations. The cooling of the cloudless atmosphere, as now computed, differs from the earlier calculation. At that time, the cooling of the entire troposphere was found to be constant at 1C per day. In the new calculation (Fig. 5, curve *A*) it increases with altitude to 2¾C per day (at 9 km). The cause of this difference lies in

part in the use of the improved radiation diagram, but principally in the realization that the water-vapor content of the stratosphere is much lower than was formerly assumed. For this reason—as mentioned above—the main emission level is shifted down to the altitude of the tropopause. The importance of reliable measurements of the water-vapor content of the stratosphere for these investigations cannot be over-emphasized, for a higher water-vapor content sharply reduces the emission at the level of the tropopause. So far only two or three measurements of the frost point have been published. They show an unchanged decrease above the tropopause [13] sometimes interrupted by thin saturated layers [5]; however, we do not know whether, for example, the water-vapor content over low-pressure areas is higher than shown by these measurements. There is considerable evidence for this supposition, for otherwise how should the mother-of-pearl clouds observed in Norway develop in the rear portion of cyclones at an altitude of 28 km, if the relative humidity remains at 1 per cent and less from an altitude of 14 km upward? If the humidity is greater, the radiation processes of the tropopause are reduced considerably.

It is also very difficult to make any reasonable statements concerning the distribution with altitude of the upper cloud boundaries. It is here assumed that this boundary lies between 0 and 2 km in 15 per cent of all cases, between 2 and 5 km in 45 per cent, between 5 and 8 km in 30 per cent, and between 8 and 10 km in 10 per cent of all cases. For the lower cloud boundaries, 80 per cent are assumed to lie between 0 and 3 km, and 20 per cent between 3 and 10 km. Consideration of these values gives a cooling of the free atmosphere on completely overcast days as shown in Fig. 5, curve *B*, that is, there is a "radiation screen" in the lowest layers, but above 2 km there is an increase of heat loss of about 0.8 to 1 degree per day as compared to a cloudless atmosphere. On the whole, however, the difference between the cloudy and the clear atmosphere is small. The calculation made in 1935 (under the assumption of a somewhat different cloud distribution) gave the greatest cooling at an altitude of 4 km. This might well be taken as an indication of the importance for these investigations of more accurate data respecting the distribution of the clouds in the atmosphere not only on the average, but also for specific weather situations.

An example from an altogether different climatic regime also shows this very clearly. An entirely different temperature and cloud distribution must be assumed for a winter month at the earth's cold pole situated approximately in northeastern Siberia. On cloudless days (50 per cent of all days) the temperature at ground level is assumed to be −50C, rising to −25C at 1.5 km, and decreasing to −60C at 8-km altitude. In the presence of a cloud cover, the temperature between 0 and 2 km is taken to be constant at −30C. The clouds, which are seldom very massive (there are only two days of precipitation per month), are assumed to be restricted to the layer between 0.5 and 2 km. On clear, as well as on cloudy days, an extremely strong cooling layer results between 1 and 2 km, with an average of 5C per day (Fig. 5, curve *D*). Aloft the cooling is comparatively slight. Thus the vertical arrangement of the heat balance in this continental winter climate deviates considerably from that of our middle latitudes under maritime influence.

As a third example, a tropical atmosphere is represented (Fig. 5, curve *C*). There is little difference up to 7 km as compared to the temperate latitudes. The maximum of the outgoing radiation lies at an altitude of 13 km and is not affected by processes resembling heat conduction as in the temperate tropopause at 10 km. This appears to be an approach to a solution of

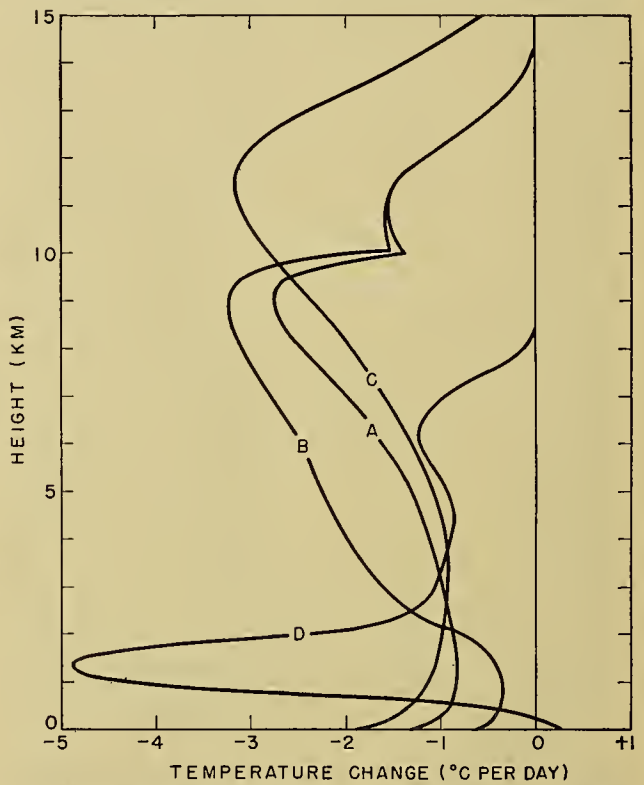


Fig. 5.—Temperature change of the atmosphere by water-vapor radiation in degrees per day.
(*A*) Normal atmosphere in middle latitude, cloudless.
(*B*) The same with an average distribution of clouds.
(*C*) Tropical atmosphere, cloudless.
(*D*) Temperature and cloud distribution at the cold pole.

the old problem concerning the origin of the tropopause. In middle latitudes we find the maximum of heat loss in the region of the tropopause. There, it may be a fair assumption to consider the water-vapor radiation as the cause of the tropopause. In the tropics, the layer of strongest cooling lies about 4–5 km below the tropopause. Thus, it is most probable that the tropopause is produced in a different manner in the tropics than in temperate latitudes. Whether steady-state dynamic phenomena play a role here, or whether the CO_2-radiation becomes important, remains as yet unexplained. Explanations which distinguish between the tropopause in temperate and in tropical regions have also been developed by Goody [20]. Finally, further indication is given by the abrupt discontinuity in the altitude of the tropopause at lat 30°, which was sus-

pected as early as 1935 [30] and which was recently demonstrated by Hess [23].

Radiation cannot be responsible for day-to-day changes in the tropopause of the temperate latitudes, since its effect is much too slow, as shown by Junge [26]; in this case dynamic processes are certainly important. On the average, however, radiation of water vapor is as responsible for the low temperature at the tropopause, as is the decrease of solar radiation from the equator to the pole for the meridional temperature distribution of the troposphere, although in the latter case, day-to-day variations are likewise considerable.

Numerical and Analytical Radiation Calculations

All calculations of radiation processes mentioned so far are based on the method of the radiation diagrams. There is no doubt that these are somewhat cumbersome. The determination of the heating and cooling processes via the radiation fluxes and their differentiation with respect to altitude seems especially laborious. Bruinenberg [8] has made a valuable contribution here toward attaining the objective directly. He puts the differentiation with respect to altitude, which can also be replaced by differentiation with respect to temperature, under the integral of the radiative flux in equation (13) or (15). Thus, after calculations of a scope analogous to those required for the construction of a radiation diagram, he arrives at expressions which are suitable for numerical integration or summation. However, these equations are less suitable for a graphical treatment. Therefore, Bruinenberg has calculated tables which permit very simply the determination and addition of the cooling or heating effects exerted on a given element by all atmospheric layers above and below it.

The principal advantage of the individual calculation is the fact that the cooling for points closely spaced along the vertical can be determined independently and with great accuracy. This leads to a result which, though not unexpected, has thus far been underestimated in its implications. Every break in the vertical temperature distribution is manifested as a sharp peak in the distribution of the cooling rate. Thus cooling peaks up to 3C per day project from the average cooling level of 1C per day in the lower troposphere at every break of the characteristic temperature curve directed toward higher temperatures, and heating peaks up to +1C or +2C per day where the characteristic curve breaks toward lower temperatures. Heating of +5C per day occurs below an inversion, whereas there is a cooling of 15C per day at its upper boundary. However all this holds true only in very thin layers. (Bruinenberg's method is so far applicable only to calculations in the lower and middle troposphere; for the tropopause, see p. 42.) At these points the radiation simply acts as a temperature equalizer in a manner similar to heat conduction. Every break in the characteristic temperature curve is equalized at an accelerated rate. From the foregoing discussion it follows again that if inversions or more or less sharp discontinuities in the temperature gradient persist for days, the ordinary radiation processes of the water vapor cannot be the cause. Thus the question whether long-wave radiation can produce inversions is decided partially in favor of the negative [4]. For the maintenance of existing inversions, some processes must be constantly active that recreate the inversion continuously against the equalizing effects of the radiation. Such processes are partly the additional radiation from the haze layers below the inversion which overcompensate the heating (see p. 42) and partly the dynamic processes of shrinking and subsidence. The intensity of these dynamic processes can then be estimated from the calculations of radiation.

An especially significant level is the earth's surface. For radiation effects, the ground may be conceived as replaced by an infinitely extended isothermal layer of water vapor or CO_2. In such a case the corresponding characteristic temperature curve extended into the ground has a break at the ground surface. If there is a temperature decrease with altitude above the ground, there must be strong cooling directly at the ground surface, whereas there is strong heating at the base of a ground inversion. Such radiative processes will scarcely become operative, since all observations disprove them. However, in every case, they will tend toward an isothermal state near the ground as an equilibrium condition. Even if this equilibrium is not reached, its quantitative consideration will possibly lead to a revision of the assumptions concerning the magnitude of the austausch near the ground as far as has been disclosed by measurements of the temperature gradient.[4]

Although the graphical and numerical calculation methods are carefully worked out, an analytical equation can be very advantageous at times because it permits combination of the radiation process with others that can be approached theoretically. Such a possibility is offered by the quasi-conduction of radiation introduced by Brunt [9]. However, this method will not include the processes at the ground surface which were discussed above. The complicated construction of a radiation diagram makes it apparent that a *complete* description by a single convenient equation is impossible. Simplifications must be made. One such simplification consists of running the temperature lines horizontally in the Möller diagram (in the water-vapor section). This means that we do not change the distribution of the absorption coefficients over the wave lengths, but that we do assume that the energy curve has the same shape for all temperatures, instead of assuming Planck's formula for black-body radiation. This would imply, for example, that, if we take the shape of the energy distribution at 273K, the radiation for a different temperature would be given by multiplying this curve by the factor $T^4/273^4$ which is not a function of the wave length. In the Elsasser chart this assumption would mean that the w-lines would be rectilinear and convergent to the point where $T = 0$K. If, in addition, we approximate the abscissa scale X of the Möller chart with any convenient function, new problems can be solved.

4. (Note added July, 1950.) In his newest publication, Robbinson [44] also comes to the conclusion that radiation processes are very important near the earth's surface.

An initial attempt of this type was made by approximating the function X by the sum of two exponential functions. This would mean that the absorption spectrum of the water vapor is not gray but "bichromatic." The investigation of the radiation equilibrium of the atmosphere was carried out by Emden [15] for gray radiation and led to an isothermal stratosphere of −68C. The calculation of the same problem for bichromatic absorption of water vapor does not lead to an isothermal stratosphere but yields a steady decrease in temperature with altitude down to −144C at the boundary of the atmosphere [31]! If it still were at all necessary to deal the death blow to the untenable theory of the radiation equilibrium of the water vapor in the stratosphere, this calculation would do so.

Another approximation of X is fundamentally more accurate. If we set $X = a \ln(w + b)$, where a and b are numerical values, we obtain an approximation that is especially good for small values of w and small amounts of CO_2. Then $dX = adw/(w + b)$, and from this simple formula the radiation equilibrium can be developed as a kind of integral equation and can be brought closer to numerical solution. This method appears important especially for the temperature distribution in immediate proximity to the ground because, under conditions of both midday adiabatic stratification and nocturnal ground inversion, this temperature distribution is a function not only of the austausch but also of radiation. Results of these calculations are not yet available.

Radiation in the Stratosphere

Investigation of radiation phenomena in the stratosphere is more difficult than in the troposphere for several reasons. First of all, the absorbing media are under very low pressure, and therefore a check must be made in every case to ascertain to what extent the absorption coefficients, as measured in the laboratory, may be used. In the second place, the concentration of the various gases is not known exactly. Finally, the radiation cannot be considered any longer as an individual process out of the context of the general physical phenomena. In the troposphere also, radiation participates everywhere, to be sure, in the general weather pattern, and in the presentation above, emphasis has been on showing that radiation participates decisively in all meteorological processes. However, the substances themselves are not changed in their molecular structure by absorption or emission. If, for instance, water vapor is brought to condensation and fog forms as a result of strong radiational cooling, it changes only its aggregate state, not its chemical structure. In the stratosphere, however, ozone offers an example of more drastic changes. It is only by the absorption of the solar radiation in the ultraviolet bands that the formation of O_3 from O_2 becomes possible, and it is under the influence of longer waves in the ultraviolet that the ozone again dissociates. It is indeed true that it is not the long-wave heat radiation but the short-wave solar radiation that controls the formation and dissociation of the absorbing medium. In addition, however, both processes participate in the heat balance with their thermal implications. Therefore, the radiation processes in the stratosphere cannot be separated from the multitude of physical atomic processes at those altitudes. If, in spite of this, such an attempt is made, it will necessarily be in full consciousness of the unreliability involved.

Above 10-km altitude the importance of water vapor as a decisive absorbing medium decreases more and more, and CO_2 and O_3 become significant. The water-vapor content of the air in the stratosphere is extremely small according to measurements by Dobson [13], Barrett [5], and others. For several kilometers above the tropopause the frost point falls steadily at the rate of its decrease with altitude in the troposphere, so that the relative humidity at 2 km above the tropopause has decreased to below 1 per cent. However, if we accept this low humidity as a generally valid fact, the radiation effect of the water vapor in the stratosphere no longer exerts any notable effect, since its total content drops to 10^{-4} g cm^{-3}. Dobson, however, thinks that its effect is to be considered equivalent to that of CO_2, but gives no numerical values for the amount of absorbed radiation.

Carbon dioxide has a very intensive absorption band around 15 μ. An extremely weak band around 10 μ absorbs only one per cent for layer thicknesses that equal the content of the whole atmosphere. An additional band at 4 μ lies at the boundary of the spectrum. The absorption curve in the 15-μ band is known. The dependence on pressure is usually estimated by the effect on the 4-μ band which is known from measurements by Wimmer [51]. From this Möller has derived a reduction factor $\mu = (p/p_0)^{0.74}$, whereas Elsasser uses the $\sqrt{p/p_0}$ law in the same manner as for the water vapor, and Goody a proportionality to p. A calculation of the processes is possible by means of the Möller diagram. Möller calculated the CO_2-radiation emitted by an isothermal stratosphere and found a maximum effect at an altitude of 26 km with a cooling of 1.5C to 2C per day. Since his assumption of a CO_2-content of 0.03 per cent is somewhat too high, the maximum lies probably a little lower.

In this calculation it was assumed that the ozone has no effect. But, as mentioned already, ozone has a very intensive absorption band around 14 μ which is situated at the same point as that of CO_2 and shows a curve with respect to wave length which is similar to that of CO_2. The line structure of ozone which probably differs from that of CO_2, is unknown, however, and probably does not enter into the calculations. Since half of the ozone lies at altitudes above 20 km, the radiation of the CO_2 is extensively shielded by its absorption, and the outgoing radiation in this range of wave lengths takes place only at higher altitudes and then as an emission of the ozone. Thus, these processes interact strongly with each other, and for this reason scarcely any attempts have been made to investigate them in greater detail [33].

Furthermore, the experimental bases for the ozone spectrum are not yet sufficient. In addition to the 14-μ band, there is another band around 9.6 μ, which

lies exactly in the "window" of the water vapor and which therefore can contribute to the effectiveness of the outgoing radiation emitted by the earth's surface despite the weak absorption and the small amount of ozone. Adel [2] measured the intensity of the solar spectrum at this wave length in an excellent experimental investigation and was able to detect the absorption in this band. Its maximum is about 50 to 70 per cent. The old laboratory measurements by Hettner and collaborators [24] were made on large amounts of ozone under high pressure. From their absorption coefficients and the normal amount of ozone in the atmosphere the maximum absorption found at 9.6 μ is only 14 per cent. This contradiction was explained by Strong [47], who made measurements with ozone under low total pressures. He was able to show that the supplementary atmospheric pressure greatly broadens the absorption lines of ozone, whereas an increased partial pressure causes only a minor broadening of these lines. It is for this reason that the absorptions measured in pure ozone without admixture of air are much too small in spite of large quantities of ozone. The application of the absorption values as measured by Strong leads to a maximum absorption at 9.6 μ, equal to that found in the solar spectrum. Though these processes are explained for the 9.6-μ band, this is not true for the 14-μ band for which similar measurements are completely lacking. In this case we must depend on analogous conclusions.

Through numerous investigations we are well informed concerning the amount of ozone in the atmosphere and its vertical distribution. Only recently were the optical determinations of this vertical distribution excellently confirmed by direct measurements. Previous calculations of the radiation phenomena in the ozone (Gowan, Penndorf) are based on the uncorrected laboratory measurements of the absorption by Hettner. The results are therefore incorrect. Recently, a new calculation was made by Oder [38]. He uses values for the absorption coefficient which in each case give only an average for the whole band. This incorrectly distorts the absorption function, and his results are therefore apparently inaccurate. However, the numerical values, which are presented in Table IV, are noteworthy. At an altitude of about 40 km the emission of radiant heat produces a cooling of about 8C per hour. It will not be very easy to explain what processes compensate for this large cooling, but they must be compensated somehow if the assumption that the temperature distribution remains stationary is correct. Goody [20] made the most important contribution to the theory of radiation of the tropopause. He assumed that in this region a discontinuity exists in the vertical temperature gradient, namely from 6C per km in the troposphere to isothermal conditions in the stratosphere. By means of separate calculations of the individual bands of the three absorbers, water vapor, carbon dioxide, and ozone, he investigated the radiation balances and their dependence on barometric pressure and temperature at the tropopause. He found that in middle and high latitudes an equilibrium exists between the heating effect of carbon dioxide and the cooling effect of water vapor; in the tropics, however, water vapor, because of its extremely small concentration, no longer has any effect. There, an equilibrium exists between the effects of carbon dioxide and ozone. At the great heights and the low temperatures of the tropical tropopause, however, carbon dioxide has a cooling effect, ozone a very faint heating effect. This led Goody to the remarkable concept that radiative equilibrium always depends on the contrast between two different absorbers, and that in the tropics the participating media are different from those in middle latitudes. The quantitative bases of these computations appear to be still inadequate. For this reason, verification would be most desirable. Also, it appears to this writer that the selection of the tropopause for these calculations is somehow not quite suitable because of the peculiarities of radiative processes at points of discontinuity in the temperature gradient. However, it is difficult to suggest altitudes that would be more suitable for such computations.[5]

Table IV. Radiational Heating of the Stratosphere (*According to Oder* [38])

h (km)	15	20	25	30	35	38	41	44	47
$\partial T/\partial t$ (deg C per day)	0.1	−0.6	−0.5	3.1	(−44)	(−120)	(−172)	−98	3.3
$\partial T/\partial t$ (deg C per hour)	—	—	—	—	−1.8	−5.0	−7.2	−4.1	—

Suggestions for Future Research

Though the foregoing exposition touched only briefly on the experimental foundations, it has nevertheless been shown that the most important lines along which research must now proceed are of an experimental or observational nature. The following appear to the author to be of particu'ar importance:

1. Absorption or emission of water-vapor layers of limited thickness must be checked by laboratory and free-air experiments. The available measurements seem insufficient to explain the variation with temperature that results from the variation of the observed atmospheric radiation. Such measurements are a very important basis for radiation diagrams and fo r all conclusions drawn from them regarding the free atmosphere.

2. Long-wave radiation, particularly in the free atmosphere, must be measured. Since there are filter substances available which are not only very good in the long-wave range but which are uniform for all wave lengths [25], there should be no basic difficulty in con-

5. (Note added July, 1950.) A recent investigation by Plass and Strong [40] may clarify this problem. However, only an abstract of this work has been published thus far.

structing instruments to be mounted in aircraft. This would shift investigations into entirely new channels.

3. Adequate data concerning the content of water vapor and carbon dioxide of the upper troposphere and the lower stratosphere are needed for investigation of tropospheric radiation by means of the familiar graphical methods. Spot checks are inadequate. Measurements are needed in such numbers that radiation changes with the weather situation become clear. The influences of the geographical latitude and of continents and oceans on the content of water vapor and carbon dioxide must also be determined. The same holds true for the determination of the upper cloud boundary for various weather patterns and various types of climate. Radiosonde observations are not sufficient for this purpose; direct observations from aircraft are needed. We must consider such observations as the principal demand which radiation research makes on the field of observational aerology.

4. The influence of pressure on the absorption by CO_2 and O_3 in the 15-μ band must still be investigated in the laboratory for application to the study of the radiation processes in the region of the tropopause. Only then can we approach an explanation of the radiation processes at these altitudes with any hope of success.

REFERENCES

1. Adel, A., "Absorption Line Width in the Rotation Spectrum of Atmospheric Water Vapor." *Phys. Rev.*, 71: 806–808 (1947).
2. —— and Lampland, C. O., "Atmospheric Absorption of Infrared Solar Radiation at the Lowell Observatory; 5.5—14μ." *Astrophys. J.*, 91: 1–7 and 481–487 (1940).
3. Ångström, A., "Messungen der nächtlichen Ausstrahlung im Ballon." *Beitr. Phys. frei. Atmos.*, 14: 8–20 (1928).
4. Ashburn, E. V., "The Vertical Transfer of Radiation through Atmospheric Temperature Inversions." *Bull. Amer. meteor. Soc.*, 22: 239–242 (1941).
5. Barrett, E. W., Herndon, L. R., Jr., and Carter, H. J., "A Preliminary Note on the Measurement of Water-Vapor Content in the Middle Stratosphere." *J. Meteor.*, 6: 367–368 (1949).
6. Bolz, H. M., and Falckenberg, G., "Neubestimmung der Konstanten der Ångström'schen Strahlungsformel." *Z. Meteor.*, 3: 97–100 (1949).
7. Brooks, F. A., "Observations of Atmospheric Radiation." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 8, No. 2 (1941).
8. Bruinenberg, A., "A Numerical Method for the Calculation of Temperature-Changes by Radiation in the Free Atmosphere." *Meded. ned. meteor. Inst.*, (B), Deel 1, Nr. 1 (1946).
9. Brunt, D., "The Transfer of Heat by Radiation and Turbulence in the Lower Atmosphere." *Proc. roy. Soc.*, (A) 124: 201–218 (1929).
10. —— "Some Phenomena Connected with the Transfer of Heat by Radiation and Turbulence in the Lower Atmosphere." *Proc. roy. Soc.*, (A) 130: 98–104 (1930).
11. Callendar, G. S., "Infra-red Absorption by Carbon Dioxide, with Special Reference to Atmospheric Radiation." *Quart. J. R. meteor. Soc.*, 67: 263–274 (1941).
12. Cwilong, T. G., "Atmospheric Absorption of Heat Radiation by Water Vapour." *Phil. Mag.*, (7) 41: 109–123 (1950).
13. Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., "Meteorology of the Lower Stratosphere." (Bakerian Lecture) *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
14. Elsasser, W. M., "Heat Transfer by Infrared Radiation in the Atmosphere." *Harvard meteor. Stud.*, No. 6 (1942).
15. Emden, R., "Über Strahlungsgleichgewicht und atmosphärische Strahlung." *S. B. bayer. Akad. Wiss.*, 43: 55–142 (1913).
16. Falckenberg, G., "Über die Abhängigkeit der Gegenstrahlung der Atmosphäre vom Temperaturgradienten und vom Wetter." *Meteor Z.*, 57: 241–249 (1940).
17. —— und Hecht, F., "Messung der infraroten Eigenstrahlung der Atmosphäre vom Flugzeug." *Meteor. Z.*, 58: 415–417 (1941).
18. Falckenberg, G., und Stoecker, E., "Bodeninversion und atmosphärische Energieleitung durch Strahlung." *Beitr. Phys. frei. Atmos.*, 13: 246–269 (1927.)
19. Fritz, S., "Measurements of the Albedo of Clouds." *Bull. Amer. meteor. Soc.*, 31: 25–27 (1950).
20. Goody, R. M., "The Thermal Equilibrium at the Tropopause and the Temperature of the Lower Stratosphere." *Proc. roy. Soc.*, (A) 197: 487–505 (1949).
21. Groen, P., "Note on the Theory of Nocturnal Radiational Cooling of the Earth's Surface." *J. Meteor.*, 4: 63–66 (1947).
22. —— "On Radiational Cooling of the Earth's Surface during the Night, Especially with Regard to the Prediction of Ground Frosts." *Meded. ned. meteor. Inst.* (B) Deel I, Nr. 9 (1947).
23. Hess, S. L., "Some New Mean Meridional Cross Sections through the Atmosphere." *J. Meteor.* 5: 293–300 (1948).
24. Hettner, G., Pohlmann, R. und Schumacher, H. J., "Die Struktur des Ozon-moleküls und seine Banden im Ultrarot." *Z. Phys.*, 91: 372–385 (1934).
25. Joos, G., "Optische Eigenschaften der festen Körper." *Naturforschung und Medizin in Deutschland, 1939–1946* (*FIAT Rev.*). Wiesbaden, Dieterich Verlag, 1948. (See Vol. 9, Pt. 2, pp. 164–184)
26. Junge, C., "Zur Strahlungswirkung des Wasserdampfes in der Stratosphäre." *Meteor. Z.*, 54: 161–164 (1937).
27. Linke, F., *Meteorologisches Taschenbuch*, 4. Aufl. Leipzig, Akad. Verlagsges., 1939. (See p. 262)
28. —— und Möller, F., "Langwellige Strahlungsströme in der Atmosphäre und die Strahlungsbilanz," *Handbuch der Geophysik*, Bd. 8, Kap. 11. Berlin, Gebr. Bornträger, 1943. (See pp. 668–721)
29. Möller, F., "Die Wärmequellen in der freien Atmosphäre." *Meteor. Z.*, 52: 408–412 (1935).
30. —— "Höhenwindmessungen und horizontales Temperaturfeld." *Beitr. Phys. frei. Atmos.* 22: 299–307 (1935).
31. —— "Die Wärmestrahlung des Wasserdampfes in der Atmosphäre." *Beitr. Geophys.*, 58: 11–67 (1941).
32. —— "Labilisierung von Schichtwolken durch Strahlung." *Meteor. Z.*, 60: 212–213 (1943).
33. —— "Zur Erklärung der Stratosphärentemperatur." *Naturwissenschaften*, 31:148 (1943).
34. —— "Das Strahlungsdiagramm." *Wiss. Abh. D. R. Reich. Wetterd.* (1943).
35. —— "Grundlagen eines Diagramms zur Berechnung langwelliger Strahlungsströme." *Meteor. Z.*, 61: 37–45 (1944).
36. —— "Wirkungen der langwelligen Strahlung in der Atmosphäre." *Meteor. Z.*, 61: 264–270 (1944).
37. Mügge, R., und Möller, F., "Zur Berechnung von Strahlungsströmen und Temperaturänderungen in Atmosphären von beliebigem Aufbau." *Z. Geophys.* 8: 53–64 (1932).

38. Oder, F. C. E., "The Magnitude of Radiative Heating in the Lower Stratosphere." *J. Meteor.*, 5: 65–67 (1948).
39. Pedersen, F., "On the Temperature-Pressure Effect on Absorption of Long-Wave Radiation by Water Vapour." *Meteor. Ann. Oslo*, Vol. 1, No. 6, pp. 115–136 (1942).
40. Plass, G. N., and Strong, J., "Radiation Equilibrium in the Stratosphere." *Phys. Rev.*, (2) 78: 334 (1950).
41. Regener, E., "Aufbau und Zusammensetzung der Stratosphäre." *Schr. dtsch. Akad. Luft.*, SS. 7–41 (1939).
42. Roberts, O. F. T., "On Radiative Diffusion in the Atmosphere." *Proc. roy. Soc. Edinb.*, 50: 225–242 (1929–30).
43. Robinson, G. D., "Notes on the Measurement and Estimation of Atmospheric Radiation." *Quart. J. R. meteor. Soc.*, 73: 127–150 (1947).
44. —— "Notes on the Measurement and Estimation of Atmospheric Radiation—2." *Quart. J. R. meteor. Soc.*, 76: 37–51 (1950).
45. Robitzsch, M., "Strahlungsstudien." *Arb. preuss. aero. Obs.*, 15: 194–213 (1926).
46. Schnaidt, F., Über die Absorption von Wasserdampf und Kohlensäure mit besonderer Berücksichtigung der Druck- und Temperaturabhängigkeit." *Beitr. Geophys.*, 54: 203–234 (1939).
47. Strong, J., "On a New Method of Measuring the Mean Height of the Ozone in the Atmosphere." *J. Franklin Inst.*, 231: 121–155 (1941).
48. —— "Study of Atmospheric Absorption and Emission in the Infrared Spectrum." *J. Franklin Inst.*, 232: 1–22 (1941).
49. Volz, F., *Untersuchungen über den Einfluss der Trübung auf die langwellige Strahlung in der Atmosphäre.* Diplom-Arbeit, Meteor. Inst. Frankfurt/M., 1950 (unpublished).
50. Wexler, H., "Observations of Nocturnal Radiation at Fairbanks, Alaska, and Fargo, N. Dak.," *Mon. Wea. Rev. Wash.*, Supp. No. 46 (1941).
51. Wimmer, M., "Über die Beeinflussung der ultraroten Kohlensäureabsorptionsbande bei 4.27 μ durch fremde Gase und ihre Anwendung zur Gasanalyse." *Ann. Physik.*, 81: 1091–1112 (1926).

ACTINOMETRIC MEASUREMENTS

By ANDERS ÅNGSTRÖM

Meteorological and Hydrological Institute of Sweden

ACCURACY OF ACTINOMETRIC MEASUREMENTS

In actinometry, as in every other field of science founded upon some kind of direct measurement, the instrumental accuracy and the method of observation must be closely related to the character of the scientific problem under consideration.

Determination of the Solar Constant. Some scientists maintain that this so-called "constant" is subject to short periodic variations amounting to approximately 0.2 per cent on the average, with a maximum amplitude of about 1–2 per cent. It is evident that in order to investigate these variations our measuring devices must be sufficiently accurate to measure amounts of radiation below the smallest amplitude of the variations which we intend to determine, in other words, the instrument must measure accurately to at least ± 0.2 per cent.

Heat Exchange at the Earth's Surface. Considerably less accuracy is required in many other problems closely connected with actinometry. Suppose, for instance, that we wish to investigate the heat exchange at the earth's surface through radiation, convection, conduction, evaporation, etc. Here, the radiation enters into an equation in which the other factors involved can hardly be determined to an accuracy greater than 5–10 per cent. Even if we could measure them more accurately, it would be of little benefit, since the values have no general applicability. Convection, conduction, reflection from the earth's surface, and evaporation are all highly variable from place to place. Local measurements are seldom representative for more than very limited areas. An accuracy of ± 3 per cent seems in general quite sufficient for such purposes as actinometric measurements aiming at an evaluation of the heat balance at the earth's surface, ablation studies on glaciers and snow covers, and studies of similar geophysical problems.

Analysis of the Atmosphere. Actinometric measurements are, however, also an important means for the analysis of the content of the various atmospheric constituents. Through rather simple measurements of the total direct solar radiation and of the same radiation within a few selected regions of the spectrum, an evaluation can be made of the total water content in the path of the solar beam as well as of the turbidity (*i.e.*, the content of solid or liquid particles which scatter light in the atmosphere). The principles on which such determinations of the turbidity and water content are founded will be briefly summarized in the following paragraphs.

If the solar "constant" is regarded as a true constant and its relatively small variations are neglected, the variations of the incoming direct solar radiation Q_m at a given solar elevation (air mass = m) may be regarded as due to four principal causes:

1. The variable distance between sun and earth.
2. Molecular scattering.
3. Scattering and absorption by liquid and solid particles in the atmosphere.
4. Selective absorption by the gases of the atmosphere.

The variations resulting from the first cause are well known and easily computed. The scattering by the molecules may be computed from the theory of Rayleigh; the scattering coefficient thus determined is a continuous function of the wave length, being inversely proportional to its fourth power.

Scattering by the solid and liquid particles in the atmosphere may, as a first approximation, also be regarded as a continuous function of the wave length. Strictly speaking, if the physical nature of the particles is known in detail, the scattering coefficient may be computed as a function of the wave length, according to the classical theory of Mie. However, since we seldom or never have a complete knowledge concerning the variable nature of the scattering particles in the atmosphere, it seems allowable to introduce a simplification based on empirical results. From Abbot's extensive observations in various parts of the world the present author found that the scattering coefficient S due to liquid and solid particles in the atmosphere may, in general, be expressed by

$$S = \beta/\lambda^{\alpha}, \qquad (1)$$

where β has a value proportional to the number of particles, and α no longer has a value of 4.0 as in the case of the molecules but varies between 0.5 and 2.0, according to the size of the particles. The smaller the particles, the larger is the value of α. In general, an acceptable average value for α, that holds for ordinary conditions, seems to be 1.3. When the atmosphere has been polluted with larger particles, as after violent volcanic eruptions or through dust storms over deserts, the value of α is sometimes as low as 0.5 or even less. The particles influencing the visibility in the atmosphere near the ground are also evidently larger than the average particle causing the scattering of the solar beam. For these lower layers Schmolinsky [10] found an average value of about 0.9 for α.

These considerations hold approximately for the visible part of the solar spectrum, that is, for wave lengths in the range from 0.4 to 0.8 μ. They are, however, not applicable to the ultraviolet and the far infrared. Finally, the incoming solar beam is weakened also by the selective absorption by the various atmos-

pheric gases, especially by water vapor and carbon dioxide. This absorption is not a simple function of the wave length, but is concentrated in certain spectral regions. The conditions are, however, simplified by the fact that the selective absorption by the variable gases is almost entirely confined to the far red and the infrared, where the principal water-vapor and carbon dioxide bands occur. Only a small part, about 1–2 per cent of

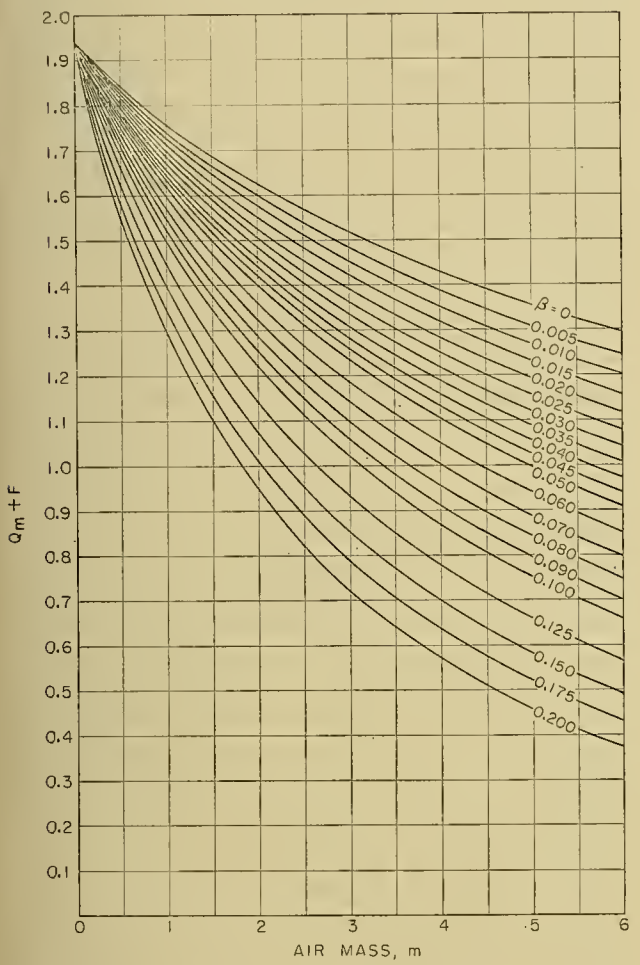


FIG. 1.—Atmospheric turbidity β according to Angström-Hoelper. Ordinate: total radiation measured (Q_m), with addition of water vapor absorption (F). Abscissa: air mass (unit vertical air mass at sea level and 760 mm). Actinometric scale: Smithsonian.

the total energy is subject to a variable absorption within the visible and the ultraviolet.

Consequently, the total incoming solar radiation Q_m, as measured for instance with a pyrheliometer, may be expressed as

$$Q_m = \int_0^\infty I_{0\lambda} q^m \exp(-\beta m/\lambda^a)\, d\lambda - F. \qquad (2)$$

In this equation we may assume Q_m to be known from measurements; $I_{0\lambda}$, the radiation outside the atmosphere at the wave length λ, is known from the elaborate investigations by the Smithsonian Institution (Abbot and collaborators);[1] q, the transmitted fraction of the incident energy for unit air mass (if only molecular scattering is considered), is computed from the theory of Rayleigh,[2] the air mass m is computed from the observed elevation of the sun (m is unity for zenith position of the sun). If we assume a to have a given value, the only unknown quantities are β and F. The total selective absorption F has been expressed by Fowle through the linear equation

$$F = 0.10 + 0.0054 e_0 m, \qquad (3)$$

where e_0 is the water-vapor pressure (mm Hg) at the surface of the earth, and m is the air mass. It is evident from several considerations that Fowle's equation must include rather rough approximations. However, if we accept it as a first approach, it is evident that the quantity β, which is a measure of the dust content, may be determined from equation (2) and a single pyrheliometric observation. In practice this determination is made by a graphical evaluation of the integral for various values of e_0 and m—made once and for all. The value of β which makes the value of the integral equal to the measured value of the radiation increased by F is then obtained from a diagram or from tables. We may effect the computation most simply by plotting the observed radiation, increased by F, on the diagram shown in Fig. 1 and interpolating the value of β.

We can, however, derive a similar result more accurately if we add another pyrheliometric measurement to the one covering the total solar spectrum. By using a colored glass filter—RG 2 or OG 1 for instance—we can examine separately a part of the spectrum covering all wave lengths longer than a given limiting value. The filter RG 2, for instance, lets through all radiation of wave lengths longer than about 0.6 μ. With due regard to the nearly constant reflection $(1 - \gamma)$ by the filter (where γ is the fraction of incident radiation transmitted by the filter, that is, its transmission coefficient), we obtain for the observed filter radiation Q_r, in perfect analogy with equation (2),

$$Q_r = \gamma \int_{0.6}^\infty I_{0\lambda} \psi(m, \beta, \lambda)\, d\lambda - \gamma F, \qquad (4)$$

where

$$\psi = q^m \exp(-\beta m/\lambda^a).$$

The value of F may here be assumed with good approximation to be the same as in (2). Hence, from equations (2) and (4), we obtain

$$Q_m - \frac{1}{\gamma} Q_r = \int_0^\infty I_{0\lambda} \psi\, d\lambda - \int_{0.6}^\infty I_{0\lambda} \psi\, d\lambda, \qquad (5)$$

or

$$Q_m - \frac{1}{\gamma} Q_r = \int_0^{0.6} I_{0\lambda} \psi\, d\lambda. \qquad (6)$$

1. Given for instance in F. Linke, *Meteorologisches Taschenbuch*, IV, Table 109, p. 238 (Akad. Verlagsges.) Leipzig, 1931.

2. See F. Linke, *Meteorologisches Taschenbuch*, IV, Table 113, p. 240.

On the basis of this simple theory, the present author, and later Hoelper [6], Tryselius [11], and Olsson [8], among others, have computed values for β and the total selective absorption from simple pyrheliometric observations.

This method for treating actinometric observations has been discussed in some detail because considerations in the foregoing discussion are apt to provide an answer to the question which follows. Suppose we wish to use the simple actinometric measurements of Q_m and Q_r to obtain some idea of the scattering and absorbing properties of the atmosphere (such a purpose is certainly an important justification for actinometric measurements in general), what is the accuracy that we should demand? It is perfectly clear from the simple theory presented here that, in order to make progress and obtain conclusions of importance, we are forced to introduce some simplifications in our assumptions concerning the laws governing atmospheric scattering. Such a simplification has been made in assuming a simple exponential expression for the dependence of the scattering on wave length. This simplification represents a rather rough approximation, and because of this there is very little use in attempting to determine values of β and α with an accuracy greater than about 5–10 per cent. This corresponds to accuracies in the values for the solar radiation Q_m and Q_r of about 1–2 per cent. For determinations of the scattering and absorption in the atmosphere by simple actinometric methods, this appears to be the desirable accuracy.[3]

Actinometry with Regard to Biological Problems. The importance of solar radiation for a number of phenomena of biological character, such as plant growth and photosynthesis, and the health of man, has to a large extent led to the organization of actinometric measurements in connection with research concerning such phenomena.

What accuracy is here required from the actinometric instruments? In trying to answer this question, we must keep certain facts in mind. It is evident that, in general, the radiation which we are able to measure or record is very seldom the radiation that is effective in the biological processes under investigation. Neither the radiation on a horizontal surface, recorded for instance by a pyranometer, nor on a spherical surface, nor on a surface perpendicular to the solar beam as in pyrheliometers, is equal or strictly proportional to the radiation falling on a given plant or other organism. Furthermore, it is seldom of much use to try to construct a perfect model of a plant or organism since the effectiveness of the radiation is, in general, quite different at different parts of the plant's surface. In addition, a number of other factors, of which as a rule we have a rather incomplete knowledge, are generally effective in producing the observed results on, for instance, photosynthesis or growth. Therefore, if we ask for the equipment which ought to be recommended for meteorological stations when we have biological or agricultural needs in mind, it seems more important that the instruments should be characterized by a certain simplicity and stability and that their results should be easily comparable with those from other stations, than that the instruments should be given some elaborate form in a rather vain attempt to imitate what can hardly be reproduced. Instruments which measure or record the radiation from the sun and sky on a horizontal surface and on a surface perpendicular to the solar beam thus seem to be able to satisfy more general needs. An accuracy of 5 per cent seems sufficient, provided that the instruments are not subject to systematic errors or systematic changes.

INSTRUMENTS AND MEASUREMENTS

Determination of the Solar Constant for the Purpose of Ascertaining Its Variations. From what has been said above concerning the accuracy which must be demanded from the instruments for determining the solar constant, it must be concluded that none of the standard types of pyrheliometers in current use strictly satisfies these requirements. The two instruments which should be given first consideration are the compensation pyrheliometer of K. Ångström and the silver-disk pyrheliometer of the Smithsonian Institution (Abbot). Both have been subjected to rather elaborate investigation and critical examination by various scientists and lately, in a very systematic way, by the International Radiation Commission. On the initiative of this Commission, Courvoisier [5] at the Observatory of Davos has made a complete theoretical survey of the principles of their construction, supplemented to some extent by experimental studies.

With regard to the Ångström compensation pyrheliometer, in the form in which it has been delivered by its manufacturers in Uppsala (Rose, now discontinued) and Stockholm (A. Lindblad), the following can be said of earlier as well as of more recent researches.

As shown by the present author in 1914, the instrument is subject to a small error called the "edge effect" arising from certain features in its construction. The "edge effect" is caused by the fact that one of the strips which constitute the sensitive part of the instrument is heated electrically throughout its entire length, while the other strip is illuminated by the solar radiation only to about $^9/_{10}$ of its length on account of the screening of a diaphragm inside the instrument. The error which thus arises is of the order of 2 per cent. It is, however, to some extent dependent on the convection in the air over the heated strips. Taking into account the conditions occurring in practice, we must conclude, from theoretical considerations supported by actual measurements, that the error arising from the edge effect may vary between the limits 2 ± 0.5 per cent.

3. For a more detailed presentation of the methods for determining atmospheric turbidity on the basis of actinometric measurements along the lines indicated above, reference may now be made to a valuable treatise by Dr. Walter Schüepp, published after the present article was written: *Die Bestimmung der Komponenten der atmosphärischen Trübung aus Aktinometermessungen.* Inaugural Dissertation, Wien, Springer, 1949.

To obtain measurements for computing the solar constant Abbot has constructed a special Ångström pyrheliometer in which the edge effect is reduced to practically zero. However, a detailed survey of the factors influencing measurements with this type of instrument leads one to conclude that errors of the order of ±0.3 per cent can hardly be avoided completely.

The silver-disk pyrheliometer has been extensively used, especially in connection with Abbot's earlier work on the solar constant. Courvoisier has subjected it to a very thorough theoretical examination, with due regard to the physical constants of the instrument. Some minor corrections must be applied to Abbot's theory on which the method of measurement is founded. These corrections may affect the absolute values obtained. What interests us here especially are the variable errors arising from the variation of different physical and meteorological factors. Strictly speaking, Abbot's theory holds true only when the instrument is maintained at semiconstant temperature during the periodical temperature variation to which it is subjected when alternately exposed to and screened from solar radiation. This puts a rather high demand on the precautions to be taken during exposure.

Further effects arise from the unavoidable fluctuations of the temperature of the surrounding air, which in turn affect the temperature of the cylindrical wooden cover by which the silver disk is protected. Courvoisier concludes that errors up to 0.3 per cent are possible on that account. Mörikofer [7] estimates the average error from various causes to be about 0.5 to 1.0 per cent.

Among the instruments rather widely used, the Michelson actinometer is especially convenient, since it is rapidly read and does not require auxiliary instruments. Its accuracy is less than that of the Ångström or the silver-disk instruments, the average error probably amounting to ±1.0–1.5 per cent. The Michelson actinometer should be checked at regular intervals since it is sometimes subject to deterioration. As the accuracy of the instrument seems sufficient for general meteorological purposes, it would be highly desirable that instruments of a similar construction should be built in greater number and at moderate cost.

Small differences exist between the indications of all actinometers in practical use for measuring direct solar radiation. They arise from the fact that the instruments receive not only direct solar radiation but also diffuse radiation from the sky in the immediate neighborhood of the sun, within certain aperture angles which vary with the instrument. Thus the silver-disk pyrheliometer has an aperture of 6° (formerly 10.5°), while the Ångström compensation pyrheliometer has a rectangular aperture with plane angles of 24° and 6° (more recently 10° and 3°), respectively. The Ångström instrument consequently receives in general somewhat more scattered light, and the difference between the readings of the Ångström ($\mathring{A}$) and the silver-disk (S) instruments can be approximately expressed by

$$\mathring{A} - S = -\Delta + f(q, \beta, a, m), \qquad (7)$$

where Δ is the difference which is to be expected at the limit of the atmosphere where no diffuse radiation is added to the direct radiation, and $f(q, \beta, a, m)$ is a function of the quantities q, etc. (notation as given earlier). The function $f(q, \beta, a, m)$ is equal to zero when q equals unity and β vanishes. According to experience recently acquired, this function seems to vary, at stations near sea level, between about 0.01 cal cm^{-2} min^{-1} for low values of β to about 0.02 cal cm^{-2} min^{-1} for high values of β. The difference $S - \mathring{A}$ therefore seems to vary between about 1.5 and 4.0 per cent, where the upper and lower limits correspond to extreme conditions (Ångström—uncorrected for edge effect; silver-disk—Smithsonian scale, 1934).

With due regard to recent investigations, we may, chiefly according to Mörikofer [7], give the following survey on the status of absolute actinometry. Setting the indication of the Ångström pyrheliometer equal to 100 per cent, we have:

Instrument	*Per cent*
Ångström (uncorrected)	100.0
Ångström (corrected for edge effect)	101.3 ± 0.5
Smithsonian scale, 1913	103.5
Solar Radiation Commissions (preliminary results) 1936 (see [3])	101.0
Guild, Physical Laboratory, 1934	101.3
Abbot (Smithsonian scale, 1934)	101.2

Actinometric Equipment at Meteorological Stations. It would be futile to attempt actinometric measurements, even at selected meteorological stations, for the purpose of computing the solar constant, and still more so for determining its possible variations. The required elaborateness and quality of the actinometric equipment are such that measurements for this purpose must be reserved for observatories with very high standards, at locations carefully selected with regard to atmospheric conditions.

On the other hand, the chief problems which actinometric observations from a meteorological network might help to solve are those which we have briefly discussed in the first section of this article. Thus, the measurements should (1) furnish data for an evaluation of the heat exchange at the surface of the earth and enable us to treat, on this basis, the fundamental problem of the energy transport within the atmosphere; (2) provide the means for a general optical analysis of absorption and scattering in the atmosphere; and (3) furnish the basis for correlation studies of the relationship between the radiation processes at the earth's surface and a number of biological phenomena such as the growth of plants, crop conditions, the health of man, and the incidence and spread of certain diseases.

It is clear from what has been said above that the accuracy required of our actinometric equipment for these purposes is much less than that for the determination of the solar constant. The accuracy and ease of operation of our present instruments are undoubtedly satisfactory for these general purposes.

The details of the organization must be based on a realization of the main problems, on experience or knowledge, otherwise gained, concerning actinometric instruments and their qualities, and finally on some acquaintance with meteorological observations in gen-

eral and with what can reasonably be demanded from the observers. On the basis of such considerations the following general pattern for the actinometric stations within a meteorological network is recommended. The proposed organization is based partly on the author's own experience, but is, to a very large extent, the result of discussions with colleagues.

Central Actinometric Observatory (CAO). It seems desirable that at least one such observatory should be established within every large country. Among its main objectives would be scientific investigations of actinometric problems of a geophysical character, or those with geophysical applications. A plan for such investigations will not be discussed here, since it must be based chiefly on the personal initiative and scientific ideas of the observatory personnel. Among the problems deserving general attention are especially the distribution of ozone and its determination by actinometric methods and also special investigations of ultraviolet radiation. The former problem is at present being vigorously pursued by G. M. B. Dobson at Oxford, and it seems probable that in the near future his investigations will result in the construction of apparatus which can be generally accepted for ozone measurements. The work at the Bureau of Standards by W. Coblentz and later by R. Stair and his collaborators will probably lead to similar results with respect to the measurement of ultraviolet radiation.

First we shall discuss the relation which should exist between the activities of the central actinometric observatory and the actinometric stations of the meteorological network, as they will be described below. The CAO must form a center for *calibration*, *standardization*, and *checking* of secondary instruments, and for this purpose such a central observatory should be equipped with absolute instruments. The actinometer perhaps most widely used for measuring direct solar radiation, the Michelson actinometer, is a secondary instrument and, because of a certain delicacy in its construction, it seems desirable that it be compared at regular intervals with an absolute instrument or at least with other instruments of greater stability. The silver-disk pyrheliometer of the Smithsonian Institution is also a secondary instrument. Its construction is much less delicate than that of the Michelson actinometer, and in several instances the instrument has retained its calibration unchanged during rather long periods. However, as a precaution against error, if the instrument is to be used daily, it should be compared from time to time with an absolute instrument or a secondary standard, preferably at least once a year.

The Ångström compensation pyrheliometer is strictly an absolute instrument. Its "constant," which is applied to the direct reading in order to yield the corresponding radiation values, can be determined through simple physical measurements of resistance, dimensions, etc., on the instrument. But in practice such measurements are very seldom made, and they can hardly be carried out without rather elaborate physical equipment. Therefore, the instrument is actually used at most stations simply as a secondary instrument relying upon the correctness of a standardization carried out prior to delivery of the instrument. It is clear, especially since the construction is rather delicate, that reliance on a single calibration involves the serious risks of collecting observational data of greatly reduced value. For this reason, a standardization, either through comparison with a central standard or through a physical determination of the constants of the compensation pyrheliometer, should be carried out at regular intervals. A central actinometric observatory would be the proper place for that purpose.

Second Order Actinometric Stations. Provided that absolute standardizations are carried out at a central observatory, second order stations do not need to be provided with absolute instruments. On the other hand, it seems advisable that they be furnished with double sets of actinometers for measuring direct solar radiation, in order that one instrument can, to some extent, serve as a check on the other. The following measurements are proposed.

1. Measurements of the direct total solar radiation at least three times a day when the sky is clear.

Instruments: Smithsonian silver-disk pyrheliometer, Ångström compensation pyrheliometer, or an instrument of the Michelson type. These are the instruments which have been most carefully examined with respect to various errors. However, other instruments may also be used, provided they are carefully standardized.

2. Measurements of direct solar radiation within special regions of the spectrum with the aid of colored glass filters of the type RG 2 or OG 1, through which the infrared radiation may be separated from the visible and ultraviolet (at least three times a day with a clear sky).

Instruments: Same as under 1.

3. Continuous measurements of the total incoming radiation from the sun and sky.

Instruments: Moll-Gorczynski actinograph with recording galvanometer, Kimball-Eppley actinograph, Ångström pyranometer with photographic recorder, Robitzsch actinograph, or instruments of similar construction.

Screening these instruments from direct solar radiation and comparing the reduction of their reading with the direct solar radiation, measured simultaneously, provides a check on these recording instruments. With this method, the first three instruments mentioned above will easily yield results accurate to about ± 5 per cent for individual days and ± 3 per cent for longer periods.

With regard to the Robitzsch actinograph, the construction of which is similar to a common thermograph with bimetallic elements, the following remarks may be made. The instrumental "constant" (the factor by which the deflection on the recording drum must be multiplied in order to give radiation values) is highly dependent on temperature. A temperature rise of one centigrade degree generally corresponds to an increase in the constant of about 1 per cent. Furthermore, the sensitivity is to some extent dependent on the elevation of the sun as well as on the orientation of the instrument

relative to the sun. Most important in the use of the instrument is a careful examination of its dependence on temperature. Under all circumstances, rather frequent comparisons with actinometers for solar radiation measurements must be made. It is difficult to avoid errors of less than about ±10 per cent for individual days and less than ±5 per cent in monthly means.

4. Measurements of the outgoing "effective" radiation.

Instruments: Ångström pyrgeometer with electrical compensation. Some of the errors inherent in this instrument are the same as for the compensation pyrheliometer. A variable edge effect introduces errors up to about ±3 per cent in single measurements. No recording device of suitable design can yet be recommended for general use. Some theoretical investigations by Prohaska and Wierzejewski [9] on the Bellani instrument for measuring incoming radiation seem to support the view that the Ångström "tulipan" instrument founded on a similar principle, that is, overdistillation of ether under the influence of outgoing heat radiation, may adequately serve its purpose of performing time integration of "effective" radiation. The necessary precautions seem, however, to be too numerous to allow a more general use.

5. Records of hours of sunshine. It is recommended that the duration of sunshine be recorded at all stations where the total radiation from sun and sky is recorded. Special studies of the relationship between the duration of sunshine and the incoming radiation at various stations will then provide a possibility of computing the incoming radiation from the hours of sunshine Q_s according to some formula, for instance, $Q_s = Q_0[\alpha + (1 - \alpha)S]$, where the constants Q_0 and α must be determined. Such studies may give us a means of interpolating radiation values for localities between actinometric stations from the number of hours of sunshine recorded at a much larger number of places.

Instruments: Most widely used are the sunshine recorders of Campbell-Stokes and the recording black-bulb thermometer used by the United States Weather Bureau. It should be emphasized that these instruments can under no circumstances be considered as instruments of precision. It is more important to correlate the data from the various types of sunshine recorders with the radiation balance than to attempt to standardize or correct the instruments so that a precise measurement of the loosely defined "hours of sunshine" may be obtained. With this guiding principle in mind, one may allow rather wide variations in the type of instrument.

Third Order Actinometric Stations. These stations should be equipped with sunshine recorders as well as with simple recording instruments such as the Moll-Gorczynski, Kimball-Eppley, or the Robitzsch type. The pyranometers should, however, be checked at regular intervals by comparison with secondary standard actinometers, either brought to the stations during inspection or kept at the station for use on special occasions. The frequency of these check measurements must, to some extent, be chosen on the basis of experience concerning the constancy of the recording device, which probably depends on the climate.

Regular Meteorological Stations Equipped with Sunshine Recorders. In general, it seems advisable that all actinometric stations of the second and third orders should also be meteorological stations of at least the second order. This will facilitate studies of the relation between radiation and other meteorological elements.

A rather wide network of stations recording simply sunshine duration is desirable so that they may act as interpolation stations with respect to the radiation balance. However, it is also recommended that other meteorological observations be made at such stations, at least observations of cloudiness, temperature, and precipitation.

CONCLUSIONS

Critique of Routine Actinometric Measurements as Hitherto Organized. It is generally realized that actinometric measurements are still in a state in which an organization according to systematic and generally accepted rules is lacking. The observations are, as a rule, limited to certain single observatories, and there is, in general, no possibility for a synoptic treatment of the results. Only very seldom are the observations made in a manner which permits separation of scattering and absorption in the atmosphere.

The recording devices, such as the Moll-Gorczynski instrument or the Robitzsch instrument, when used at field stations, are in many cases too infrequently checked, a fact which especially for the latter instrument is rather fatal, since its indications are highly dependent on the manner of exposure, and its constant is highly variable with temperature. Many observational data of no value have been collected in this way.

Measurements of the effective radiation, important as they may seem, are almost totally lacking, with the exception of the results from only a few observatories. Regular checks of the instruments in use are seldom made.

It is my opinion that our knowledge and experience of the instruments available are such, after the work on the subject particularly by the Smithsonian Institution and the International Radiation Commission in close collaboration with the Observatory of Davos, that the time is ripe for a systematic organization of an international actinometric station network within the meteorological organization. The technical and instrumental problems which still need to be solved and the improvements to be expected need not delay an organization which now seems highly desirable.

Special Problems. We have already briefly indicated the problems which such an actinometric network would primarily help to solve. They include the important problem concerning the energy exchange within the earth's atmosphere and the factors influencing it. If we consider a given vertical column extending from the

earth's surface to the upper limit of the atmosphere, the temperature of the air within this column is determined primarily by the incoming and outgoing radiation, by the transfer of energy through horizontal advection, and, to some extent, by evaporation and condensation processes. If we take a column of a sufficiently large cross section and consider a sufficiently long time interval, our problem of analyzing the factors influencing the temperature coincides with the problem of finding the causes for climatological temperature changes in general. Here, incoming and outgoing radiation are the most fundamental elements. Without a thorough knowledge of these factors, their distribution and variations, all speculations on climatic variations are reduced to rather vague guesses.

Another equally important problem, closely connected with the climatic variations, concerns the transmission of the atmosphere and its fluctuations. A clear

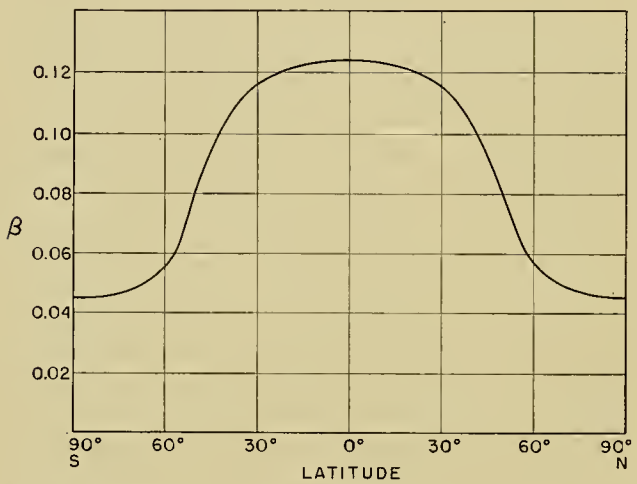


Fig. 2.—Preliminary curve showing variation of scattering coefficient β with latitude.

separation between scattering and absorption is necessary. The simplest way of accomplishing this is indicated above. If we take the coefficient β as a measure of the scattering of the atmosphere, the following remarks may be made. A rather summary treatment of available pyrheliometric data already shows that β is on the whole much larger in tropical and subtropical regions than at higher latitudes. The graphical representation in Fig. 2, taken from a previous paper by the author [2], shows this, but it may be taken only as a very rough indication of the general conditions. More extensive observational data, systematically treated, must be available before the distribution of β over the whole globe can be mapped. Intensive investigations may here solve the problem concerning the dust-producing centers at the earth's surface, the nature of the scattering particles, and their origin under various conditions. Investigations near desert regions seem to show that the scattering particles directly produced by storms over the desert are much larger than those which generally occur in the atmosphere. At all Northern Hemisphere stations investigated for an annual variation of β, the scattering coefficient reaches a pronounced maximum in the spring (April or May). The values of β obtained at Spitsbergen by Olsson [8] and at Abisko (northern Lapland) by Tryselius [11] are so low that the scattering there must be almost totally due to the molecules. Therefore, if we consider only the effect of scattering, the ultraviolet radiation at, for instance, 30° solar elevation in Lapland, must be almost as intense as at about 2000-m height in Switzerland for the same solar elevation. Only a few of the problems related to a more detailed synoptic investigation of scattering in the atmosphere have been indicated here.

The outgoing "effective radiation" should be the object of a closer investigation especially with respect to its distribution over the earth's surface. In general, it seems to be very closely related to two factors: (1) the temperature at the place of observation and the temperature distribution in the atmosphere, and (2) the content of water vapor in the atmosphere. This relationship is so close that it is doubtful whether there are any other factors whose effects do not fall within the errors of observation. On the other hand, there are some indications of a variation during the night, which might not be explained by a variation of the previously mentioned elements. Here is an important field of research from which perhaps some factor influencing the climatic variations may be discovered.

Finally, for the important studies of the relation between biological phenomena and radiation, an actinometric network has an essential task to fulfill. In all these studies in which special portions of the spectrum must be considered effective, a clear separation between scattering and absorption is necessary. When we know the scattering coefficient and its variations, we will be able to give a much more detailed picture of the variations in the different spectral regions of the sun's radiation. Such a detailed knowledge is probably necessary before the relation between biological phenomena and solar radiation can be investigated with success.

REFERENCES

1. Ångström, A., "On the Atmospheric Transmission of Sun Radiation and on Dust in the Air." *Geogr. Ann., Stockh.*, 11:156–166 (1929).
2. —— "On the Atmospheric Transmission of Sun Radiation, II." *Ibid.*, 12:130–159 (1930).
3. —— "Survey of the Activities of the Radiation Commissions of the International Meteorological Organisation and of the International Union of Geodesy and Geophysics." *Tellus*, Vol. 1, No. 3, pp. 65–71 (1949).
4. —— "Atmospheric Circulation, Climatic Variations and Continentality of Climate" in *Glaciers and Climate* (Geophysical and geomorphological essays dedicated to Hans W:son Ahlmann). *Geogr. Ann., Stockh.*, Vol. 31, Nos. 1 and 2, pp. 316–320 (1949).
5. Courvoisier, P., und Wierzejewski, H., "Beiträge zur Strahlungsmessmethodik. I—Die physikalischen Grundlagen der kalorischen Strahlungsmessmethoden." *Arch. Meteor. Geophys. Biokl.*, (B) 1:45–53 (1948). "II—Die

Berechnung der Wirkungsweise kalorischer Strahlungsmessinstrumente." *Ibid.*, (B) 1:156–199 (1949).

6. Hoelper, O., "Der atmosphärische Trübungszustand über Aachen nach kalorischen Strahlungsmessungen im Polarjahr." *Dtsch. meteor. Jb. Aachen* (1933). Aachen, 1935.
7. Mörikofer, W., "Meteorologische Strahlungsmessmethoden." *Handb. biol. ArbMeth.*, E. Aberhalden, Hsgbr., 2.4005–4245 (1939).
8. Olsson, H., "Meteorological Observations at Mt. Nordenskiöld, Spitzbergen, during the International Polar Year, 1932–33." *Medd. meteor.-hydr. Anst., Stockh.*, Vol. 6, No. 5, 83 pp. (1936).
9. Prohaska, F., et Wierzejewski, H., "Théorie et pratique du pyranomètre sphérique de Bellani." *Ann. Géophys.*, 3:184–221 (1947).
10. Schmolinsky, F., "Die Wellenlängenabhängigkeit der Sichtweite und des Koeffizienten der Dunstextinktion." *Meteor. Z.*, 61:199–203 (1944).
11. Tryselius, O., "On the Turbidity of Polar Air." *Medd. meteor.-hydr. Anst., Stockh.*, Vol. 5, No. 7, 10 pp. (1936).

METEOROLOGICAL OPTICS

General Meteorological Optics *by Hans Neuberger* 61

Polarization of Skylight *by Zdeněk Sekera* 79

Visibility in Meteorology *by W. E. Knowles Middleton* 91

GENERAL METEOROLOGICAL OPTICS

By HANS NEUBERGER

The Pennsylvania State College

The phenomena of meteorological optics are not as generally well known as those of other branches of meteorology. For this reason, the subsequent sections are introduced with a brief description of the major phenomena and a summary of known facts. For abnormal variations the reader is referred to the multitude of meteorological, astronomical, and general science periodicals. Visible sound waves in the sky, frequently observed during World War II,[1] could not be considered here.

In order to keep the number of references within space limits, authors whose contributions have been discussed in standard works are, where possible without ambiguity, cited from these works; this is indicated by a "c" preceding the respective reference number.

SUBJECTIVE PHENOMENA IN THE ATMOSPHERE

The Apparent Shape of the Sky. When scanning the daytime sky, an observer perceives a more or less flattened vault. For some observers, this impression persists even at night, although the multitude of stars in their different magnitudes tends to convey the idea of a rather undefinable space. The variety of impressions is the result of a complex psychometric coordination of the visual and the physical space. This coordination is not a simple transformation of a geometric reality such as obtains for an overcast sky, because, in the case of the clear daytime sky, we look into an indefinite depth of the luminous atmosphere. Indeed, if we fix our sight in any elevated direction, we fail to perceive a "surface" on which our eyes may rest. However, if we let our eyes wander between zenith and horizon, the impression of such a surface is inescapable. The problem of the sky shape, although some of its aspects lie within the realm of psychophysics, is of meteorological interest because of its bearing on the practice of estimating cloudiness.

Smith [c. 42] introduced a method by which a numerical value can be assigned to the apparent flatness of the sky, by measuring the elevation angle α (Fig. 1) of the estimated position of point M that bisects the imaginary arc ZH from zenith to horizon. The *half-arc angle* α is then a measure of the flatness of the sky, becoming smaller as the sky appears flatter. Another method for expressing the sky flatness numerically is the estimation of the ratio of the apparent distances OH/OZ, which increases with increasing flatness of the sky. The estimation of this ratio is a more difficult task than bisecting the arc ZH and is a much coarser measure. Assuming, for example, a circular sky profile, we can compute corresponding values of OH/OZ and α [c. 42]; $OH/OZ = 1, 2, 3, 4$, corresponds to $\alpha = 45°, 33°, 25°, 20°$, respectively. Since most observations fall in the range $20° < \alpha < 40°$ and are reproducible within 1°, while OH/OZ can be estimated, at best, only in whole numbers, OH/OZ is obviously an inadequate measure.

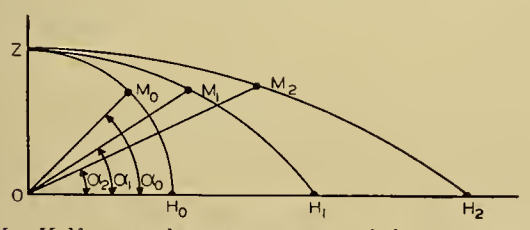


Fig. 1.—Half-arc angle α as a measure of the apparent shape of the sky.

Various authors have deduced the curvature of the sky profile as circular, elliptic, parabolic, hyperbolic, or helmet-shaped [8, 9, 22, 23, c. 42]. The angle at which the sky appears to meet the horizon plane is one criterion of the geometric form. This angle would be 90° only in the cases of elliptic and parabolic shapes, but more or less acute in all others. A parabolic profile would also be distinguished by the pointed appearance of the zenith sky. Another criterion is derived from comparison of observed overestimation of elevation angles of objects with those computed under the assumption of various sky shapes [42]. Variations of a psychological nature among different observers and physical variations of sky conditions undoubtedly preclude the assumption of any unique sky shape. For the discussion of observational results, consideration of the half-arc angle α may suffice.

Table I shows the effect of general sky brightness and cloudiness on the depression of the sky as seen by various observers. Although the large differences among observers reflect the strong subjectivity of the phenomenon, the same trend appears in the effect of the various sky conditions. The cloudy daytime sky looks flattest, the moonless night sky most arched. Some of the individual differences may be due to effects of locality, since the observations cited were made at different places and altitudes.

Table I. Average Half-Arc Angles for Various Sky Conditions

Observer	Daytime		Twilight clear	Clear night	
	cloudy	clear		moonlit	moonless
Dember and Uibe [8]	29.0°	32.0°	32.2°	36.7°	40.1°
Miller [37]	29.9	34.0	—	—	—
Reimann [c. 42]	21.0	22.5	—	26.6	30.0
Mendelssohn and Dember [33]	—	31.9	—	36.6	—

1. For example, see W. E. K. Middleton, *J. R. astr. Soc. Can.*, 38: 432 (1944).

The effect of the amount of cloudiness is demonstrated in Table II by the most recent observations [37]. In qualitative agreement with observations by Reimann [c. 42], these data show that even a few clouds in the sky materially increase its apparent flatness, whereas with sky covers of $> \frac{5}{10}$, the observer refers chiefly

Table II. Average Half-Arc Angle for Various Steps of Cloudiness

Clouds *tenths*	0	<1	1–3	4–7	8–9	>9
α (°)	34.0	32.6	31.5	30.6	30.2	29.9

to the cloud layer in forming his impression of the sky shape. Therefore, for cloudiness $> \frac{4}{10}$, the increase in flatness becomes practically negligible.

Miller [37, 38] seems to be the only observer to investigate the effect of cloud types (and ceiling height) on the apparent sky shape. Table III shows his results for various cloud types arranged in order of increasing average cloudiness for these types. The groups *Ac* and *As*,[2] and *Sc* make the sky appear flatter than would result from the implicit cloudiness effect. This has been attributed to the structure of the underside of *Ac* and *Sc*, which facilitates depth perception and makes the observer aware of the extension of these cloud layers beyond the terrestrial horizon. This apparent increase of the horizontal extent of the cloud layers causes a decrease in α.

Table III. Relationship between Cloud Type and Half-Arc Angle

Cloud type	Average cloudiness, *tenths*	Average α(°)
Cu, *Fc*	4	30.0
Ci, *Cs*	5	29.1
Ac, *As*	7	27.8
Sc	8	28.4
St	10	29.0

From Table III it is also obvious that the cloud height does not influence the apparent flatness of the sky, in the manner which might be expected from

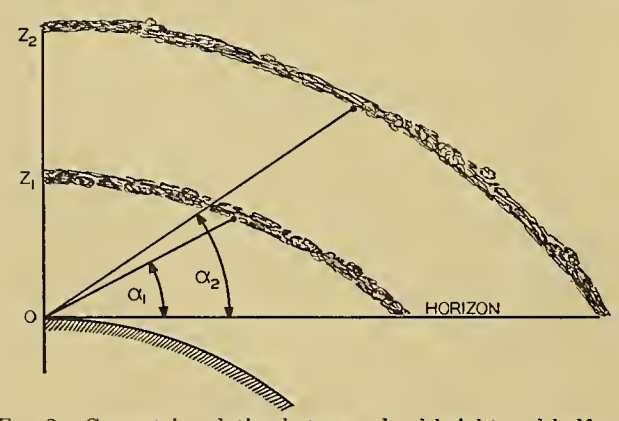


Fig. 2.—Geometric relation between cloud height and half-arc angle.

purely geometric considerations (Fig. 2). The *Ac*- and *As*-group, for example, is associated with a smaller α than is the *St*-group, whereas the reverse should be true according to Fig. 2 if the visual space were a simple transformation of the physical space. A comparison between Table III and the last three columns of Table II shows that the type of clouds has a greater effect than the amount of clouds.

In contrast to Dember and Uibe [9], Miller [37] found that the visual range has only a minor influence on the apparent shape of the sky, whereas the effect of the distance to the terrestrial horizon is very strong, as shown in Table IV. The differences in α between various

Table IV. Average Effects of Visual Range (V) and Horizon Distance (D) on the Half-Arc Angle (α)

V (km)	$D = 0.4$ km α(°)	$D = 12$ km α(°)
<6	33.3	30.7
6–10	32.1	28.8
>10	32.6	28.9
Mean	32.6	29.2

groups of visual range are considerably smaller than between different horizon distances. This result was confirmed by measurements of α at various distances from a mountain range [38]. In addition to the actual distance of the horizon, the facilities for subconscious estimation of this distance, such as is offered by suitable foreground configuration, have also proven of strong influence on the perceived sky shape [23, 37, 38]. For this reason, the sky dome cannot generally be considered a rotational geometric figure.

Related Phenomena. Closely related to the shape of the sky is the well-known enlargement of sun or moon when near the horizon [23, 42]. As can be seen from the shaded angles in Fig. 3, the same angular subtense intersects a greater portion of the flattened sky near the

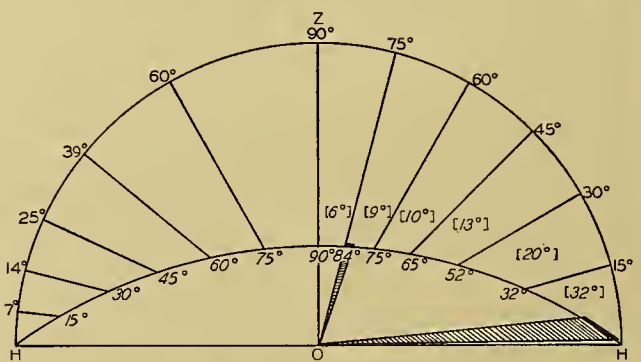


Fig. 3.—Relationship between true and apparent (slanted numbers) elevation angles and angles of subtense (bracketed numbers).

horizon than at higher elevations, and this portion becomes greater as the sky becomes flatter. The same holds true for the angular distance between stars. Jeffreys [25] raised an objection to this projection theory on the grounds that the sun or moon should appear elliptical with a long vertical axis, because only this axis would be distorted by the shape of the sky. However, the angular subtense is too small to permit detection of such a deformation; moreover, this effect is probably compensated by an opposite effect of the

2. *Ac* and *As* occurred simultaneously in almost all cases.

physical distortion owing to astronomical refraction. The size variation can also be explained by the general overestimation of angular elevations of terrestrial and celestial objects. On the left side of Fig. 3 the flat arc of the sky is divided into six equal segments corresponding to 15°-intervals of estimated elevations (slant numbers). The true elevations show by comparison that the relative overestimation is > 100 per cent for small elevation angles and decreases to zero at 90°. The absolute overestimation reaches a maximum at true elevations between 30° and 40° [23, 42]. The right side of Fig. 3, showing true 15°-intervals and their estimated equivalents (in brackets), indicates that a given angular subtense is overestimated when below roughly 30° elevation, underestimated when above 30°. Accordingly, the ½° - subtense of sun or moon appears larger near the horizon, smaller at higher elevations.

The overestimated height and steepness of mountains [22, 42] and the apparent ellipticity of circular halos [23, 34] are also related to this phenomenon, as is the incorrect estimation of the amount of clouds. The latter is of practical significance, because an observer tends to underestimate the amount of clouds overhead and to overestimate the amount of clouds near the horizon. This fact is qualitatively known, and in the observer's manual, measurements of angular elevations of clouds are suggested to eliminate this error for advancing or receding cloud layers and those surrounding the station. However, no such expedient is available for estimations of the more frequent nonuniform states of sky. This problem can be summarized as follows: The estimates of cloud covers of 0 and 10 tenths are usually correct. For all other cloud amounts, the error made by the observer is a function of subjective factors (experience, etc.), of the amount and type of the clouds, the distance of the terrestrial horizon, and the general brightness level (Tables I–IV).

Theories and Problems. All attempts to formulate an all-inclusive theory of the apparent shape of the sky have thus far been unsuccessful, chiefly because of the simultaneous involvement of physical and psychophysiological factors. Although physical and geometrical variables modify our impression of the sky shape, they do not suffice, in themselves, to explain the observed facts. The simple consideration of the geometric properties of a cloud layer at 2000 m, for example, would result in an $\alpha = 1°$, whereas α is actually observed between 20° and 30°, that is, the sky does not appear as flat as it should. A similar discrepancy arises for the clear sky [38]. From a physical standpoint, Dember and Uibe [9] attribute the sky shape to the distribution of sky brightness. They assume that the maximum distance from which scattered light reaches the observer is proportional to the square root of brightness. However, their theory not only implies that the brighter objects appear to be farther away, which is not true [42] (*e.g.*, brighter stars appear nearer), but also precludes any variations of the sky shape with the distance of the terrestrial horizon. Similarly, Humphreys [21] believes that the greater haziness near the horizon produces the impression of greater distance than is the case at more elevated regions of the sky. While this effect is undoubtedly present, its magnitude has been shown in Table IV to be of secondary order only. Purely psychophysiological explanations are similarly unsatisfactory. Gauss and others [c. 42] were of the opinion that, because our line of sight is habitually horizontal and normal to the body axis, the illusion of a variable moon size should disappear when the direction of sight is changed by mirrors or the body orientation altered. Not all results of pertinent experiments supported this theory. Also, the effects of physical variables would remain unexplained. Various attempts at a combination of geometric and psychophysical explanations have also been made. For example, v. Sterneck based his transformation of physical into visual space on the underestimation of distances according to a hyperbolic function, while Witte assumed that a straight line in physical space also appears rectilinear in visual space [c. 42]. According to Exner [42], the moon's visible area is compared to that of the fovea when the moon is high, but at low moon the respective linear dimensions are compared because of the attention commanded by the horizon line. This hypothesis, however, is incapable of explaining the sky shape or the variations in the moon's apparent size caused by external physical factors. Isimaru [23] combined, in his theory of the shape of the cloudy sky, the underestimation of distances with the overestimation of elevation angles. The same idea was followed by Hüttenhain [22] who reverted to v. Sterneck's theory. Both authors [22, 23] make the implicit assumption that the overestimation of angular elevations is independent of the sky shape, because it can also be observed in rooms [22]. However, since both phenomena are due to the properties of our visual space, they cannot be independent.

Fundamentally, all of the theoretical approaches to the general problem have hitherto been based on the assumption that the visual space is Euclidean and can be obtained by some unique transformation of the three-dimensional manifold of the physical space. However, Luneburg [30] has recently shown that the sensations produced by binocular vision represent a Riemannian manifold. From the analysis of certain optical illusions he concluded that the geometry of our visual space is the hyperbolic geometry of Lobachevski as had already been indicated by Škreb [49]. With this theory a differential equation for the apparent size of a line element in the horizontal plane was developed, which is, however, not immediately applicable to the problem of the apparent shape of the sky; at any rate, an extension of this theory to include this group of phenomena appears desirable. The major physical factors whose effects must be explained by such a theory may be tabulated as follows:

1. General brightness of the visual field and brightness distribution over the sky.
2. Cloud types and amounts.
3. Distance of terrestrial horizon, and facilities for estimating this distance.

In addition, there are several phases of the problem that have been insufficiently explored or are entirely

unknown as yet. For example, the practical problem of the mutual influence of sky shape and cloudiness estimation needs further exploration. Simultaneous photographic records of the sky, estimation of cloudiness, and determination of the shape of the sky (under all possible states of sky) may furnish an individual correction factor for adjustment of cloudiness estimates. In particular, the quality of cloudiness estimations for individual cloud layers in the presence of other cloud decks needs special attention, as does the effect of the configuration of the terrestrial horizon on such estimations. More observations are also needed on the following effects: possible seasonal variations; the terrestrial horizon distance in conjunction with the configuration of the visual field between horizon and observer; the measured brightness level and brightness distribution over the sky.

Wholly unexplored is the possible effect of an observer's altitude above the earth's surface on his impression of the sky shape and consequently the accuracy of his cloudiness estimation; this phase is especially of interest for meteorological observers on mountains or in airplanes. In this connection, the apparent shapes of the terrestrial surface and of cloud layers, as seen from above, deserve attention, because the quality of estimations of cloudiness below an aerial observer hinges on this problem.

REFRACTION PHENOMENA IN THE CLOUD-FREE ATMOSPHERE

In the subsequent discussions reference is made only to the visible portion of the electromagnetic spectrum, although analogous phenomena occur with other wave lengths. The refractive index n_λ of air for various wave lengths and its dependence on the air density is well known from laboratory determinations [31, 42]. An example of the magnitude of the dispersion and temperature effect on the refractive index in air at normal pressure (760 mm *Hg*) is given in Table V. Obviously,

Table V. Variation of $(n_\lambda - 1)10^6$ with Wave Length and Temperature
(*After Meggers and Peters* [31])

λ(A)	0C	15C	30C
3983	297	282	268
5569	293	278	263
7664	290	275	261

the effect of dispersion is small in comparison with that of temperature. The refractive indices of the various gaseous constituents of air differ even more significantly from each other, but since the composition of air varies only slightly, this effect is negligible for all practical purposes.

Phenomena due to Average Density Gradient. When penetrating the atmosphere, which is assumed to consist of concentric, equidistant isopycnic surfaces, an extraterrestrial light ray describes a curve whose well-known equation is

$$nr \sin i = k = \text{const}, \qquad (1)$$

where r is the distance of an isopycnic surface from the earth's center, i the angle of incidence, and the refractive index n (for white light) is a function of r, $n(r)$. In Fig. 4, $r_o = CO$ is the earth's radius; an observer at O sees a light ray L, that enters the atmosphere at P with the angle of incidence i, at the apparent zenith distance θ_o instead of the true zenith distance θ. The difference $\theta - \theta_o = R$, the *astronomical refraction*. In its general form, the equation of the ray curve PO in polar coordinates is

$$\varphi = \int_{r_o}^{r} \left(\frac{n^2 r^2}{k^2} - 1 \right)^{-\frac{1}{2}} \frac{dr}{r}, \qquad (2)$$

or expressed in terms of astronomical refraction:

$$R = \int_{n_H}^{n} \frac{n_o r_o \sin \theta_o}{n \sqrt{n^2 r^2 - n_o^2 r_o^2 \sin^2 \theta_o}} \, dn. \qquad (3)$$

The integrals, to be taken between the upper limit of the atmosphere (where $n = n_H$) and the earth's surface

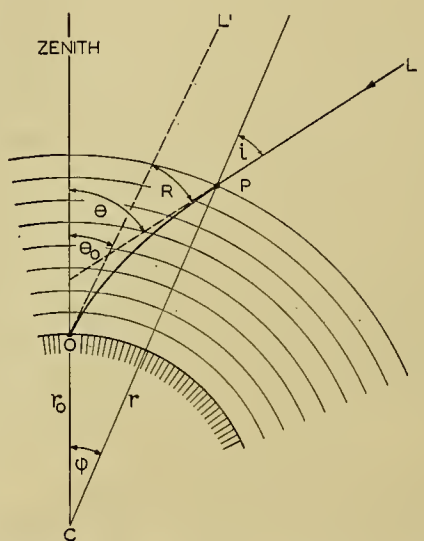


Fig. 4.—Schematic diagram of astronomical refraction.

(index o), have no general solutions, because the change of n with altitude must first be known.

Various authors made different assumptions regarding the function $n(r)$ and computed the astronomical refraction for various apparent zenith distances. Wünschmann [59], who compared several of the results, prefers Gylden's refraction data, because they are based on a hypothesis which agrees closely with data obtained from aerological ascents. Recently Link and Sekera [28] computed tables of various characteristics of the ray path for zenith distances from 75° to 90° and altitudes up to 60 km. They considered the vertical density distribution separately for summer and winter, using average density conditions as revealed by balloon soundings, sound propagation, and twilight phenomena. Correction tables for the effect of deviations from normal in local conditions of pressure, temperature, and humidity have also been variously computed [31, 42]. These corrections are important chiefly for apparent zenith distances up to 70°, for which the values of R

are sensibly independent of the assumption regarding the vertical density distribution and the concentricity of isopycnic surfaces [28, 42, 59]. Court[3] pointed out the inadequacy of these correction tables particularly for cold climates; he suggests the use, instead of air temperature, of "refractive temperature," *i.e.*, that temperature for which the correction obtained from tables is the one actually required, and he presents rules for estimating these refractive temperatures.

For zenith distances > 70°, Esclangon [12] attributes the greatest optical effect to the layer up to about 15 km, whereas according to Wünschmann [59] the mass distribution in the stratosphere between 15 and 45 km height and the orientation of the isopycnic surfaces in this layer play a dominant role. More frequent and systematic soundings of the upper atmosphere with modern equipment would enable us to solve this problem and to obtain more direct values of stratospheric density variatio ıs.

Wünschmann, by means of upper-air soundings, constructed charts showing the topography of the optical surfaces. He found that the influence of density variations in the lower troposphere up to 6.5 km is generally compensated by an opposite influence in the upper tropospheric region between 6.5 and 15 km, leaving an insignificant net effect of the troposphere. The inclinations of isopycnic surfaces owing to local horizontal temperature gradients, fronts, or orographic features, generally change the astronomical refraction by only a fraction of a second of arc.

The large astronomical refraction for zenith distances of 90° or more generally lengthens the day, since it causes the sun to rise somewhat earlier and set somewhat later than is computed from purely geometrical considerations. This is of practical importance for the prediction of illumination conditions in polar regions, where the strong density gradient, built up in the lower atmosphere during the winter night, may advance the date of sunrise by several days [42]. The great decrease in normal refraction with slight elevation above the horizon causes a deformation of the sun's or moon's disk; for the difference in refraction between the lower limb, touching the horizon, and the upper limb amounts to 6′, so that the vertical axis of the disk appears shorter by 1/5. The large refraction is connected with a relatively large prismatic dispersion which is of the order of 20 to 40 seconds of arc between blue and red wave lengths [19, 42]. This dispersion occasionally causes the last segment of the setting sun or planets [54] to appear green for a few seconds, the so-called *green flash* [21] or *green segment* [19]. Sometimes the color is blue or it changes continuously from yellow to violet. This phenomenon can occur only when the atmosphere is so clear that the shorter wave lengths are not attenuated. Whereas Hulburt [19] believes that normal dispersion is sufficient to cause this phenomenon, most observations seem to be associated with refractions in excess of the normal [35, 54]. To what extent selective absorption by water vapor [42] or other gases contributes to this phenomenon is still undecided.

The curvature of the rays from artificial lights is due to *terrestrial refraction*. In Fig. 5, an observer at A sees a point B in the direction of incidence AC of the curved light ray; the angle α between the straight line AB and the tangent AC is the terrestrial refraction. Similar conditions obtain for an observer at B sighting point A; in this case the terrestrial refraction is measured by angle β. The sum $\alpha + \beta = \epsilon$ is called the total refraction. For an average density gradient the terrestrial refraction increases roughly from 2″ to 42″ when the distance between the two points increases from 1 to 20 km [42]. The curve of the light ray can, in most cases, be assumed as a circular arc, so that $\alpha = \beta = \epsilon/2$. The path of the light ray is then determined by its radius of curvature r_L. In Fig. 5 the angular distance between points A and B is φ at the earth's center, and ϵ is the central angle of the arc AB. Since the heights of A and B above the earth's surface are small as compared to the

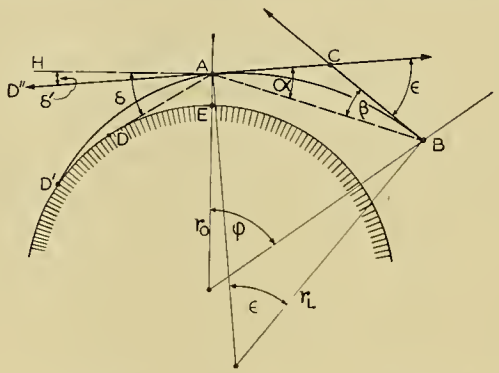


Fig. 5.—Schematic diagram of terrestrial refraction.

earth's radius r_o, and the angles ϵ and φ are small, the ratio

$$\frac{\epsilon}{\varphi} = \frac{r_o}{r_L} \qquad \text{or} \qquad \frac{\epsilon}{2} = \frac{\varphi r_o}{2r_L}. \tag{4}$$

As φ and r_o are known, the terrestrial refraction can be computed from the radius of curvature of the light ray. Its reciprocal, the curvature of the ray, is (according to Wegener [56]) determined by

$$\frac{1}{r_L} = (n_\lambda - 1)\,\frac{p}{T_v^2}\,\frac{273}{760}\,(\gamma' - \gamma), \tag{5}$$

where n_λ is the refractive index, p the barometric pressure in mm Hg, T_v the virtual temperature in °K, γ the temperature lapse rate (counted as negative in case of inversions), and γ' the autoconvective lapse rate. The curvature of the light ray is proportional to the barometric pressure and inversely proportional to the square of the virtual temperature, showing the dominating influence of temperature on refraction. The curvature decreases as the lapse rate increases, and the light ray becomes rectilinear for a homogeneous atmosphere ($\gamma = \gamma'$). For large lapse rates, however, the convective activity causes scintillation. For strong inversions ($\gamma << 0$) the curvature of the light rays approaches

3. See A. Court, "Refractive Temperature," *J. Franklin Inst.*, 247: 583–595 (1949).

that of the earth's surface, which ordinarily is several times larger, and total reflection (mirages) may occur.

The expansion of the horizon and the decrease of its angular depression is a usual consequence of terrestrial refraction. In Fig. 5, an observer at A whose horizontal plane is AH would, in case of rectilinear light rays, see the horizon at D, where his line of sight is tangent to the earth's surface and the geodetic depression is δ. Because of terrestrial refraction he normally sees the horizon at D' in the direction AD'' (tangent to the curved ray AD'), that is, at the depression δ'. The difference $\delta - \delta'$ represents the terrestrial refraction, and equations (4) and (5) apply to this case also. However, δ is generally determined by direct observation rather than computed from meteorological data and the known geometric quantities, because temperature and lapse rate in the air layer below the observer vary with the nature and contour of the underlying surface. The effects of these variables for air layers not in immediate contact with the ground were investigated by Brocks [2, 3, 4] who found that the terrestrial refraction can be quite accurately computed from meteorological data and that, in turn, the average lapse rate can be determined from the observed curvature of light rays. For a ray-path length of 30 km, a change in zenith distance of 1″ corresponds to a lapse-rate change of 0.04C/100 m. By means of mutual sighting from both ends of a ray path the absolute values of the lapse rate can be determined. However, this method is of limited practical value, because for steep lapse rates, with their attendant convection currents, the image fluctuations of the light beam would considerably decrease the accuracy of measurement.

Phenomena Due to Special Density Gradients. When the decrease in density upwards is greater than normal, a condition which may be caused by smaller than normal lapse rates, terrestrial refraction is increased and objects that are usually beyond the horizon come into view. This excessive extension of the normal horizon is called *looming*. The opposite phenomenon of *sinking* is due to an abnormally small vertical density gradient. As can be seen from application of equation (5) to the average conditions between observer and horizon point, the curvature $1/r_L$ of the light ray (AD' in Fig. 5) becomes smaller with increase in lapse rate γ; for $\gamma = \gamma'$, the curvature becomes zero and the observer sees horizon at D; for $\gamma > \gamma'$, his horizon would further shrink and end at a point between D and E, while the curvature becomes negative (ray convex toward surface). In this case it is not necessary that $\gamma > \gamma'$ over the whole range, although this often occurs over strongly heated surfaces; it is sufficient that the isopycnic surfaces be inclined upwards toward the observer, so that the air density is greater there than at the distant point on the horizon [42]. The great increases in, or the reversals of, the normal vertical density gradients are generally confined to the air layers near the ground.

When the light rays from the upper portion of a distant object LU in Fig. 6 (A) have a different curvature than those from the lower portion, the geometric angle s, that the object subtends at the observer O, appears changed. Exner [42] has shown that a linear change in refractive index n with height cannot lead to a noticeable change in the angle of subtense. When the decrease of n with height is slower than it would be according to a linear function as shown by case (A), the curvature of the ray LO is greater than that of UO, and the apparent angle of subtense s' between the tangents to the respective rays becomes smaller. This phenomenon, in which the object also appears elevated, is called *stooping*. An enlargement of s' combined with an apparent lifting takes place, when the decrease of n with height is more rapid than according to a linear function, as shown by case (B) which represents *towering*. Whereas these phenomena result from vertical density gradients (drop in density per unit height) that decrease (A) and increase (B), respectively, with height, case (C) shows a negative density gradient that decreases, and case (D) one that increases with height, causing s' to be enlarged and diminished, respectively.

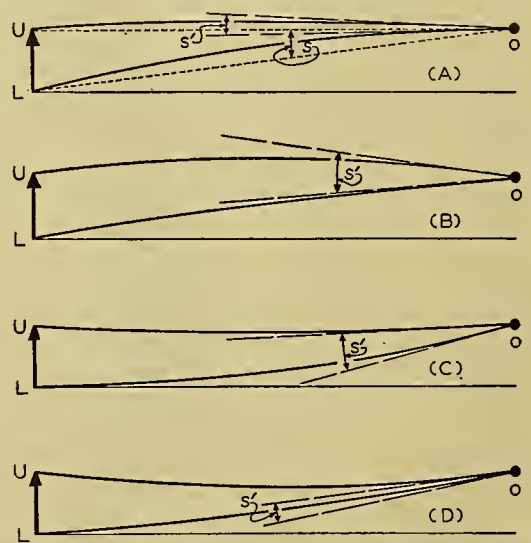


Fig. 6.—Effect of abnormal density gradients on the curvature of light rays.

Because of the increase in density with height, the rays are convex toward the ground, and the objects appear depressed. The mathematical theory for these phenomena, as well as for the corresponding ones involving horizontal objects, was developed by Exner [42].

In case the density distribution in the lower layers is such that the rays from an object reach the observer along two or more different paths, so that he sees one or more images of the object, we speak of *superior*, *inferior*, or *lateral mirages*, depending on whether the image appears above, below, or to the side of the object. The latter case can occur only when the isopycnic surfaces are vertical or nearly so, for example, in proximity to strongly heated walls. This phenomenon was theoretically treated by Hillers [17]. In case of a complicated density distribution in the lower layers, complex distorted images of distant objects, the *Fata Morgana*, may appear. General theories of mirages were developed by Nölke, A. Wegener, and others [c. 42], and

more recently by Fujiwhara and others [c. 21]. There is, however, still a thermodynamic problem connected with inferior mirages such as can be observed over heated highway surfaces. There, the vigorous stirring of the air by passing vehicles apparently has little effect on the existing density distribution. It would be interesting to study the "tenacity" of the mirage-producing air layer. Ives [24] has investigated larger-scale mirages of this nature and found that phenomena caused by steep lapse rates re-form, after disturbance, within a few minutes, while mirages produced by low inversions are more readily disturbed and "heal" much more slowly. Laboratory experiments, which permit control of the variables, and theoretical study of the heat transfer rate seem desirable to expand our knowledge of mirages.

Scintillation. Scintillation is due to temporal and spatial variations of density in the atmosphere. It consists of one or more of the following characteristics, depending on the nature of the object viewed. If the object is a point source, scintillation causes (1) apparent directional vibrations (unsteadiness in position of fixed objects), (2) apparent intensity fluctuations (light source may even appear to flash on and off), or (3) color changes (white light shows alternately its individual chromatic components). For extended objects, inhomogeneities in air density will cause (1) varying distortions of the contours or of internal line-structure of distant objects, thereby producing apparent expansion, contraction, or even disruption of the visible area; or (2) inhomogeneous brightness distribution, so-called *shadow bands*, over the surface of an object that is illuminated by a collimated light beam.

Astronomical scintillation involves extraterrestrial light sources. Its effect, in general, decreases with increase in angular elevation of the source. The amplitude of the vibratory motion of stars (or of the edge of the sun's or moon's disk) amounts to a few seconds of arc at the most, with a frequency of roughly 2 to 30 sec^{-1}, although the vibrations are seldom of truly periodic nature [7, 42, 55]. The relationship between the quality of star images in telescopes and weather elements has long been recognized [42]; the image quality deteriorates as wind speed, turbulence, or temperature lapse rate in the lower atmosphere increase [2, 55]. Respighi [c. 42] ascribed a greater effectiveness to the rotation of the earth than to wind, because he observed spectroscopically that stars on the western horizon pass through the spectral color sequence from red to violet, while those on the eastern horizon show the reverse. Pozdèna [43] shares this opinion, stating that the linear speed of the earth's rotation is much greater than the relative speed of the winds. Exner [42], however, pointed out that the observed phenomenon was due to the prevailing westerlies at higher altitudes and suggested that the argument could be decided by means of observations in regions with prevailing easterly circulation. There, the phenomenon observed by Respighi should appear reversed if wind is the dominant factor. Such a test has apparently not yet been made.

The explanation of chromatic scintillation, which has frequency characteristics similar to those of directional vibrations, was given by Montigny [c. 42] and is briefly outlined (Fig. 7). The difference in refractive index for the extreme visible wave lengths causes a white light ray L_1, entering the atmosphere at A, to send its red component toward O_1 at the ground, its violet component toward an observer at O. Another light ray L_2, which is lower than L_1 by a distance D and enters

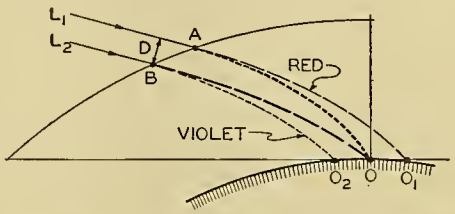


Fig. 7.—Dispersion of light rays by the atmosphere.

the atmosphere at B, sends its red beam toward O, while its violet beam falls at O_2. Other rays between L_1 and L_2 send the intermediate colors toward O, so that the observer there sees the entire color mixture as white, because the angular subtense of the spectrum is generally too small to be resolved by the eye. The distance D between rays of the extreme colors varies with the angular elevation of the light source and the height above the earth's surface. Table VI represents an abstract of the corresponding table by Exner [42]. The

Table VI. Variation of the Distance between Violet and Red Rays (in cm) with Zenith Distance (Θ) and Height (*After Exner*)

Θ(°)	Height in km				
	0.1	1	5	10	40
50	—	—	2	3	5
60	—	—	5	8	12
70	—	3	14	22	31
80	—	15	58	92	127
84	4	37	142	224	311
88	17	151	580	915	1273
90	50	442	1698	2680	3727

color separation for zenith distances of $< 50°$ is extremely small so that chromatic scintillation of stars is generally not perceptible. For greater zenith distances and for increasing heights, the rays' separation rapidly increases. Any air parcel that has a density different from that of its environment (density schlieren) and a diameter less than the rays' separations, will be capable of diverting individual color components into a different direction at different instants, thus causing chromatic scintillation. The size of these air parcels was variously determined as of the order of a few centimeters to a few decimeters [36, 42]. The size of the schlieren in relation to altitude, and the schlieren velocity, obviously determine the possibility and the frequency, respectively, of color fluctuations. For example, in order to produce chromatic scintillation of a star at 80° zenith distance, an air parcel must have a diameter of < 15 cm if at 1 km height, < 58 cm if at 5 km height, etc., whereas near

the ground such a parcel must have an extremely small diameter to cause scintillation.

The mechanism of intensity fluctuations was explained by K. Exner [c. 42] as follows: In Fig. 8, density schlieren around S are embedded at certain intervals in an otherwise more or less homogeneous field of density and cause concavities (or convexities) in an originally plane wave-front of light and, thus, divergence of the rays at A and A' and convergence at B and B'. If the system of schlieren moves horizontally, an observer—say at B—will perceive alternate increases in flux density where the rays converge, and decreases where they diverge. He will also see the light come from slightly varying directions, that is, apparent vibratory motions of the star. It is easily envisioned that the same variations occur if the schlieren system moves vertically. K. Exner measured the radius of curvature r of the wave-front deformation to be roughly between 2 and 20 km. In addition to discrete schlieren, we may also consider the wavy structure of surfaces of temperature or wind discontinuity as a cause of variations in flux density.

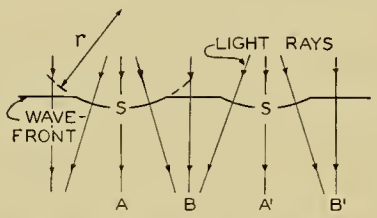


Fig. 8.—Scintillation resulting from density schlieren (*after K. Exner*).

The short-term fluctuations of images received from terrestrial light sources or objects, or *terrestrial scintillation*, is of great practical importance; in particular, strong scintillation may interfere with blinker signaling. Scintillation also limits the precision of telescope pointing and the useful magnification of telescopic devices [46]. Siedentopf and Wisshak[4] recently investigated the case of collimated light sources over a range f km long and 1 m above the ground, employing an objective receiver. With strong scintillation, the frequency of apparent intensity fluctuations was most often observed between 5 and 9 sec^{-1}, with lesser scintillation between 1 and 3 sec^{-1}. The whole range covered the frequencies between 1 and 50 sec^{-1}. The relative variability of the apparent intensity ranged between 0 and 100 per cent and did not materially increase with lengthening of the ray path[5] beyond 1 km. The mean frequency of intensity fluctuation was found to be independent of the path length. Shadow bands of from 5 to 10 cm width and several seconds duration were also observed moving horizontally with the wind across a screen several hundred meters from the searchlight.

The fact that an air column of 1 km length produces almost the entire effect of scintillation shows that the density discontinuities near the receiver are optically the most effective. Brocks [2] similarly found that the atmospheric conditions in the first tenth of a horizontal ray path (counting from the observer) are nineteen times as effectual in producing scintillation as the same conditions in the last tenth of the path. This seems to be the reason why various meteorological elements, observed near the receiver, correlate so well with the degree of terrestrial (and to a certain extent, astronomical) scintillation that predictions of atmospheric optical conditions are possible [4, 24, 42, 55]. Such predictions may not hold where, for example, thermal turbulence is present near the light source but not near the receiver. In this case, scintillation presumably could still occur, if, at the receiver, the angular subtense of the responsible air parcels were large as compared to that of the light source. Also, in cases in which the central portion of very long rays grazes the earth's surface, this portion is particularly exposed to density discontinuities that may not exist at the elevated end points of the rays.

The optical distortions by the atmosphere of nonluminescent, diffusely reflecting sources are generally referred to as *shimmer*, or *atmospheric boil*. Riggs and others [46] measured photographically the apparent lateral displacement of vertical linear targets and the distortion of rectangular grids at several hundred meters distance. They found angular deviations from the mean position of line elements of the order of 1″ to 5″. For targets separated by more than 3′ to 5′, there was no appreciable coherence of the observed deviations, that is, the horizontal dimensions of the schlieren subtended, at the camera, angles of less than 3′ to 5′. Unfortunately, no exact linear dimensions of the experimental arrangement are given, so that only the minimum size of the air parcels can be estimated (roughly 10 cm). In general, the characteristics of terrestrial scintillation appear to be very similar to those of astronomical scintillation; this fact indicates that the cause of the latter must lie predominantly in the lower layers of the atmosphere. Nevertheless, the extent to which the upper atmosphere contributes to astronomical scintillation must still be considered an open question, whose answer must come from direct exploration of the properties of the upper air.

The theories of scintillation by Montigny, K. Exner, and others [c. 42] explain qualitatively the observed phenomena. However, an exact mathematical expression of the relationship between the frequency and amplitude of apparent object motion, apparent intensity fluctuations, and chromatic effects on the one hand, and periodic time-space variations of meteorological factors on the other, remains undeveloped. There are also experimental problems as follows: The scintillatory behavior of uncollimated and diffuse light sources needs further investigation, although it is to be expected that diffuse sources will show a much lesser degree of scintillation than do collimated sources. Of particular interest would be an investigation into the size, shape, spacing, and transport velocity of schlieren in relation to the size, distance, and optical characteristics of the light source for various degrees of scintillation. Such studies

4. R. Meyer [36] cites their paper as unpublished and gives a concise summary of their results.

5. According to K. Exner's results, the minimum source distance that chromatic terrestrial scintillation is observable is about 10 km [c. 42].

could be made, for example, with the aid of a spatial arrangement of a field of light sources and receivers in connection with a dense micrometeorological network. For recording the apparent vibration of objects, motion picture cameras could be employed, whereas photoelectric devices seem to be preferable for measuring the apparent intensity fluctuations. The variation of scintillation with the altitude of the ray path above the ground, as well as with oblique upward and downward direction of the rays, is another problem which seems of practical interest for flight operations.

PHENOMENA DUE TO ATMOSPHERIC SUSPENSIONS

In this section, the sun is considered as the source of light, although the moon or artificial luminants may also produce the phenomena.

Halos. The term "halo," although implying ring shape, is generally applied to all optical phenomena that are produced by ice crystals suspended in the atmosphere and, occasionally, by those deposited on the ground [34, 36].

In Fig. 9, the sun is roughly 25° above the horizon *HH;* ring *A* represents the *22°-halo* (radius 22°), ring *B* the *46°-halo.* The two *parhelia CC* lie to the right and left of the sun at the same elevation, but at angular distances from the sun that vary between 22° and 32° for sun's elevations between 0° and 50°, respectively. The *parhelic circle DD* through the sun and parallel to the horizon is rarely seen as a complete ring; often only short segments of it extend outward from the parhelia. Tangent to the 22°-halo are the *upper* and *lower tangent arcs EE* and *E'E'*, respectively, which are only one of the metamorphic forms of the *circumscribed halo.* This halo is truly circumscribed only for sun's elevation of > 30°. For the various forms of this halo see pertinent literature [21, 34, 42]. The *lateral tangent arcs of the 22°-halo,* or *Lowitz-arcs, FF,* curve concavely toward the sun from the parhelia and touch the 22°-halo below its equator. The vertex (in the sun's meridian) of the *Parry-arc G* has an angular distance from the sun that decreases from 43° at 0° sun's elevation to tangency with the 22°-halo for sun's elevations between 40° and 60°, and then increases again for higher sun's elevations. The *circumzenithal arc J* is centered around the zenith and very near, or even tangent to, the 46°-halo. The *infralateral tangent arcs of the 46°-halo* are represented by *KK;* these arcs also are metamorphosed as their points of tangency move downward and meet at a sun's elevation of 68°, while at the same time their curvature (convex toward the sun) decreases and reverses itself for sun's elevations > 58°. The single arc separates from the 46°-halo when the sun is higher than 68°. The *sun pillar LL* lies in the sun's meridian and is, like the parhelic circle, generally white because of its origin by reflection, whereas the other halos are produced by refraction and thus more or less colored.

There are also other halos, such as the *circumhorizontal arc* that corresponds to the circumzenithal arc, but lies about 46° below the sun. *Supralateral tangent arcs* of the 46°-halo correspond to the infralateral arcs. On the sky opposite the sun, the *anthelion,* a bright spot on the parhelic circle, is sometimes obliquely crossed by *anthelic arcs.* Also rings of unusual radii, 8–9°, 17–19°, 23–24°, etc., have been observed on rare occasions, as well as skewed forms such as inclined pillars and parhelic circles, and secondary phenomena caused by reflection or refraction of light emitted from primary halos.

The geometrical optics of the various halo phenomena was theoretically treated by various authors [21, 34, 42, 44, 58]. In general, most phenomena can be explained by refraction with minimum deflection and/or by reflection involving simple hexagonal ice crystals of columnar or platelet shape, with various attitudes and, in some cases, oscillating motion while falling. The principal genetic features of the major halo phenomena are summarized in Table VII, in which the optical relationship, for example, between the circumzenithal arc and the infralateral arcs of the 46°-halo, becomes evident. There are, however, many phenomena that have been explained by different patterns of ray paths, or by more complicated crystals or crystal aggregates. Thus, for example, the optical origin of the anthelion, its oblique arcs [53], Hevelius' parhelia at about 90°

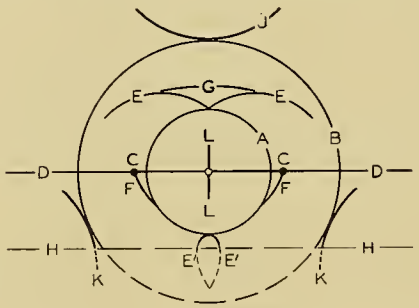


FIG. 9.—Schematic view of major halo phenomena.

from the sun, and other phenomena, is still uncertain; Bouguer's halo, a white ring of about 38° radius around the anthelic point, may even be a fogbow produced by very small supercooled water droplets [34, 47]. Decisive explanations will not come from additional theories, but from an accumulation of better observations. In particular, sampling of the ice crystals producing rare halos or those of uncertain origin would be most desirable. Many halos, theoretically established, have never been observed [58]; on the other hand, some phenomena, such as those observed by Arctowski [c. 42], are still unexplained. Meyer [34] discusses the various geometric problems of halo phenomena, and outlines as prerequisites for a complete theory the various physical aspects, such as the concentration of ice crystals in the clouds, and the brightness and polarization of the halos relative to those of the clouds in the environment. The physical optics of halos is almost entirely unexplored; in addition to refraction and/or reflection by ice crystals, diffraction plays a role in the formation of halos [34, 42]. Photometric observations, that have been introduced into halo investigations by E. and D. Brüche [5], would advance our pertinent knowledge of halos and the constitution of ice clouds [34].

The frequency of different halos, their diurnal and annual or seasonal variations, and their relation to weather situations, have been investigated [1, c. 34, 52]. The similarities or differences in the results can generally be ascribed to climatic characteristics. The relationship between halo occurrence and solar activity is, however, still quite problematic. Visser [52] found a decrease in halo frequency with increasing relative number of sunspots[6] up to 90 and then an increase for higher sunspot numbers. Archenhold [1] arrived at a direct linear relationship, and a halo periodicity of 27–28 days, which he associates with the rotation of the sun. This periodicity was shown to be spurious [39]. Although a certain degree of correlation with solar activity seems to exist, there is hardly a direct relationship (*e.g.*, corpuscular solar radiation furnishing sublimation nuclei), considering the prerequisites that must be fulfilled for a halo to be observable [34, 39].

Table VII. Principal Genetic Features of Major Halo Phenomena

Halo	Refracting angle	Orientation of principal crystal axis	Special characteristics (s = sun's elevation)
22°	60°	Random	Incident ray at 90° to principal axis.
Parhelia	60°	Vertical	Intensity rapidly decreases for $s > 50°$.
Circumscribed to 22°-halo	60°	Horizontal	Upper and lower tangent arcs join at $s \geqq 30°$. At $s \geqq 55°$ elliptic with long horizontal axis. At $s = 90°$ circular and coincides with 22°-halo.
Parry-arcs	60°	Horizontal	Pair of crystal sides vertical for $s < 30°$, horizontal for $30° < s < 50°$.
46°	90°	Random	Incident ray at 90° to principal axis.
Circumzenithal arc	90°	Vertical	Ray entrance at crystal base; limited to $s \leqq 32°$; also possible with horizontal axis and pair of sides horizontal, but rare.
Cricumhorizontal arc	90°	Vertical	Ray entrance at crystal side; limited to $s \geqq 58°$; also possible with horizontal axis and pair of sides horizontal.
Lateral tangent arcs of 22°-halo (Lowitz-arcs)	60°	Oscillating	Oscillation $< \pm 30°$ about vertical in plane that is parallel to sun's meridional plane. Limited to $25° < s < 55°$.
Infralateral tangent arcs of 46°-halo	90°	Horizontal	Ray entrance at crystal base. Limited to $0° < s < 68°$.
Supralateral tangent arcs of 46°-halo	90°	Horizontal	Ray entrance at crystal side. Limited to $0° < s < 32°$.

Rainbows. The colorful arcs around the antisolar point that appear on sheets of water droplets are called rainbows, although these phenomena may be produced by dew droplets on the ground or water sprays. The *primary rainbow* has an angular radius to the red outer border of roughly 42°; concentric to it is the *secondary rainbow*, whose radius to the red inner border is about 51°. Inside the primary and outside the secondary bow, *supernumerary bows* showing fewer and fainter colors are often visible. When the sun's rays are reflected by a smooth water surface before striking the suspended droplets, primary and secondary *reflection rainbows* may appear whose center lies as much above the horizon as that of the regular bows lies below it. Thus, the reflection bows intersect the respective regular ones at the horizon. When the reflecting water surface is undulated by a smooth swell, the reflection bow may deform into vertical shafts [57]. Droplets of radii $< 30\ \mu$ may produce a broad white band with faintly tinted borders, the *fogbow*, between about 37° and 40° distance from the antisolar point. Rainbows or fogbows produced by drops in a horizontal plane appear in the form of conic sections, that is, hyperbolic, elliptic, or parabolic arcs, depending on whether the sun's elevation is respectively smaller than, larger than, or equal to, the aperture of the cone (42° or 51°).

The well-known explanation by Descartes considered geometrical optics alone. Airy, basing his theory on wave optics, explained the variation of colors with droplet size and the supernumeraries as interference rings. His rainbow integral was also solved by others. The distribution of intensity, color, and polarization of

6. According to Schindler [c. 36], a relationship between sunspots and halos begins only at a spot number of 35 to 40.

the light after refraction and reflection by the droplets was computed, notably by Pernter and Möbius [c. 42]. Bucerius [6] developed an asymptotic method (in analogy to Debye's treatment of cylindric functions) by means of which Mie's theory of scattering of electromagnetic waves by dielectric spheres [c. 29] can be extended to large waterdrops. Applying this general theory to the rainbow phenomenon, he showed that it contains the older rainbow theories as approximations, and that the rainbow is an areal phenomenon that actually covers the entire region between the antisolar point and the first ring. Meyer [36] considers the classical diffraction theory still adequate, particularly for the intense rainbows that originate from comparatively large droplets. He developed this theory further to permit the determination of the luminous density of rainbows. He takes into consideration the total optical effect of all droplets contained in a surface element of the cloud deck, the thickness of the cloud, and the attenuation of the rays to and from the cloud. He finds the luminous density of the primary rainbow to be twelve times that of the secondary bow.

The theoretical advancement of our knowledge of rainbows appears to have surpassed our fund of observational data. Aside from older visual measurements, there are, to my knowledge, no results of up-to-date colorimetric photometry of rainbows available for application to the theoretical findings. In this connection an almost forgotten problem may be recalled, the fluctuations of colors during lightning and thunder [42], an observation requiring objective verification and explanation.

Corona and Related Phenomena. The sun shining through relatively thin clouds often produces one or more sets of colored rings, the *corona*, having diameters of a few degrees. When poorly developed, only an *aureole*, a bluish-white disk with brownish rim, may be visible. After violent volcanic eruptions, a broad reddish-brown ring of large radius (20° and more), *Bishop's ring*, has been observed in dust clouds [11]. We speak of *iridescent clouds*, when the colors are not arranged concentrically around the sun, but are irregularly distributed over, or follow the contours of, the cloud. This group of coronal phenomena around the sun is paralleled around the antisolar point by a similar group: An observer, seeing his slightly enlarged shadow, the *Brocken-specter*, on a fog bank or cloud, often finds the shadow of his head surrounded by one or more sets of colored rings, the *anticorona* or *glory*, well-known to pilots. If the shadow falls on a bedewed surface on the ground at some distance from the observer, the shadow of his head may be rimmed by a narrow white sheen, the *heiligenschein*, which also can be observed around one's head-shadow on a beaded projection screen.

The classical diffraction theory applied to the corona, under the assumption that the droplets are opaque, has been found to be in fair agreement with observations [42]. The well-known approximation formula by K. Exner [c. 42] for the angular radius θ of the circular intensity minima produced by particles of the diameter d in light of wave length λ, is

$$\sin \theta = (N + a)\lambda/d, \qquad (6)$$

where N is the order of the minimum counting from the center, $a = 0.22$ for spherical, $a = 0$ for nonspherical particles. Ramachandran [45] based his new corona theory on wave optics and included the wave-front portion that is transmitted through the droplets. In Fig. 10 the results of his calculation of intensities ($\lambda = 0.5\ \mu$) at various diffraction angles (θ) for small droplets (radii in μ ascribed to the curves) are reproduced. These curves indicate that the ring systems oscillate as the small droplets increase in size. Only relatively large droplets diffract like opaque disks of the same size, which explains some of the discrepancies formerly noted between the classical theory and observations. The position of the ring systems appears

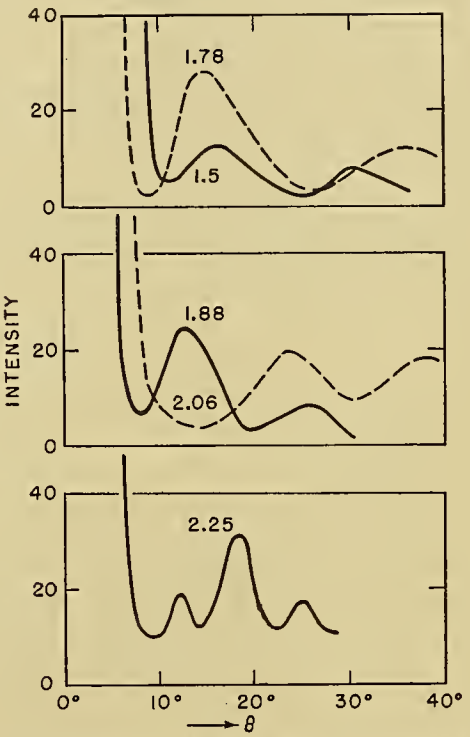


Fig. 10.—Intensity of diffracted light as function of angular distance from light source. (*After Ramachandran.* Ordinate scale presumably in relative units; numbers on curves are drop radii in microns.)

unaffected by the thickness and density of the clouds. Bucerius' work [6], mentioned above, also includes the application of the rigorous diffraction theory by Mie [c. 29] to both corona and anticorona. The anticorona was similarly treated by van de Hulst [20]. This theory yields the intensity and polarization of the diffracted light and the position of intensity maxima and minima. Table VIII gives a comparison of the values for the argument of the Bessel function at which corona maxima occur according to the old and the newer theories. It is noteworthy that in Ramachandran's theory the location and intensity of the maximum for small drops also depends on the value of $(\sin \xi)/\xi$, where ξ is a function of d/λ. For this reason the maxima oscillate as shown in Fig. 10. According to Bucerius [6, equation (47)], the argument contains twice the sine of half the angle between primary and diffracted rays.

instead of the sine of the whole angle θ. This is also true for the anticorona [6, equation (48)], the theory of which shows that it is produced by rearward diffraction, not by reflection of the primary ray with subsequent forward diffraction [Richarz, c. 42]. Bucerius' results may be summarized as follows: The intensity of the corona is $x^2 = (\pi d/\lambda)^2$ times as great as that of the anticorona; the intensity at the center of the anticorona is a minimum, that of the corona a maximum. Values for the angle θ of the successive intensity maxima of the anticorona[7] are determined by $2x \sin(\theta/2) = 3.05, 6.7, 10.0, 13.2, \cdots$; the minima are located at relatively the same position as are the corona maxima (Table VIII). This explains why the application of the classical corona formula (6) to the minima of the anticorona yielded values of the drop diameter d, that varied with the order of the minimum [c. 42, 47]. Also the decrease in intensity of successive maxima is much greater for the corona than for the anticorona; therefore, multiple glories are more frequently observable than multiple coronas.

The old controversy regarding the possibility of coronas and glories in ice-crystal clouds [21] now stands as follows: Multiple-colored rings generally indicate the presence of water droplets; however, the production of faintly colored glories by ice clouds has been established [47]. A statistical survey by Peppler [41] revealed that 78 per cent of glories were simultaneously observed with fogbows at temperatures between 0C and −4C; a maximum frequency of glories occurs at about −4C, that of halos at about −12C. Nevertheless, the frequency curves of anticoronas and halos overlap in a wide range of temperatures from about −2C to < −20C. At any rate, this problem cannot be considered solved. Statistical analyses of the occurrence of diffraction rings simultaneously with halos or fogbows reveal, at best, the relative frequency of diffraction rings in ice and water clouds, respectively, but are entirely inconclusive regarding the physical possibility of these phenomena in ice clouds. Moreover, Stranz [c. 36] by means of photronic cells, detected multiple coronas that were invisible to the eye. The theoretical objections to the possibility of diffraction phenomena produced by ice clouds were mainly based on the optical properties and orientation of ice needles, but other possible crystal forms must also be considered. Moreover, the occurrence of Bishop's ring in dust clouds shows that nonspherical particles are capable of producing diffraction rings. Unfortunately, none of the new theories considers diffraction by nonspherical particles, so that no final decision can be made.

Iridescence of clouds is explained as diffraction patterns produced by groups of uniform droplets that vary in size in different portions of the cloud. In view of the great sensitivity of the diffraction patterns to slight differences in size of small droplets [45] (Fig. 10), it is no longer difficult to explain the occurrence of iridescence at relatively large angular distances from the sun [c. 42]. The heiligenschein is considered the result of external reflection by the dew droplets [21, 42]; to what extent diffraction plays a role in this phenomenon is not known. Experimental data or intensity measurements are completely lacking.

The corona and anticorona have been widely employed in the study of cloud and fog elements. Measurements were largely confined to the angular radius of prominent rings and subsequent evaluation in terms of droplet radius. In view of recent theoretical developments [6, 20, 45] this geometric method seems unreliable; also the difficulty in visually locating diffuse rings, produced by inhomogeneous fogs or clouds, causes considerable uncertainties (see [5]). In the future it would be preferable to resort to objective monochromatic photometry of the entire zone around the light source (or shadow center), and perhaps, to determine the intensity of the diffracted light separately for the two components of polarization. Simultaneous determination of the droplet size by other means could serve as a check of the theory by making possible a comparison between observed and theoretical intensity distributions, rather than a comparison of only the minima or maxima.

Considering the rarity of homogeneous fogs, a theoretical and experimental study of inhomogeneous fogs appears of particular practical importance. Also, a final answer to the question of the possibility of coronas in ice clouds would give the observer on the ground a tool for the identification of the physical state of the clouds.

Table VIII. Comparison of Position of Corona Maxima of Various Orders According to Different Theories

Author	Argument of Bessel Function	Order of Maximum				
		1	2	3	4	5
Mascart [c. 21]	π d(sin θ)/λ	5.14	8.46	11.67	14.84	17.98
Ramachandran [45]	π d(sin θ)/λ	5.14	8.42	11.62	14.78	17.96
Bucerius [6]	2π d(sin $\theta/2$)/λ	5.1	8.4	11.6	14.8	—

TWILIGHT PHENOMENA

The investigation of twilight phenomena is closely connected with the study of the optical properties of the upper atmosphere, at least to a height of 60 km [18] and, thus, indirectly with the study of its density and dust content. The discussion of the temporal developments of the various phenomena is based on the sunset and the sun's meridian. Figure 11 shows the nomenclature for the significant astronomic-geometric features pertaining to the half-space above the observer and a schematic view of the major phenomena.

7. Bucerius (also [36]) gives here $2\pi x \sin(\theta/2)$, but refers to it as the argument of the Bessel function which, however, appears as $2\pi d(\sin\theta/2)/\lambda = 2x \sin(\theta/2)$.

Description. At, or shortly after, sunset, the *antitwilight arch*, a purplish band of some 3° or more in width, can be seen to rise above the solar counterpoint on the eastern horizon. At about 1° sun's depression, the gray-blue *dark segment* or *earth's shadow* begins to rise beneath the antitwilight arch. At approximately 2° sun's depression, the *purple light* appears as a purplish area above the solar point in the western sky. This area has a vertical angular extent of 10° to 50°, a lateral extent of 40° to 80°. Its upper boundary, which has an elevation of about 50° at the beginning, descends steadily to the horizon. The purple light reaches its maximum intensity at about 4° sun's depression and usually disappears at about 6° sun's depression. The rising antitwilight arch usually fades from view when the purple light is at its height, and shortly afterwards, the dark segment becomes indistinguishable from the rest of the darkening sky. Its transit through the zenith generally cannot be observed, but it reappears as the *bright segment*[8] or *twilight arch* above the solar point, when the sun's depression is about 7°. The bright segment disappears below the western horizon at about 16° sun's depression.

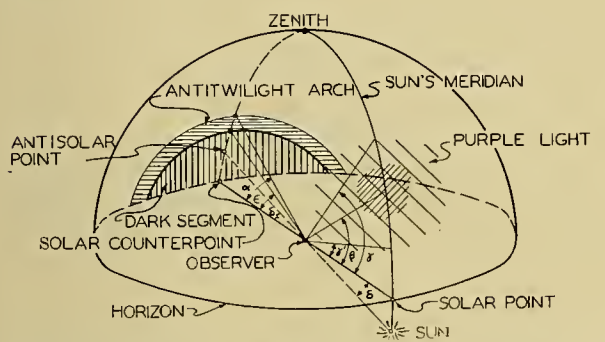


FIG. 11.—Schematic diagram of major twilight phenomena (elevation angles are α = antitwilight arch, β = maximum purple light intensity, γ and γ' = upper and lower boundary of purple light, δ = antisolar point, and ϵ = dark segment).

When the sun's rays are partially obstructed by clouds or mountain peaks, the purple light may assume a ray-structure because of the interruption by the shadow bands (crepuscular rays) which seem to converge towards the sun. Similarly, the continuation of these shadow bands (anticrepuscular rays) on the eastern sky may give the antitwilight arch a fanlike appearance. Colored illustrations of the various twilight phenomena can be found in [16].

During brilliant twilight phenomena, a secondary purple light may become visible after the main purple light has disappeared. Also a secondary antitwilight arch and dark segment may develop within the primary dark segment. These secondary phenomena are much more diffuse in outline and show fainter colors.

8. The descriptive term "bright segment" is preferable because of its fundamental identity with the dark segment and the basic difference from the "antitwilight arch." The term "earth's shadow" is physically incorrect [40] and does not readily permit differentiation between the phenomena on either side of the zenith.

For the measurements of twilight phenomena the following spatial and temporal aspects are generally considered: The elevation angle of the upper boundary of the antitwilight arch, of dark and bright segments, and of purple light; and the sun's depression at the time of beginning, end, and maximum intensity of the purple light. In addition, photometric measurements of the light intensity in various spectral ranges along significant portions of the sun's meridian are of major interest.

Results of Observations. The development pattern of the purple light has been found to be practically the same everywhere except at altitudes above 2000 m where the purple light ends at somewhat greater sun's depressions and where its upper boundary reaches greater elevations. A greatly detailed analysis of visual observations, such as made by Gruner [14] and Dorno [11], seems hardly warranted in view of the fact that the sun's depressions are computed without regard to the variable refraction, that the intensity is estimated according to a memory scale, and that the measurement of elevation of the diffuse boundary of the delicately tinted phenomenon is affected by subjective factors [50]. In Europe, a maximum frequency of bright purple lights occurs in autumn, a minimum in spring; this fact has been attributed to the corresponding frequency of anticyclones with clear skies in that area [51]. Otherwise no significant relationship between weather and purple lights has been found.

Secular variations of intense purple lights have been observed associated with dust produced by volcanic eruptions [c. 42, 50]. There exists, however, a time lag; for example, after the Katmai eruption in summer 1912, purple lights did not reach their maximum intensity until summer 1913 [11]; this delay was obviously due to the time involved in the sedimentation of dust particles necessary to produce the optimum concentration and size distribution for the formation of the purple light. For this reason, an absence of dust layers cannot be deduced from an absence of intense purple lights [50].

Although visual observations have long been recognized as inadequate, objective methods have been employed only in relatively recent times [13, 14, 32]. The techniques involved still need improvement and standardization. The results obtained at different stations from photoelectric [13] and photographic [32] intensity-measurements with color filters show a maximum red content of the sky light between 4° and 5° sun's depression, corresponding to the visually observed maximum relative intensity of the purple light. The absolute luminous density of the sky light decreases steadily in all spectral ranges with increasing sun's depression in contrast to the visual impression [14]. Whereas the results from visual observations were essentially confirmed by objective methods [13], the latter have shown the presence of the purple light when the spectral differences in intensity were below the threshold of visual perception [32].

Gruner [14] has indicated a method for determining

the height of the layer responsible for the purple light. By graphical approximation he arrived at values between 25 and 31 km for the upper boundary of the layer, and 18 km for the thickness of the generating ray beam, that is, a thickness of the layer of roughly the same magnitude. Smosarski [c. 14, 50] estimated the lower boundary at 8 km and the upper boundary at 17 km. The relationship between purple lights and high dust layers of volcanic origin seems to confirm at least the order of magnitude of these values. The possibility of a connection with the ozone layer is an unexplored question. Incidentally, Gruner [14], assuming the purple light to be a corona, also determined the order of magnitude of the particle diameter as between 1 and 1.5 μ. For these small particles, however, the classical diffraction theory fails [45] as was shown in the preceding section.

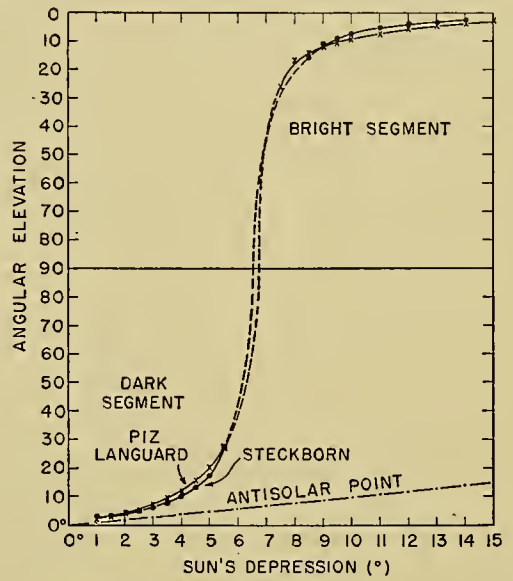


Fig. 12.—Average course of dark and bright segments at two Swiss stations.

Several series of geometric measurements of the elevation of the upper boundary and the width of the antitwilight arch have been made. The course of its angular elevation is similar to that of the dark segment (Fig. 12). Mendelssohn and Dember [33] made a few spectral measurements by photographic photometry, but their results confuse, rather than elucidate, the visual observation. The annual and secular variations of antitwilight-arch occurrence were determined by Smosarski [50] who found, in Poland, a maximum frequency in autumn and winter, a minimum in summer, and a good correlation with volcanic activities. A direct connection with the occurrence of purple lights does not seem to exist. According to computations by Smosarski [c. 14], as the sun's depression increases, the antitwilight arch is produced by the rearward scattering of sunlight by smaller particles at higher altitudes. However, according to Mendelssohn and Dember [33], the antitwilight arch is supposedly fixed in a layer between 4 and 9 km. This problem will be discussed below.

Of the many observations on the movement of the dark and bright segments, only the averages of two series obtained at Piz Languard (3280 m) and Steckborn (430 m) in Switzerland, as reported by Gruner [14], are shown as examples in Fig. 12. As the sun sinks below the horizon, the dark segment rises more rapidly than does the antisolar point; the ascent rate increases with the sun's depression, until the segment fades from view between 5° and 6° depression. Between 7° and 8° depression the bright segment appears at an elevation of roughly 25° above the solar point and descends, first rapidly then more slowly, to the western horizon. According to the interpolated (dashed) portions of the curves, the invisible transit through the zenith occurs at a sun's depression between 6° and 7°. This agrees well with observations of the zenith brightness by Brunner [c. 14] and Hulburt [18], and of global illumination by Siedentopf and Holl [48]; these authors present curves of brightness and illumination, respectively, versus sun's depression, that show a definite inflection point between 6° and 7° sun's depressions. This fact is involved in the problem of the height at which this phenomenon occurs. In this connection the change in relative variability of the dark segment's elevation at various sun's depressions deserves attention. It has been shown [26, 40] that the variability decreases rapidly until the sun's depression is about 2.5°, then more slowly, although the opposite trend was to be expected in view of the decreased accuracy of measurement at greater sun's depressions. This fact was interpreted as being caused by a transition of the dark segment at about 2.5° sun's depression into the stratosphere, where marked changes in turbidity from day to day are less frequent [40].

The sun's depressions at which the last traces of the bright segment disappear below the horizon have been variously used to compute the height at which the density of the atmosphere is sufficiently great to produce visible scattering of the direct sunlight. However, the resulting values of 50–65 km [14] are still quite problematic because of subjective factors involved in the perception of faint light and of the effects of attenuation of the direct light rays at various altitudes. Except for a slight increase in elevation of the dark segment in summer and with diminished transparency of the lower layers of the atmosphere, no clear-cut relationships between the turbidity of the air and the bright segment, nor a definite seasonal variation of either segment have been established [14, 16, 18, 26, 40].

Theories and Problems. Although many theoretical approaches to the problems of twilight phenomena have been made, no complete theory exists as yet. The theories of the dark and bright segments and of the antitwilight arch [15, 16] agree qualitatively with, but differ quantitatively from, the observations. For example, the elevations of the dark segment, computed from the spectral intensities of light scattered by a Rayleigh atmosphere (disregarding multiple scattering), were considerably smaller than the observed ones [14, 15]. The principal ideas underlying the explanations of

the segments and antitwilight arch are schematically demonstrated with the aid of Fig. 13 as follows: R_0 to R_3 are sun's rays passing through the atmosphere whose optically effective height may be E_3D_3. The lowest ray R_0 touches the earth's surface at O_0. Owing to scattering and absorption on their way through the air, any rays between R_0 and R_1 lose part of their short-wave components and their intensity is depleted, so that R_0 may not reach much beyond O_0, the next higher ray a little farther, and so on. The envelope of the end points of all these rays is represented by the curve $O_0D_1D_2D_3$. Consequently, the atmosphere to the right of this curve lies in the shadow of the earth. An observer O_1, for whom the sun's depression is δ_1, sights the vertex of the dark segment at D_1, the point of tangency of his line of sight O_1S_1 to the ray envelope. If he raises his line of sight slightly, he perceives light scattered from the still illuminated portion of the atmosphere near the envelope. This prevalently reddish light constitutes the antitwilight arch. Its upper boundary is seen when the line of sight at O_1 is further raised so that the diminishing light from the zone of red rays is compensated by the increasing light of shorter wave lengths from the higher

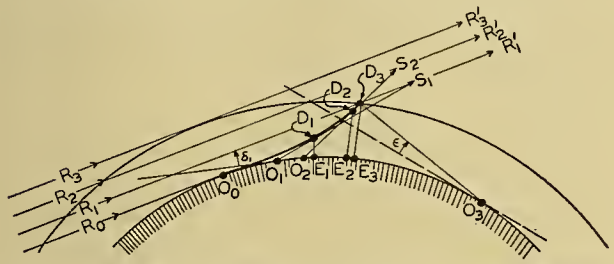


Fig. 13.—Schematic diagram for the dark and bright segments.

layers of the atmosphere. When the sun sinks further, the relative position of the observer shifts to O_2, where the line of sight O_2S_2 touches the envelope at D_2. For an observer at E_3, the passage of the dark segment through the zenith is imperceptible, because there is not enough contrast between the sky brightness to the right and left of the line of sight E_3D_3. Finally, the observer at O_3 sees the bright segment at an elevation ϵ.

The geometric aspects of this problem have been studied by various authors [c. 14, 40]. The results of computations of the heights of points D at various sun's depressions, although based on different assumptions, agree fairly well as shown by the examples in Table IX. According to these heights, point D moves into the stratosphere at a sun's depression between 2° and 3°, in agreement with the deductions made from the variability of the dark segment. However, the problem is essentially a photometric one [42]. An attempt at a physical solution was made by Dember and Uibe [10], who took into consideration the visibility as proportional to the square root of the measured sky brightness. Application of this theory by Mendelssohn [33] to photometric measurement yielded a constant height of the dark segment between 2 and 4 km. However, the transit of the dark segment would then have to occur not later than at about 3.5° sun's depression [40], which is contrary to observation. Nevertheless, the basic idea of including the attenuation of light along the line of sight is correct. This was suggested by Exner [42] who, however, based his formula on Rayleigh scattering alone and disregarded secondary scattering which undoubtedly plays a role in the brightness of the sky below the ray envelope [18].

Table IX. Height of Dark Segment at Various Sun's Depressions

δ(°)	Mohn [c. 14] (km)	Neuberger [40] (km)
1	—	2
2	11	8
3	21	17
4	31	27
5	40	38

The major problematic factors pertaining to the segments and antitwilight arch are as follows: Two influences on the ray envelope must be considered, that of atmospheric refraction which causes a vertical divergence of the sun's rays (shown as parallels in Fig. 13), and that of atmospheric attenuation of the rays near the ground. This attenuation tends to counteract the effect of refraction by eliminating the lowest, most refracted rays [27]. The position and shape of the ray envelope is, thus, primarily a function of the transparency of the atmosphere at and above the point of tangency with the earth's surface. As regards the intensity and color of the scattered light, most computations have been based on an idealized atmosphere in which Rayleigh's theory with its symmetric scattering function is valid [14, 15, 18]. However, the real atmosphere contains a large number of particles, especially in the lower layers. For this reason, the agreement between theory and observations is not satisfactory, and, in particular, the variations from day to day observed in dark and bright segments and antitwilight arch remain unexplained. According to Linke [29], the rigorous theory by Mie-Debye is more suitable for the theoretical approach to the twilight colors. However, this theory is difficult to apply to the problems at hand, because it still involves the assumption of spherical particles. In view of the new theories of the anticorona [6, 20], the consideration of rearward diffraction should be extended to the theory of the antitwilight arch.

The theory of the purple light which was recognized as a diffraction phenomenon by Kiessling [c. 42] was established by Gruner [14] along the lines suggested by Pernter [42] and others. This theory adequately describes the temporal and spatial development of the purple light; the basic ideas may briefly be outlined with the aid of schematic Fig. 14. The sun's rays R_0 to R_3 pass through a dust layer DD' in which they are deprived of their short-wave components. While R_0, passing through the dense lower layers, and R_3, having the longest path through the dust layer, may be completely extinguished, the rays around R_1 will emerge as a reddish beam of light and enter the lower boundary of the dust layer again at P_1, where the particles will

be relatively large. These particles may diffract the light dominantly at the angle φ_1 toward the observer O, whose horizon is HH', and for whom the sun's depression is δ. At the higher points P_2 and P_3, the prevailing sizes of the dust particles become successively smaller and the diffraction angles φ_2 and φ_3 correspondingly larger. Since the incident beam is reddish, the diffracted rays converging toward O will outline a reddish area in the sky. The observer sees only those diffracted rays that fall in his cone of vision (and are sufficiently intense); for example, a particle at P_1' of the same size as that at P_1 and one at P_2' of the same size as the one at P_2, will also diffract light toward O. It may be noted that an observer, for whom the sun is still above the horizon, may see Bishop's ring around the sun. Also, the light scattered and diffracted by the dust layer in the region of the purple light is sometimes sufficiently intense to cause a secondary purple light for an observer located farther on the night side of the earth (*i.e.*, to the right of observer O in Fig. 14). The red light diffracted by the dust layer and augmented by blue light scattered by the air in the region $S_1S_2S_3$ above gives rise to a purplish tone.

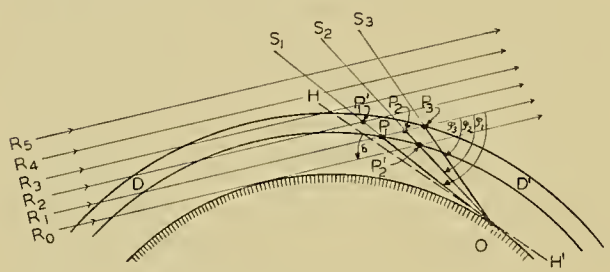


Fig. 14.—Schematic diagram for the purple light.

From the geometric aspects we can see that the areal extent of the purple light depends on the particle size distribution, the thickness of the layer, and absorption. The latter generally prevents the purple light from appearing as a circular arc. The theory does not predict the exact color and intensity of the phenomenon; these depend on the modification of the incident beam by its first passage through the dust layer, its further fate in the air below this layer, the specific effect of the dust on the beam after re-entrance into the layer, and, finally, the modification of the diffracted light on the way to the observer, in conjunction with the sky light from above. For a homogeneous dust layer, an optimum particle concentration and size distribution must exist, whereby the red component of the incident sunlight experiences a minimum depletion on its first passage through the dust layer and produces maximum intensity of diffracted light when again penetrating the layer.

In general, there are too many unknown variables, in particular the scattering function of the dust particles and attenuation of the direct rays, to render the problem as a whole accessible to an analytical solution, especially when multiple dust layers are involved. For the present, it appears most expedient to provide reliable observational material by means of which the available theories may be checked more adequately. In order to eliminate from such material all the subjective variables that are involved in visual observations, only objective methods of observation should be employed. The design of the spectrophotometric equipment should incorporate the features of very high sensitivity, to enable the use of filters with narrow transmission bands, and of rapid response, so that the major portions of the sky area could be scanned within a few minutes. By means of a wide station network, the question regarding the cause of asymmetric twilight phenomena that have been variously attributed to the shape of the atmosphere as a whole or the slope of the tropopause [14] could be answered. The problem of the height of the effective layers of the atmosphere and of the shape of the ray envelope could be approached by means of airborne photometric instruments to furnish vertical cross sections of the light flux at various altitudes along a latitudinal line to represent various sun's depressions. Spectrophotometry of clouds of known height may furnish information on the geometrical and optical properties of the sun's rays tangent to the earth's surface.

As regards determination of the terrestrial or possibly cosmic origin of dust layers [11, 14] that periodically produce striking twilight phenomena, only a long-range observational project on an international basis will lead to success.

SCATTERING OF LIGHT IN THE ATMOSPHERE

Scattering is the deflection of light quanta in a transparent medium such as the atmosphere. The atoms and molecules of the gaseous constituents cause the quanta of the incident light beam to be scattered more or less in all directions. In addition, there is scattering by minute particulate suspensions such as condensation nuclei. When a beam of this scattered light encounters further matter, it is again subject to (multiple) scattering; however, the contribution of multiple scattering to the total intensity of scattered light is small, except in very turbid air or in the absence of primarily scattered light, (*e.g.*, in the case of the dark segment).

The classical theory by Rayleigh [c. 21] was found to be only in approximate agreement with pertinent observations [c. 42]. The later theory by Mie and Debye [c. 29], which contains Rayleigh's theory as a limiting case, is more general, but difficult to apply to atmospheric scattering because it requires a knowledge of specific constants, the distribution, and concentration of the scattering substances. For a thorough summary of the theories of scattering as well as of methods and results of observations, the reader is referred to the work by Linke [29].[9]

It was shown that the scattering process was involved in some of the previously discussed optical phenomena; its more immediate manifestations are much less spectacular, but nevertheless of great practical importance. The essential consequences of scattering are: The restric-

9. See also Chap. 7 on "Die kurzwellige Himmelsstrahlung" in the same volume, pp. 339–415.

tions in visual range; the polarization of sky light; the depletion of direct sunlight; and the luminance of the sky in daytime. Only some aspects of sky luminance will be briefly mentioned here, as the other topics are treated elsewhere in this Compendium.[10]

The luminance of the sky represents a considerable portion of that direct sunlight which has been depleted in its passage through the atmosphere. A practical aspect of this luminance is the resultant illumination without which the earth's surface would be in darkness except where reached by direct or reflected sunlight. The most obvious characteristic of the sky's luminance is its blue color which is caused by the preference of the scattering process for the shorter wave lengths of the incident radiation. Far from being a pure blue, the color is composed of other wave lengths to an extent that varies with the state of atmospheric turbidity, because with increase in size of the scattering particles the longer wave lengths increasingly participate in the scattering process. In the sky light, the ultraviolet component, whose intensity may exceed that of the direct sunlight, has biological (erythematous and bactericidal) and technological (dye-fading, photographic) effects.

The luminance of the sky is significant from a meteorological viewpoint because it enters as a factor in the appraisal of the atmospheric radiation balance and serves as a criterion of the turbidity of the air. In this respect, even mere estimations of the variations in the sky blue have been shown to be of practical value.[11]

REFERENCES

1. Archenhold, G., "Untersuehungen über den Zusammenhang der Haloerscheinungen mit der Sonnentätigkeit." *Beitr. Geophys.*, 53: 395–475 (1938).
2. Brocks, K., "Vertikaler Temperaturgradient und terrestrische Refraktion, insbesondere im Hoehgebirge." *Veröff. meteor. Inst. Univ. Berlin*, Bd. 3, Nr. 4 (1939).
3. —— "Eine Methode zur Beobachtung des vertikalen Dichte- und Temperaturgefälles in den bodenfernen Atmosphärenschichten." *Meteor. Z.*, 57: 19–26 (1940).
4. —— "Lokale Unterschiede und zeitliche Änderungen der Diehteschichtung in der Gebirgsatmosphäre." *Meteor. Z.*, 57: 62–73 (1940).
5. Brüche, E., und Brüche, D., "Über die Photometrie von Sonnenringen." *Meteor. Z.*, 49: 289–294 (1932).
6. Bucerius, H., "Theorie des Regenbogens und der Glorie." *Optik*, 1: 188–212 (1946).
7. Couder, A., "Mesure photographique de l'agitation atmosphérique des images stellaires." *C. R. Acad. Sci., Paris*, 203: 609–611 (1936).
8. Dember, H., und Uibe, M., "Über die scheinbare Gestalt des Himmelsgewölbes." *Ber. sächs. Ges. (Akad.) Wiss.*, 69: 139–148 (1917).
9. —— "Über die Gestalt des siehtbaren Himmelsgewölbes." *Ber. sächs. Ges. (Akad.) Wiss.*, 69: 391–411 (1917).
10. —— "Über eine physikalische Theorie der Bewegung des Erdschattens in der Atmosphäre." *Ber. sächs. Ges. (Akad.) Wiss.*, 71: 227–239 (1919).
11. Dorno, C., "Dämmerungsbeobaehtungen Herbst 1911 bis Anfang 1917." *Meteor. Z.*, 34: 153–165 (1917).
12. Esclangon, E., "Sur les réfractions géodésiques." *C. R. Acad. Sci., Paris*, 216: 137–139 (1943).
13. Gruner, P., "Photometrie des Purpurliehtes." *Beitr. Geophys.*, 50: 143–149; 51: 174–194 (1937).
14. —— "Dämmerungserscheinungen," *Handb. Geophysik*, Bd. 8, Kap. 8, SS. 432–526. Berlin-Zehlendorf, Gebrüder Borntträger, 1942.
15. —— und Klee, T., "Numerische Berechnung der Helligkeit des Himmels im Sonnenvertikal." *Helv. phys. Acta*, 11: 513–530 (1938).
16. Gruner, P., und Kleinert, H., "Die Dämmerungserscheinungen." *Probl. kosm. Physik*, Bd. 10. Hamburg, H. Grand, 1928.
17. Hillers, W., "Über eine leicht beobachtbare Luftspiegelung bei Hamburg und die Erklärung solcher Erscheinungen." *UnterrBl. Math. Naturw.*, 19: 21–38 (1913).
18. Hulburt, E. O., "The Brightness of the Twilight Sky and the Density and Temperature of the Atmosphere." *J. opt. Soc. Amer.*, 28: 227–236 (1938).
19. —— "The Green Segment Seen from an Airplane." *J. opt. Soc. Amer.*, 39: 409 (1949).
20. Hulst, H. C. van de, "A Theory of the Anti-coronae." *J. opt. Soc. Amer.*, 37: 16–22 (1947).
21. Humphreys, W. J., *Physics of the Air*, 3rd ed. New York, McGraw, 1940. (See pp. 451–556)
22. Hüttenhain, E., "Zur scheinbaren Gestalt des Wolkenhimmels." *Himmelswelt*, 46: 94–101 (1936).
23. Isimaru, Y., "On the Form of the Firmament, Apparent Shape of Cloud and Visual Magnitude of the Sun and Moon." *J. meteor. Soc. Japan*, 4: 3–10 (1926).
24. Ives, R. L., "Meteorological Conditions Accompanying Mirages in the Salt Lake Desert." *J. Franklin Inst.*, 245: 457–473 (1948).
25. Jeffreys, H., "The Shape of the Sky." *Meteor. Mag.*, 56: 173–177 (1921).
26. Jensen, C., "Atmosphärisch-Optische Messungen in Ilmenau." *Beitr. Geophys.* 35: 166–188 (1932).
27. Link, F., "L'éclairement de la haute atmosphère et les tables erépusculaires de M. Jean Lugeon." *C. R. Acad. Sci., Paris*, 199: 303–305 (1934).
28. —— und Sekera, Z., "Dioptrische Tafeln der Erdatmosphäre." *Publ. Pražské Hvězdárny*, No. 14, pp. 1–28 (1940).
29. Linke, F., "Die Theorie der Zerstreuung, Extinktion und Polarisation des Liehtes in der Atmosphäre." *Handb. Geophysik*, Bd. 8, Kap. 5, SS. 120–238. Berlin-Zehlendorf, Gebrüder Borntträger, 1942.
30. Luneburg, R. K., *Mathematical Analysis of Binocular Vision*. Princeton, Princeton University Press, 1947.
31. Meggers, W. F., and Peters, C. G., "Measurements on the Index of Refraction of Air for Wave Lengths from 2218 A to 9000 A." *Bull. nat. Bur. Stand.*, 14: 697–740 (1919).
32. Mendelssohn, T., "Photometrie des Purpurlichtes." *Rev. Fac. Sci. Univ. Istanbul (Istanbul Üniversitesi Fen Fakültesi Mecmuasi)*, Ser. A, 6: 94–115 (1941).
33. —— und Dember, H., "Über die Bewegung des Erdschattens in der Atmosphäre." *Rev. Fac. Sci. Univ. Istanbul (Istanbul Üniversitesi Fen Fakültesi Mecmuasi)*, N. S. 4: 1–30 (1939).
34. Meyer, R., "Die Haloerscheinungen." *Probl. kosm. Physik*, Bd. 12. Hamburg, H. Grand, 1929.
35. —— "Der 'Grüne Strahl.'" *Meteor. Z.*, 56: 342–346 (1939).
36. —— "Optische Erscheinungen in der Atmosphäre" in *FIAT Rev. German Sci. 1939–1946, Meteorology and*

10. Consult articles by W. E. K. Middleton, Z. Sekera, S. Fritz, and E. O. Hulburt.

11. See H. and M. Neuberger, "Color Memory in Cyanometry." *Bull. Amer. meteor. Soc.*, 24: 47–53 (1943), and references there.

Physics of the Atmosphere, R. Mügge, sen. author. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948. (See pp. 73–81)

37. Miller, A., *Investigations of the Apparent Shape of the Sky*. B. S. Thesis, The Pennsylvania State College, October 1943.
38. —— and Neuberger, H., "Investigations into the Apparent Shape of the Sky." *Bull. Amer. meteor. Soc.*, 26: 212–216 (1945).
39. Neuberger, H., "Über die Beziehung zwischen der Sonnentätigkeit und dem Auftreten von Haloerscheinungen." *Beitr. Geophys.*, 51: 343–364 (1937).
40. —— "Some Remarks on the Problem of the Dark Segment." *Bull. Amer. meteor. Soc.*, 21: 333–335 (1940).
41. Peppler, W., "Über Glorien und Halos." *Z. angew. Meteor.*, 56: 173–186 (1939).
42. Pernter, J. M., und Exner, F. M., *Meteorologische Optik*, 2. Aufl. Wien u. Leipzig, W. Braumüller, 1922.
43. Pozdèna, R., "Das Funkeln der Sterne und zwei damit zusammenhängende, bisher unzureichend gelöste Probleme." *Beitr. Geophys.*, 41: 203–208 (1934).
44. Putniņš, P., "Der Bogen von Parry und andere unechte Berührungsbogen des gewöhnlichen Ringes." *Meteor. Z.*, 51: 321–331 (1934).
45. Ramachandran, G. N., "The Theory of Coronae and of Iridescent Clouds." *Proc. Ind. Acad. Sci.*, (A) 17: 202–218 (1943).
46. Riggs, L. A., and others, "Photographic Measurements of Atmospheric Boil." *J. opt. Soc. Amer.*, 37: 415–420 (1947).
47. Schwerdtfeger, W., "Wasserglorien und Eisglorien." *Meteor. Z.*, 55: 313–317 (1938).
48. Siedentopf, H., und Holl, H., "Helligkeitsabfall und Farbänderung während der Dämmerung." *Reichsber. Physik (Beih. z. Phys. Z.)* 1 (1): 32–40 (1944).
49. Škreb, S., "Zur Kritik der sogenannten Referenzfläche." *Meteor. Z.*, 50: 310–311 (1933).
50. Smosarski, W., "Dämmerungsfarben-Intensität in den Jahren 1913–1936." *Beitr. Geophys.*, 50: 252–263 (1937).
51. —— "Intensité et polarisation de la lumière atmosphérique." *Ann. Géophys.*, 2: 1–18 (1946).
52. Visser, S. W., "Sunspots and Halos at Batavia." *Beitr. Geophys.*, (Köppen-Band I), 32: 192–201 (1931).
53. —— "Die schiefen Bogen der Gegensonne." *Meteor Z.*, 53: 336–340 (1936).
54. —— and Verstelle, J. T., "Groene Straal en Kimduiking." *Hemel en Dampk.*, 32: 81–87 (1934).
55. Wahl, E., "Die Bildruhe bei astronomischen Beobachtungen, ein Turbulenzkriterium." *Beitr. Geophys.*, 58: 370–384 (1942); and 59: 49–73 (1943).
56. Wegener, K., "Bemerkungen zur Refraktion." *Beitr. Geophys.*, 47: 400–408 (1936).
57. Whipple, F. J. W., "Rainbow with Vertical Shaft." *Meteor. Mag.*, 71: 259–260 (1936).
58. Woolard, E. W., "The Geometrical Theory of Halos." *Mon. Wea. Rev. Wash.*, 64: 321–325 (1936); 65: 4–6, 55–57, 190–192, 301–302 (1937); 69: 260–262 (1941).
59. Wünschmann, F., "Über die Konstitution der Aerosphäre und die astronomische Inflexion in ihr." *Beitr. Geophys.*, 31: 83–118 (1931).

POLARIZATION OF SKYLIGHT

By ZDENĚK SEKERA

University of California, Los Angeles

Introduction

Since the discovery by Arago in 1809 that the light of the clear sky is polarized, interest in the problems of skylight polarization has varied greatly. At first, attention was concentrated more on the development of a suitable measuring technique to determine the magnitude of the polarization and its distribution over the sky. Arago first discovered a neutral point, named after him the *Arago point*, where the polarization disappears, about 20° above the antisolar point. The other neutral points were discovered in 1840 about 20° above the sun by Babinet and below the sun by Brewster. Because of the simplicity of the determination of neutral points, more attention was paid later on to their study and soon a complete picture of diurnal, seasonal, and secular variations in the position of the Arago and Babinet points was obtained.

Arago's discovery of a point of maximum polarization 90° from the sun in the sun's vertical was followed during the next two or three decades by studies of the variability of the maximum polarization at this point (Bernard, Rubenson, Crove, Cornu, etc.).

For several reasons, interest in atmospheric polarization culminated at the end of the last century. The photopolarimeter, constructed by Cornu in 1882, represents the highest point in the development of the visual measurement of polarization. The accumulated results of polarization measurements gave rise to a series of attempts to explain the observed facts theoretically, culminating in Lord Rayleigh's theory of molecular scattering.

The famous eruption of Krakatao (1883) showed the extraordinary sensitivity of skylight polarization to the presence of volcanic dust in the upper atmosphere. For several decades thereafter, the investigation of polarization was considered almost exclusively as a suitable tool for the study of perturbations of a similar kind. But since the last eruption of Katmai in 1912 no anomalies of this type occurred, and the interest in problems of atmospheric polarization rapidly decreased. Smaller fluctuations in the polarization were studied and the measurements were extended to rather narrow spectral ranges. Surprisingly, a great discrepancy was found between the results of different authors, and the only possible explanation for this is a large variability of the dispersion of polarization (variation with the wave length) with the size and number of scattering particles. As these quantities vary greatly with local conditions (weather, season, etc.), the corresponding variations in polarization follow. But if the presence of scattering particles of different size is admitted, the question arises whether it is the secondary scattering or the presence of larger particles (excluded by the assumptions of Rayleigh's theory) which is responsible for the observed deviations of the atmospheric polarization from expectations based on theory.

After Ahlgrimm's successful attempt to compute the effect of secondary scattering, in which he explained most of the facts better than he explained the observed variations, Milch in 1924 presented a theory in which the secondary scattering was neglected and the effect of larger particles was considered responsible for the deviations from Rayleigh's theory. The close relationship between Linke's turbidity factor and the degree of polarization, predicted by Milch's theory, was demonstrated in 1934 by Blickhan from simultaneous observations of polarization and turbidity. But he obtained much better agreement between the observed and theoretical values of the maximum polarization when he considered the effects of both secondary scattering and the presence of larger particles.

Even though Blickhan's results clearly pointed the way for further investigations, no appreciable increase of interest in this direction has followed. The reason for a recent slight increase of interest in problems of skylight polarization is quite different. First, the introduction of objective methods, such as the photoelectric techniques in photometry, showed the possibility of more accurate and systematic measurements. Then, after the first attempts to use optical methods for the exploration of the upper atmosphere, attention was brought to the polarization of the skylight during twilight and during the night. The great technical difficulties in the measurement of the extremely low intensities of the skylight during these hours led even to the use of searchlight beams in the study of atmospheric scattering. But since, during twilight, direct illumination from the sun is less and less intense, the secondary and multiple scattering become more and more important. And just at the present moment when there is a need for computation of the effect of secondary and multiple scattering, recent research in theoretical astrophysics offers help. The scattering by free electrons produces polarization of stellar radiation, the theoretical study of which led Chandrasekhar to develop an excellent method for computing multiple scattering exactly, suitable for application to the problems of atmospheric polarization. Thus we are now in a position to use a highly developed modern experimental technique, together with excellent theoretical tools for solving many problems of skylight polarization in such a manner that very useful information can be obtained.

Measurement of Skylight Polarization

In the measurement of skylight polarization, there occur two different problems: measurement of the degree of polarization and measurement of the position of neutral points.

The visual measurements of the degree of polarization were highly developed by Cornu in his photopolarimeter [17], and slightly improved by Martens [42]. This photopolarimeter contains a Wollaston prism as polarizer and a Nicol prism as analyzer. Both prisms can be rotated around the same axis and their mutual position and the position of the polarizer can be read off.

With the analyzer fixed 45° to the principal axis of the Wollaston prism the plane of polarization is determined by the position of the Wollaston prism in which both halves of the field in the eyepiece have the same intensity. The plane of polarization is then inclined 45° from the principal axis of the Wollaston prism. If the Wollaston prism is now rotated by 45°, one half of the Wollaston prism transmits the intensity normal to the plane of polarization, the other the intensity in the plane of polarization. While the Wollaston prism is kept fixed the Nicol prism is rotated until both halves of the field have the same intensity. The degree of polarization P is then equal to $\cos 2\omega$, if ω is the angle between the principal plane of the Wollaston and of the Nicol prism, usually readable by means of a scale outside of the instrument.

The precision of this method was discussed by Smosarski [62] and found to be most accurate for large values of P; for small values of P this method requires some modification. Errors due to incorrect settings of the polarizer or of the analyzer can be eliminated by taking successive readings with the prism in the position 90°, 180°, and 270° from the original position. In this way a precision of 1–2 per cent can easily be reached, provided the illumination of the field in the eyepiece is sufficient to enable one to distinguish the unevenness of its halves. Another disadvantage is a relatively long time interval (about six minutes) necessary for all settings and readings. Because of the failure of this method in the case of rapid fluctuations in polarization or of inadequate illumination (during twilight or night), visual methods are being replaced more and more by objective methods.

Because of the cumulative effect during longer exposures, photographic photometry is often used when the illumination is inadequate. By means of a Wollaston prism a double picture of the measured field is obtained, and the degree of polarization is computed from the ratio of the intensities of the separate pictures. The precision of this method is limited by the precision in setting the principal section of the prism in or normal to the plane of polarization and in keeping it in this direction during the exposure. Especially during the night, when the illumination is very weak, this procedure is very difficult. Another limitation of the photographic method is the great variability of the photographic material, which makes necessary a special sensitometric arrangement for the determination of a characteristic curve for each exposure. For exact measurement, a standard intensity scale must be simultaneously exposed on the plate or film, and after development, the characteristic curve from the measured intensities is determined by a visual or photoelectric photometer. This procedure, unfortunately omitted by several authors, makes photographic polarization measurement rather complicated without any gain in precision over the visual method (the accuracy seldom exceeds 5 or 10 per cent). However, if this procedure is not followed, the results are quite unreliable. On the other hand the photoelectric method seems to be more convenient for an objective polarization measurement. It offers not only the possibility of a fast and continuous measurement but also of measurement at low intensities. The continuous measurement can easily be made by measuring the intensity of light passing through a rotating Nicol prism or other analyzer [57]. If the prism rotates around its axis with a constant angular speed ω, then the intensity of the photoelectric current from the photocell situated behind the prism is proportional to $\frac{1}{2}I_n + I_p \cos^2 \omega t$, if the time is counted from the moment when the principal plane of the prism (the plane of transmitted vibrations) is normal to the plane of polarization of the measured partially polarized light, the polarized and unpolarized components of which have intensities I_p and I_n, respectively. For light of greater intensity the current can be recorded continuously, and the degree of polarization determined from two intensities (maximum and minimum) if the period and the decrement of the galvanometer or other recording system used is known. If the rotation of the prism is uniform and sufficiently slow, the direction of the polarization plane can also be determined by a very simple arrangement. By a suitable choice of the recording system even very rapid fluctuations in the polarization can be studied in this manner. The measurement of the polarization for low intensities can also be made without any basic difficulty if very sensitive photocells (photomultipliers) are used or if the photoelectric current is properly amplified. Since a-c amplification is more effective, it is desirable to produce photoelectric alternating current, for which purpose fast rotation of the prism can conveniently be used.

Because the spectral sensitivity of the human eye differs from that of the photocell or the photographic material, the results of the objective methods will in general be different from that of the visual measurement. It can be shown [58] that, if there is no dispersion of the polarization, that is, if P is constant for all wave lengths, there will be no difference in the results of these different methods. But if the dispersion occurs, that is, P is different for different wave lengths, the difference between P measured by different methods will be greater, the greater the dispersion of the polarization. Since the dispersion depends upon the turbidity of the atmosphere, the difference between the results obtained by the different methods may vary appreciably from day to day.

The measurement of the position of neutral points is

much simpler. Their elevation above the horizon is measured by standard procedures (usually by a pendulum quadrant or a theodolite); for finding the neutral point any type of polariscope can be used. The most convenient type is the Savart polariscope and its modifications. The main part of the polariscope is Savart's double plate (two plates of the same thickness cut under 45° from the optical axis of a quartz or any uniaxial crystal; one of them turned through a right angle from the other). If a polarized light passing through the double plate is observed through an analyzer with the plane of polarization bisecting the angle between the principal planes of transmittance of the plate, parallel color fringes appear. They have a dark or bright central band, depending on whether the incident light is polarized at right angles or parallel to the plane of transmittance of the double plate. The fringes disappear if the incident light is unpolarized. The modifications of Savart's polariscope differ with the type of analyzer used. In the original model a tourmaline plate was used. Its great disadvantage was a strong absorption resulting in a dark green color of the field. By using a Nicol or similar prism this disadvantage can be removed, but the field is then very small. Much larger fields and an extraordinary brightness of fringes can be reached in Voss's modification [69] with a Wollaston prism as analyzer. By a suitable adjustment of the thickness of the double plate and the Wollaston prism, the deviation of the ordinary and extraordinary rays emerging from the prism can be made exactly equal to the angular distance of the interference fringes. In this way the fringes in the ordinary and extraordinary system of rays coincide and their intensity is doubled. Because of its great luminosity the Voss polariscope is very useful for measuring neutral points late after sunset. Its colorless field makes it particularly useful for measurements within a narrow spectral zone. The advantage of a colorless field can also be achieved by using a polaroid plate as analyzer [45].

If the polariscope is set up with fringes parallel to the sun's vertical in the vicinity of a neutral point, the dark central band above the point continues as a bright one below with an interruption in the middle in the exact position of the neutral point. The elevation of this point is then measured. The position of the interruption in the fringes can also be determined photographically [6].

Distribution and Magnitude of the Polarization over the Sky

As already mentioned, the degree of polarization of skylight reaches its maximum in the sun's vertical, 90° from the sun. Mean values of a large number of measurements of the polarization at this point taken in different years and at different places, agree relatively well, showing a decrease of the degree of polarization with increasing elevation of the sun (Fig. 1). Measurement of polarization at the zenith was introduced by Jensen [31] and performed by several authors because of the simplicity of having a fixed position of the observed direction independent of the sun's elevation. With the sun at the horizon, the zenith coincides with the point of maximum polarization. With the sun above the horizon, the polarization at the zenith decreases rapidly as the point of maximum polarization descends towards the horizon. The daily variation of the polarization at the zenith has thus the same character as that of maximum polarization, but with a much larger range of variation, as may be seen in Fig. 2. The measurement of polarization at the zenith, extended for negative sun's elevations h_s, gives an interesting result: the maximum polarization at the zenith is reached for

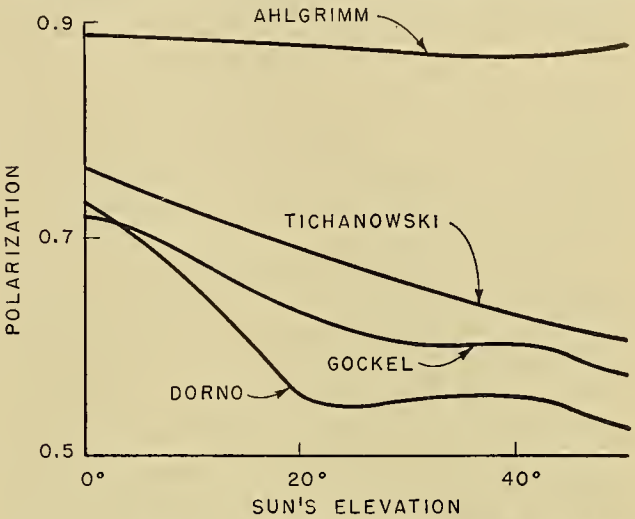


Fig. 1.—Polarization at the point of maximum polarization (in the sun's vertical, 90° from the sun) for different sun's elevations h_s. Observed values (Dorno, Gockel, Tichanowski) compared with theoretical values (secondary scattering according to Ahlgrimm).

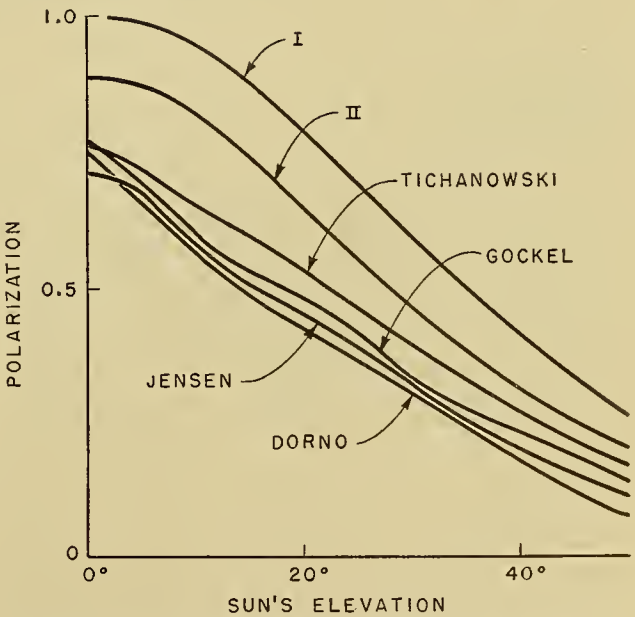


Fig. 2.—Polarization at zenith for different sun's elevations h_s. Theoretical values (I—Rayleigh's theory of primary scattering, II—secondary scattering according to Ahlgrimm) compared with observed values (Tichanowski, Gockel, Jensen, Dorno).

a sun's elevation between $-2°$ and $-4°$. The increase of polarization for $h_s < 0$ was found to be more rapid, the lower the value of the polarization at $h_s = 0$.

The distribution of the polarization over the sky was studied extensively by Dorno [22]. In a stereographic projection with the sun at its pole, the lines of equal polarization are nearly concentric circles around the sun. For corresponding P, the circles on the sun's side are closer to the equator (the line of maximum polarization) than on the antisolar side, showing a slight asymmetry of P in the sun's vertical, a fact studied and proved by Smosarski [62]. The distribution—surprisingly—varies very little with the elevation of the sun above the horizon.

The position of the plane of polarization was also measured and, at first, lines were drawn parallel to the direction of this plane. Later on they were replaced by the lines connecting points with the same inclination of the plane of polarization to the vertical (polarization isoclines). Since the inclination 45° to the vertical, called also a neutral line or Busch lemniscate, can easily be measured by the interruption of the vertical fringes in a polariscope (similar to the neutral points), it was studied extensively by Dorno [22] and Mentzel. Mentzel's measurements were recently revised by Dalhkamp and Kantus [20] and discussed with respect to the possibility of using the shape or the area inside the Busch lemniscate as a measure of the turbidity.

More attention has been devoted to the measurement of the position of neutral points than to the measurement of the degree of polarization. The reason for this, besides the great simplicity of the measurement, is the great variability of these positions, and their greater sensitivity to the turbidity of the atmosphere. The measurements mostly relate to the Arago and Babinet points; the position of the Brewster point has not been studied systematically because it is difficult to measure (below the sun, close to the horizon). The distance of the Arago point from the antisolar point and of the Babinet point from the sun vary in a characteristic way for small positive and negative elevations of the sun. If the distances of these points (A- and Ba-points) are plotted against the sun's elevations, then the curve for the A-point shows a minimum (or a secondary minimum) for small negative h_s, while—under normal conditions—the Ba-point curve shows a secondary maximum (*cf.* Fig. 3). These extremes in both curves are followed by a rapid increase for larger negative h_s. The position and the values of those extremes (for the Ba-point, even the whole character of the curve) change greatly with the turbidity; these quantities were therefore suggested by Jensen [30] as reliable indicators of atmospheric turbidity. Comparing the mean values for winter and summer from a period with a fairly normal condition, Jensen concluded that, with increasing turbidity, (1) the antisolar distances of the A-point increase, (2) the difference between the maximum and the secondary minimum for the A-point increases, and (3) the position of the minimum in the A-point curve is shifted to the negative h_s. The variations in the distances of the Ba-point do not follow such simple rules, being more sensitive to the conditions at much higher levels which are unaffected by the seasonal variations. Neuberger [45] studied the interrelationship between the extremes in the A-point curve and found a very high correlation ($+0.95 \pm 0.01$) between the difference in (2) and the distance of the A-point for $h_s = 10.5°$ or 13.5°, suggesting that a single measurement of the A-point distance for these values of h_s can be used as an indicator of the turbidity. When he compared the distance of the A-point at $h_s = 10.5°$ with the direct measurement of solar radiation, Neuberger found an increase of this distance with decreasing intensity of solar radiation; this would agree with the statement in (1) above.

As another indicator of turbidity, the difference between the points of the A- and Ba-point curves for the same h_s can be used. This difference increases with increasing turbidity, as is clearly shown in Table I, where the values of this difference, measured in a turbid atmosphere (Hamburg, Arnsberg), are compared with those measured in a much less turbid atmosphere (Davos).

Table I. Difference Between A-point and Ba-point Distances (in degrees)

	$h_s = 5.5°$	$h_s = 4.5°$	$h_s = 3.5°$	$h_s = 2.5°$	$h_s = 1.5°$	$h_s = 0.5°$	$h_s = -0.5°$
Hamburg, 1909–11	5.9	5.0	4.1	3.3	2.3	1.4	0.9
Arnsberg, 1909–11	6.3	5.6	4.6	3.8	2.9	2.1	1.6
Davos, winter 1911	0.1	0.3	0.5	0.5	0.1	−0.1	−0.2
Davos, spring 1912	0.9	0.0	1.0	0.6	0.1	—	—

Quite distinct is the effect of ground reflection on the curve of A-point distances. The maximum, which at a land station is usually very flat and is observed around $h_s = 12.5°$, shifts to smaller sun's elevations and decreases in value in the vicinity of large water surfaces [30]. If the reflection is strong, new neutral points appear, either below the ordinary points (as observed by Rubenson [54] below the Ba-point, and by Jensen [3] and Neuberger [44] below the A-point) or on both sides of the sun at the same elevation (Soret [63]). In a polariscope the fringes are visible even over the water surface, in the sun's vertical, with a dark central band (positive); in other directions they show a bright central band (negative). In a position closer and closer to the observer as the direction approaches 90° from the sun's vertical, the negative fringes change rapidly into the positive, suggesting the existence of a series of neutral points, or better, the existence of a transition zone (*Umkehrzone*, Jensen [3, 30]) between the negative po-

larization (the plane of polarization horizontal) on the horizon and the positive polarization (the plane of polarization vertical) over the water surface around the observer. Similar phenomena over land were observed by Brewster as early as 1841.

The biggest anomalies in the described distribution of polarization were observed after the volcanic eruptions of 1883–1885, 1902–1903, and 1912–1914. The effect of the volcanic dust present in the atmosphere could be noticed in the extraordinarily low values of the degree of polarization. In 1884, Cornu [18] observed the rapid decrease of the maximum polarization from 0.75 to less than 0.48; Dorno's mean values for the zenith and $h_s = 0°$ were $P = 0.557$ for 1913 and $P = 0.739$ for 1915. A very rapid increase of P during twilight also appeared (Kimball [38]). Much larger effects could be observed in the positions of neutral points.

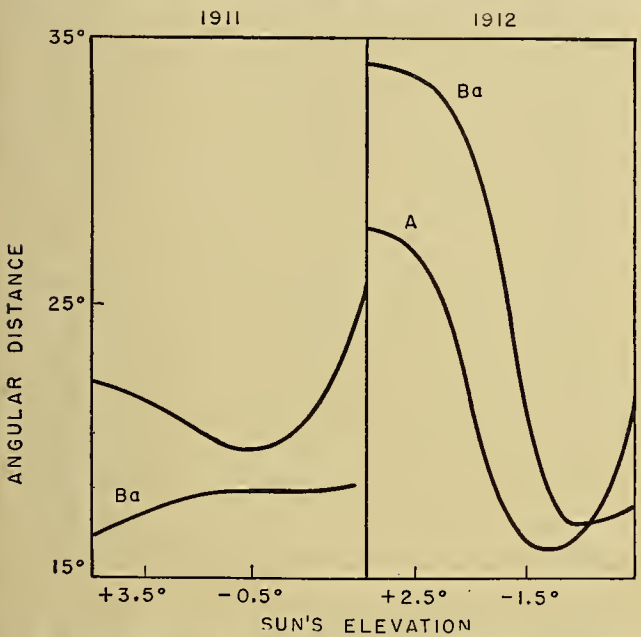


FIG. 3.—Distance of the Babinet point (*Ba*) from the sun and of the Arago print (*A*) from the antisolar point for different sun's elevations h_s. (Normal conditions—1911; after volcanic eruption of Katmai—1912).

Besides the ordinary *A*- and *Ba*-points, Cornu [18] observed four neutral points symmetrically situated at the same elevation on both sides of the sun and the antisolar point. The distances of the *A*-point and *Ba*-point increased; the largest increase, however, was observed in 1902, and was more pronounced for the *Ba*-point, as was also observed to be true in 1912. In Fig. 3 the mean values are compared for years with and without this effect; with respect to the *A*-point, the effect mentioned above, namely the increase of the distance, is clearly shown and the shift of minimum towards larger solar depressions also appears. The effect on the *Ba*-point curve is so great that its character is completely changed; the curve is shifted to the other side of the *A*-point curve.

Theory of Skylight Polarization

The first correct step toward the explanation of skylight polarization was made by Lord Rayleigh [50] in 1871. He explained skylight polarization as the scattering of sunlight on molecules and submicroscopic particles with diameters much smaller than the wave length of the incident ray of light. If it is assumed that the scattering process takes place only once (primary scattering) and if refraction is neglected, the degree of polarization of partially polarized light in the direction φ from the sun's rays,

$$P = (\sin^2 \varphi)/(1 + \cos^2 \varphi), \tag{1}$$

is a function of φ only. The maximum polarization occurs in the direction 90° from the sun, where the light is totally polarized ($P = 1$). There are two neutral points: one in the direction towards the sun, and one toward the antisolar point. Elsewhere the light is partially polarized with the plane of polarization defined by the sun, the observer, and the observed point in the sky. The theory agrees quite well with the observations with respect to the position of the point of maximum polarization and of the plane of polarization. But it does not explain the partial polarization at the point of maximum polarization and the existence of the observed neutral points. The assumptions of Rayleigh's theory are apparently not satisfied exactly in the atmosphere. The scattering particles are not isotropic and the theory should be modified for anisotropy of molecules (Cabannes [12]). The expression (1) then takes the form

$$P = (1 - a)\ (\sin^2 \varphi)/(1 + \cos^2 \varphi + a \sin^2 \varphi) \tag{2}$$

in which $a = 0.043$ is the coefficient of depolarization. Thus the maximum polarization at $\varphi = 90°$ is $P = 0.922$, but the position of the neutral points is not affected by the anisotropy of molecules.

The effect of secondary scattering, omitted in Rayleigh's theory, was studied as early as 1880 by Soret [64]. In the first approximation he considers only the light scattered by particles assumed uniformly distributed in a ring around the horizon. In the center of the ring, with the sun at the horizon, the intensity of the light scattered by all particles in the ring has the components

$$i_x = \pi b/4 = i_z/8, \qquad i_y = 3\pi b/4 = 3i_z/8, \qquad i_z = 2\pi b, \qquad (b = \text{const}). \tag{3}$$

Since the vertical component is predominant, the scattered light is negatively polarized in all directions along the horizon. In the direction 90° from the sun the component i_x is added to the primary scattered light, that is, the light is partially polarized. The neutral points are displaced to the positions where the positive polarization due to the primary scattering is compensated by the negative polarization of particles in the ring around the horizon. The distances of neutral points can be computed from relation (2) (*cf.* van de Hulst [68]). If P_1 and P denote the intensities of primary scattered light normal or parallel to the plane of polarization, and S_1 and S the intensities of secondary scattered light from the ring around the horizon, then a neutral

point appears in the sun's vertical at the elevation h, provided that in this direction

$$P_1 + S_1 = P + S. \tag{4}$$

For the sun at the horizon, $P/P_1 = \cos^2 h$. The intensities S_1 and S are normal to the direction h, and thus $S_1 = i_y$, $S = i_x \sin^2 h + i_z \cos^2 h$, and from (3)

$$S/S_1 = (1 + 7 \cos^2 h)/3. \tag{5}$$

In (5) the intensities can be expressed by the total intensities $(P_1 + P, S_1 + S)$, and then if the ratio $R = (S_1 + S)/(P_1 + P)$ is known, the elevation of the neutral point is determined from the equation

$$\sin^2 h/(1 + \cos^2 h) = R(7 \cos^2 h - 2)/(4 + 7 \cos^2 h). \tag{6}$$

The most probable value of R lies within the limits $R = 0.1$ and $R = 0.2$, which gives $h = 16.7°$ and $h = 22.4°$, in good agreement with observation.

Ahlgrimm [7] extended Soret's computation for arbitrary solar elevation, neglecting the extinction and assuming only that the distribution of scattering particles is the same in all azimuths. The unknown distribution of scattering particles with height appeared in the integrand of integrals which could be evaluated by means of the transmission coefficients. Using the measured values by Abney,[1] Ahlgrimm was able to compute the degree of polarization due to the primary and secondary scattering in any arbitrary direction. The values of maximum polarization as a function of solar elevation are reproduced in Fig. 1, values of the polarization in the zenith in Fig. 2.

Consideration of secondary scattering thus brings a great improvement in the qualitative agreement of the theory with observations under normal conditions, but a quantitative agreement still cannot be reached. Tichanowski [66] extended Ahlgrimm's computations, considering the anisotropy of molecules and even the reflection from the ground, but without any quantitative improvement. The effect of atmospheric extinction and refraction was taken into account by Link [39]. In such a case the integration for secondary scattering cannot be performed in any analytic form and requires a tedious quadruple numerical or graphical quadrature. Unfortunately the computations were made for the depressions of the sun for which no observations are available yet.

Comparison of the theoretical curve of the *Ba*-point with that obtained during the period after volcanic eruption suggests the presence of another mechanism which may be even much stronger than the effect of secondary scattering. The appearance of Bishop's ring and other twilight phenomena proved the presence of larger particles (according to Pernter [47] of a diameter from 3.2 λ to 6 λ) than assumed in Rayleigh's theory. The theoretical and experimental investigations of the scattering by particles of such a size, by Schirmann [55, 56] and later by Blumer [10], showed the possibility of neutral points already in the primary scattered light, and led Milch to the idea of using the presence of larger particles as an explanation of the deviation of the observed values from those given by Rayleigh's theory. Neglecting the effect of secondary scattering, Milch developed the following expression for the degree of polarization:

$$P = \frac{m_0 p + \mu_0 \psi f(h_s, T)}{m_0(p + n) + \mu_0(\psi + \nu) f(h_s, T)}, \tag{7}$$

where m_0 and μ_0 denote the number of molecules and large particles per unit volume at the ground; p and ψ, the intensity of the polarized part of the light scattered by molecules and by large particles respectively; $p + n$ and $\psi + \nu$, the total intensity in both cases of scattering; and finally, $f(h_s, T)$ is a function of the sun's elevation and the turbidity coefficient computed from the extinction values for different turbidities. This expression can explain all observed variation of the polarization with turbidity, but its use for the computation of P is very limited by the lack of knowledge of the functions ψ and ν. Assuming that at the point of maximum polarization ψ can be neglected with respect to p, Milch determined ν from the measured value of P for a given h_s, and computed the variation of maximum polarization with h_s. But the computed values showed a systematic deviation from the observed P.

In continuation of Milch's work, Blickhan [9] studied the correlation between the turbidity coefficient and the maximum polarization measured simultaneously. The values of P lie along a hyperbola $P = A/(B + T_k)$ in a (P, T_k) – diagram (T_k is the turbidity coefficient for short-wave radiation, and A and B are constants), in good agreement with the simplified formula (7). Extrapolating P by means of this empirical formula for an atmosphere without large particles ($T = 1$), he obtained a value of P for $h_s = 50°$, which is larger than that obtained for $h_s = 32.5°$, in agreement with Ahlgrimm's computation. From the difference between the extrapolated values and those computed by Ahlgrimm, assuming further that the light reflected from the ground is not polarized, Blickhan was able to compute the albedo $a = 0.132$, which is in good agreement with Dorno's values. For the computation of the daily variation of maximum polarization he used Milch's procedure, but with the difference that in (7) he inserted for p and n the values computed by Ahlgrimm. The values obtained in this way agreed with the measured ones much better than if the effect of the secondary scattering had been neglected. Other simultaneous measurements of polarization and turbidity were made by Wörner [70]. The polarization was measured this time at the zenith and was compared with the normal values computed by Jensen [2] (Fig. 2). The degree of polarization expressed in per cent of normal values

Table II. Polarization at Zenith (per cent of normal value) as a Function of Turbidity

T	2.0–2.5	2.51–3.0	3.01–3.5	3.51–4.0	4.01–5.0	>5.0
P	114.4	112.0	104.0	102.8	93.2	68.3

1. Abney's values are reproduced in [5].

shows a close correlation with the turbidity factor, as may be seen from Table II. Jensen's normal values correspond thus to the turbidity factor 3.9.

In connection with Milch's and Blickhan's work the recent investigation of Tousey and Hulburt [67] should be mentioned. The brightness and the polarization of the daylight sky were measured at different altitudes up to 10,000 ft. The curves of polarization with height showed clearly a much slower rise after a rapid increase within the first 2000 or 3000 ft, closely resembling the distribution of the turbidity coefficient with height [41]. They compared the measured values with theoretical values obtained on the assumption that the secondary scattered light is unpolarized, but taking full account of the extinction defined by a mean value of the theoretical extinction coefficient $\beta = 0.0126$ km^{-1}. They found that a somewhat better agreement for larger distances from the sun can be obtained with β increased to 0.017, 0.018, or 0.021 km^{-1}. The systematic deviations in the vicinity of the sun, expected by the authors, are caused by the assumption above, eliminating the neutral points around the sun. The increase of β proves without doubt the presence of larger particles, sufficient to increase the theoretical scattering by about 35 per cent. The increased values of β mentioned above correspond to the turbidity coefficient $T = 1.35$, 1.43, and 1.67, respectively. These values are in very good agreement with the measured values mentioned above [41], indicating the real presence of larger particles rather than a systematic error in taking too wide a spectral range, as suggested by van de Hulst [68]. In the theoretical computation the reflection by the ground was taken into consideration and the variation of the maximum and zenithal polarization due to the different values of the albedo is given in Table III.

Table III. Effect of Ground Reflection on Skylight Polarization

Albedo a	0.0	0.1	0.2	0.4	0.6
Tousey and Hulburt [67]					
P (maximum): $h_s = 30°$	0.85	—	0.755	0.667	0.612
P (at zenith): $h_s = 30°$	0.527	—	0.482	0.439	0.408
Chandrasekhar [16]					
P (maximum):					
$h_s = 0°$	0.918	0.903	0.885	—	—
$h_s = 13.9°$	0.906	0.860	0.801	—	—
$h_s = 39.8°$	0.906	0.796	0.673	—	—
P (at zenith):					
$h_s = 0°$	0.918	0.903	0.885	—	—
$h_s = 13.9°$	0.824	0.779	0.727	—	—
$h_s = 39.8°$	0.392	0.360	0.315	—	—

The mean value of the measured albedo, $a = 0.20$, was taken for the computation, and for this value the theoretical degree of polarization at zenith, $P = 0.482$ for $h_s = 30°$, and $P = 0.724$ for $h_s = 15°$ may be compared with the values of P computed by Ahlgrimm (0.469, 0.738); and for $h_s = 25°$ the observed value $P = 0.58$ may be compared with the theoretical value $P = 0.572$ (Ahlgrimm 0.558).

From the discussion above it is quite evident that a better quantitative agreement between measurement and theory can be achieved when the original Rayleigh theory is extended by a consideration of (1) the effects of multiple (at least secondary) scattering, extinction, and reflection by the ground, and (2) the effect of the presence of large scattering particles. The effects mentioned first can lower Rayleigh's theoretical values to the observed values and explain the existence of neutral points in observed positions; but for the explanation of the great variety and magnitude of diurnal, interdiurnal, seasonal, and secular variations the highly variable content of larger particles in the atmosphere must be considered. This is more evident if the dispersion of polarization is taken into account. The presence of larger particles can best be taken into account in the quantitative analysis, however, by separating and subtracting the effect of molecular scattering as a simpler and more nearly constant factor. For this purpose the recent theoretical investigations of similar problems in astrophysics offer excellent help. In a very elegant way, Chandrasekhar succeeded in reducing the exact solution of quite general multiple scattering to a solution of two relatively simple integral equations of a form suitable for successive iteration. Once the solution of these equations is known, the exact problem is solved. The great advantage of this method is not only that the extinction is very simply taken into account, but mainly that the effect of ground reflection can be included, as proposed by van de Hulst, and that the method can be extended for a more general law of scattering than in Rayleigh's theory [13, 14, 15].

Chandrasekhar has recently accomplished the numerical computation of the effect of multiple scattering and of the ground reflection in the skylight polarization for a special value of the optical thickness $\tau = 0.15$, corresponding to $\lambda = 450$ $\mu\mu$ under normal conditions [16]. The reflection on the ground affects the position of neutral points very little, in agreement with observation (Neuberger [46]). A much larger effect is noticeable in the degree of polarization at zenith or at 90° from the sun. It is evident from Table III that the ground reflection is responsible for the daily variation of the maximum polarization, namely for the decrease of P with the sun's elevation, in the sense of Fig. 1. The theoretical values for P are still much higher than the measured ones.

The observed distances of the neutral points are also much higher than the theoretical values obtained by Chandrasekhar (for $h_s = 0°$: A-point, 19.4° and Ba-point, 19.4°; for $h_s = 13.9°$: A-point, 20.9° and Ba-point, 18.7°). However a remarkable agreement was found between the theoretical shape of the neutral lines and the shape of neutral lines observed by Dorno [22]. Larger scattering particles apparently affect primarily the magnitude of the polarization, and the position of the neutral points, but have only a slight effect on the position of the plane of polarization. This fact is an interesting aspect of the physics of scattering by larger particles and as such should be studied more extensively.

Chandrasekhar's method of exact evaluation of the molecular scattering makes possible a quantitative

study of the presence and the nature of larger scattering particles. If the exact values of molecular scattering are subtracted from the observed values, the remaining part is the effect of larger particles, and it is quite evident that, for such a purpose, observations should be used in which the deviation from the theory is most pronounced. This seems to be the case in the dispersion of skylight polarization.

Dispersion of Atmospheric Polarization

The discussion of polarization in the preceding section refers to "white" light, as it is observed directly by the human eye. The measurement of polarization in much shorter spectral ranges was started very early in skylight investigations. In 1884, Cornu [18] found the degree of polarization to be different for different colors. During this period of volcanic anomalies the polarization at shorter wave lengths was much larger than for longer wave lengths. The dispersion of polarization has been studied since that time by several au-

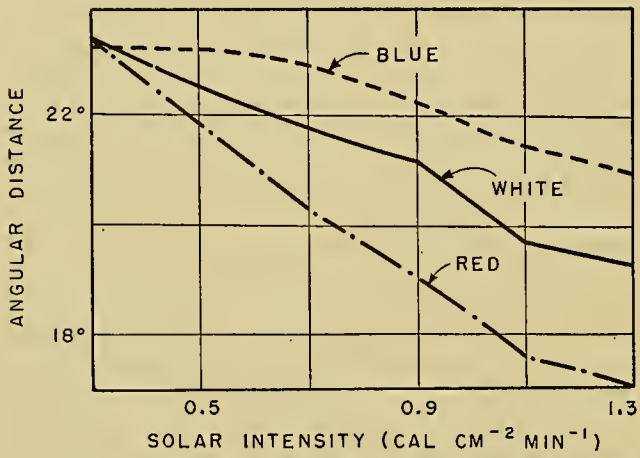


FIG. 4.—Distance of the Arago point from the antisolar point for $h_s = 10.5°$, in different colors for different intensities of solar radiation (*according to Neuberger*).

thors with one result confirmed by all, namely, the great variability of the dispersion with the turbidity, and thus with the weather, location, etc. Since the turbidity was not actually measured (except in the latest studies), the different results which have been attained are very difficult to compare. If the relativity of the terms "pure" and "turbid" atmosphere is admitted, then the contradictory results of different authors, even recently considered inexplicable [3], can be ordered to show a definite trend, confirmed by theoretical considerations. For very low turbidity (high altitude) the degree of polarization at the point of maximum polarization increases with the wave length [65]. With increasing turbidity, the maximum is shifted to the central part of the visible spectrum and the difference between polarization in the red part of the spectrum (P_r) and in the blue part (P_b) is decreased (for "pure air" in Gockel's definition the difference $P_r - P_b$ becomes smaller than the errors of observation). With still larger turbidity the maximum is shifted to the blue part, that is, $P_b > P_r$ [22, 25, 34, 48, 65].

The measurement of distances of neutral points in different colors was started by Jensen [1] in 1909. The results of his measurements were confirmed by Busch [11], who found that the distances of the A-point were larger the shorter the wave lengths. During the abnormal period 1912–14, just the opposite was found. This result was recently confirmed by Neuberger [45, 46] in his measurement of the A-point at $h_s = 10.5°$. Along with the A-point distances, the intensity of solar radiation, the blue color of the sky, etc., were measured and the variations of A-point distances with the turbidity were proved to be as shown in Fig. 4.

The dispersion of polarization is thus very sensitive to the degree of turbidity and could be used as another indicator of turbidity. But it can also be used for answering the question about the prevailing effect of the molecular scattering, or of the presence of larger particles in the atmosphere. It is evident that in Rayleigh's theory of primary scattering there is no dispersion, since the degree of polarization P is independent of λ. If it is assumed with Milch that the light scattered by large particles is unpolarized, the component of the total intensity due to the scattering on large particles can be expressed in the form $F_g = \lambda^{-\alpha} Nf(\varphi)$, where α decreases from 3.5 to 1 with increasing size of particles, N denotes the number of such particles per unit volume, and φ is the scattering angle. It can easily be shown [59] that the sign of the difference $P(\lambda) - P(\lambda_0)$ is determined by the sign of the expression

$$[p/P(\lambda_0) - 1](1 - \chi^4) + (\chi^4 - \chi^{\alpha-4})F_g(\lambda_0)/I_g(\lambda_0), \quad (8)$$

where p denotes the degree of polarization in Rayleigh's theory of primary scattering, given in equation (1), $\chi = \lambda_0/\lambda$, and I_g is the total intensity of the component due to the primary scattering. If the turbidity is small, $F_g \to 0$ and $P(\lambda) > P(\lambda_0)$ whenever $\lambda > \lambda_0$. With increasing turbidity ($F_g > 0$), the second term in (8), having an opposite sign from the first one, reduces the difference $P(\lambda) - P(\lambda_0)$ and for a sufficiently large turbidity reverses the sign of $P(\lambda) - P(\lambda_0)$. The variation of the dispersion with increasing turbidity can thus be explained in agreement with observation. But the discussion of the distance of the neutral points for different wave lengths shows clearly that the assumption made by Milch, that the light scattered by large particles is unpolarized, is not justified. If the light scattered by large particles is assumed to be unpolarized, the distance of neutral points is given by equation (4) written in the form

$$P_1(\lambda_0) \sin^2 h_0 = S(\lambda_0, h_0) - S_1(\lambda_0, h_0),$$

and for $\lambda \neq \lambda_0$,

$$P_1(\lambda_0) \sin^2 h = \chi^4 [S(\lambda_0, h) - S_1(\lambda_0, h)].$$

Since S_1 is actually independent of h, S can be expressed by S_1 and the ratio $S_1(\lambda_0)/P_1(\lambda_0)$ can be eliminated from these two equations. The resulting equation can be written in the form

$$\sin^2 h - \sin^2 h_0 = (\chi^4 - 1) \sin^2 h_0 \frac{5 - 7 \sin^2 h_0}{5 + 7(\chi^4 - 1) \sin^2 h_0}. \quad (9)$$

Since $h_0 < 30°$, the numerator and the denominator in (9) are always positive, and $h > h_0$ whenever $\lambda_0 > \lambda$. The distances in blue are larger than in the red part of the spectrum, in agreement with the observation made for very low turbidity. However, the computed difference $h - h_0$, for a given h_0 and given wave lengths λ, λ_0, is about twice as large as observed. What is more serious, the great variability of this difference with increasing turbidity cannot be explained by (9). The effect of increasing turbidity can be taken into consideration only by increasing h_0, but the right-hand side of (9) increases with h_0 instead of the observed decrease to negative values. This may serve as a proof of the incorrectness of the assumption made above. For a complete discussion it is necessary to include the polarized component due to the scattering by large particles also. This can be done only by using the Mie-Debye theory, as discussed in detail elsewhere [59]. By constructing a special model of the distribution, size, and optical properties (refractive index) of the large particles, the dispersion of polarization can be computed and compared with the observed values through a procedure similar to that used in the study of atmospheric haze [28, 51, 60] and thus a model of the distribution which best fits the observations can be found.

Polarization Anomalies During Twilight

The same information concerning the size, nature, and distribution of scattering particles in the atmosphere can be obtained from any deviations of observed values of skylight polarization from those to be expected from Rayleigh's theory. Particular emphasis has been placed on anomalies during twilight (because of the easily determined changes in illumination along the vertical line to the zenith) with the hope that more information can be obtained about the vertical distribution of scattering particles in this way. But the use of twilight anomalies is not as simple as it would seem. The first difficulty is the rapid decrease of the intensity of skylight, which causes serious difficulties in polarization measurement. Visual methods quickly become uncertain and are very seldom reliable for solar depressions beyond $h_s = -5°$ or $-6°$. The photographic method requires longer exposures during which the eventual fluctuations in the degree of polarization and in the position of the plane of polarization may cause large systematic errors.

In theoretical investigations the atmosphere can no longer be considered as plane-parallel, and refraction must be taken into consideration at least to the extent of estimating the limit of the earth's shadow. The ground reflection, acting for low solar depressions only on one side of the horizon, and the different extinction values in the solar and antisolar regions make the computation of the sky polarization rather complicated. With respect to the effect of secondary scattering, it is valid to offer the same criticism which was presented against the use of the zenith intensity for optical sounding of the upper atmosphere. As Hulburt [27] pointed out, the intensity of the secondary scattering increases rapidly in comparison with the intensity of primary scattering, so that for solar depressions larger than 8°, the secondary scattering from the lower level has a greater intensity than that of the primary scattering from the upper levels still illuminated by direct rays from the sun. Hulburt's estimate of this effect was based on the measured intensity of skylight near the western horizon; the presence of larger particles was thus taken into consideration. This may explain the much larger values for the ratio of the intensity of the secondary and primary scattering (I_{prim}: I_{sec} = 1:185, $h_s = -10°9'$) than computed by Link [39] (I_{prim}: I_{sec} = 2:1) under the assumption of molecular scattering only. For this reason it is quite difficult to explain the high correlation of the polarization anomalies (sudden or nonmonotonic decrease of P in the zenith for sun's depressions larger than 10°) with the changes in the ionization of the E- or F-layer, as observed by Khvostikov and a group of Russian scientists. The first attempt to explain the anomalies as being associated directly with the increase of anisotropy of the ions as compared to neutral particles [35, 36, 52] was found by Ginsburg [23, 24] to be unacceptable because of the predominant effect of the secondary scattering. The polarization caused only by secondary scattering under such conditions was recently estimated by Rosenberg [53] as being even larger than observed. The observed rotation [49] of the plane of polarization from the direction given by the position of the sun cannot be explained simply by the asymmetry in the solar illumination and should be studied more closely in relation to the problem of fluorescent luminescence or other types of emission of the night sky. For the study of the emission layer the scattering of the emitted light is very important and can be used for the determination of the height of the emission layer [8]. Since the secondary scattered solar radiation may be superimposed upon the light from the emission layer, the dispersion of polarization of the twilight or night sky could be used for separating the phenomena of the lower atmosphere from those of the upper levels. Because of experimental difficulties there is little hope for the effectiveness of this method in the near future. The only possible way of separating the intensity of the polarization produced in lower levels from the phenomena related to upper levels is to compute these quantities using the extinction coefficient and other parameters of scattering in lower levels, obtained by independent measurements. For this purpose the method of an artificial light source seems to be quite adaptable. The searchlight beam has been used and information about the type and the law of scattering has been derived mainly from the total intensity measurement [28, 29, 33, 51, 60]. More information can be obtained, however, if the measurement of the polarization is added, as has been done by Khvostikov [35]. But since the brightness of the searchlight beam decreases with the distance from the source, the secondary scattering in lower levels should be carefully taken into consideration before any conclusions are made about scattering in higher levels.

The searchlight-beam method definitely offers quite new possibilities and if properly used may contribute

largely to the solution of the problem of light scattering in the atmosphere.

Unsolved Problems and Future Research

As was shown in the preceding sections, much valuable information can be obtained from a comparison of measured and theoretical values. Therefore future research depends primarily upon the development of both experimental and theoretical methods. First, there is a definite need for better equipment for measuring skylight polarization. Such equipment should permit objective, fast, and precise measurements of the degree of polarization and of the position of the plane of polarization, in comparatively narrow spectral ranges, in all directions over the sky, with special provision for measurement of weak polarization, thus being adaptable also to the measurement of the neutral points. Speed in measurement is required not only for almost simultaneous measurements at different wave lengths and at different places in the sky, but mainly for the possibility of recording the shorter fluctuations observed in the degree of polarization [57] as well as in the position of neutral points [45, 46]. These fluctuations, not observable by the older methods or eliminated by improper smoothing, seem to precede cloud formation in a clear sky, and might be used in the study of condensation processes in incipient clouds. With sensitive photomultipliers and with a proper amplification, the sensitivity of measurement can be increased beyond that of the human eye, and observations can be extended far into the twilight or used for optical sounding by searchlight beams.

From the theoretical standpoint, Chandrasekhar's computations can easily be completed, and skylight polarization due to multiple molecular scattering can be computed, in all directions and for all wave lengths, for larger solar elevations when the atmosphere can be considered as plane-parallel. But the greatest advantage of this method is that it can be extended in such a way that almost all currently unsolved problems of atmospheric polarization can be solved. In the case of reflection at the ground, if, for example, the polarization according to Fresnel's law were taken into consideration as observed in the vicinity of large water surfaces [26], the anomalies mentioned earlier could be explained theoretically. If the Mie-Debye theory of large-particle scattering could be simplified in such a way that the corresponding scattering matrices could be computed without essential difficulties, the effect of large particles could also be taken into consideration. So far Chandrasekhar's theory has been limited to a plane-parallel atmosphere, and twilight anomalies have been unaccessible to theoretical investigation of similar type. Chandrasekhar [14, 15] already has shown the way to extend the theory for a spherical atmosphere. It is necessary only to work out the suggested method in more detail. The exact theory of twilight phenomena cannot, however, be developed without the consideration of refraction and the corresponding bending of light paths in the atmosphere. This part of the problem does not represent any special difficulty since complete refraction tables, with all the parameters necessary for such a computation, are now available and can be conveniently used for this purpose [40].

Hence, many problems quite accessible by modern facilities are open to further investigations. In the review given above, the list of problems to be solved is not exhaustive; new problems may easily arise as the study of skylight polarization progresses further. For this reason atmospheric polarization deserves more attention than it has received up to now.

REFERENCES

A detailed discussion of problems of skylight polarization will be found in references [1] through [5]. Reference [6] gives a detailed description of instruments used in polarization measurements. Other references cited in the text follow.

1. Busch, F., und Jensen, C., "Tatsachen und Theorien der atmosphärischen Polarisation." *Jb. hamburg. wiss. Anst.*, Bd. 28, (1910), Berlin, Duemruler, 1911.
2. Jensen, C., "Die Himmelsstrahlung," *Handbuch der Physik*, Bd. 19. Berlin, Springer, 1928. (See pp. 70–152)
3. —— "Die Polarisation des Himmelslichtes," *Handbuch der Geophysik*, Bd. 8. Berlin, Gebr. Bornträger, 1942. (See pp. 527–620)
4. Linke, F., "Die Theorie der Zerstreuung, Extinktion und Polarisation des Lichtes in der Atmosphäre," *Handbuch der Geophysik*, Bd. 8. Berlin, Gebr. Bornträger, 1942. (See pp. 120–238)
5. Pernter, J. M., und Exner, F. M., *Meteorologische Optik*, 2. Aufl. Wien-Leipzig, Braunmüller, 1922. (See pp. 644–725)
6. Jensen, C., "Die Apparate zur Untersuchung der atmosphärischen Polarisationserscheinungen," in Kleinschmidt, E., *Handbuch der meteorologischen Instrumente.* Berlin, Springer, 1935. (See pp. 666–693)
7. Ahlgrimm, F., "Zur Theorie der atmosphärischen Polarisation." *Jb. hamburg. wiss. Anst.*, 32:1–66 (1915).
8. Barbier, D., "Sur la correction de diffusion dans les mesures d'altitude des couches atmosphériques émettant la lumière du ciel nocturne." *Ann. Géophys.*, 1:144–156 (1944).
9. Blickhan, F., "Vergleichende Messungen der Polarisation des Himmelslichtes in Frankfurt a. M. und am Taunusobservatorium." *Beitr. Geophys.*, 42:208–227 (1934).
10. Blumer, H., "Strahlungsdiagramme kleiner dielektrischer Kugeln." I—*Z. Phys.*, 32:119–134 (1925); II—*Ibid.*, 38: 304–328 (1926).
11. Busch, F., "Beobachtungen über die atmosphärisch-optische Störung des Jahres 1912." *Meteor. Z.*, 30:321–330 (1913).
12. Cabannes, J., "Sur la diffusion de la lumière par les molécules des gaz transparents." *Ann. Phys., Paris,* (9) 15:5–149 (1921).
13. Chandrasekhar, S., "On the Radiative Equilibrium of a Stellar Atmosphere." XXI—*Astrophys. J.*, 106:152–216 (1947); XXII—*Ibid.*, 107:48–72, 188–215 (1948).
14. —— "The Transfer of Radiation in Stellar Atmospheres." *Bull. Amer. math. Soc.*, 53:641–711 (1947).
15. —— *Radiative Transfer.* Oxford, Clarendon Press, 1950. (See Chaps. IX, X)
16. —— and Elbert, D., "Polarization of the Sunlit Sky." *Nature*, 167:51–55 (1950).
17. Cornu, A., "Sur l'application du photopolarimètre à la météorologie." *C. R. Ass. franç. Av. Sci.*, Session à Limoges, pp. 267–270 (1890).
18. —— "Observations relatives à la couronne visible actuelle-

ment autour du Soleil." *C. R. Acad. Sci., Paris,* 99:488–493 (1884).

19. Dahlkamp, V., "Die Lage des Arago Punktes in Abhängigkeit von Sonnen und Himmelsstrahlung und den Dämmerungserscheinungen," *Z.Meteor.*, 1:130–138 (1947).
20. —— und Kantus, H., "Untersuchungen über die Verwendbarkeit der Polarisationisoklinen zur Bestimmung des atmosphärischen Reinheitsgrades." *Ann.Hydrogr., Berl.,* 72:25–35 (1944).
21. —— "Untersuchungen über den Reinheitsgrad der Atmosphäre mit Hilfe der neutralen Punkte der atmosphärischen Polarisation." *Z. Meteor.*, 1:253–263, 303–310 (1947).
22. Dorno, C., "Himmelshelligkeit, Himmelspolarisation und Sonnenintensität in Davos (1911 bis 1918)." *Meteor. Z.*, 36:109–124, 181–192 (1919).
23. Ginsburg, V. L., "On the Anomalies in the Polarization of Twilight." *C. R. (Doklady) Acad. Sci. URSS,* 38:301–303 (1943).
24. —— and Soboleff, N. N., "On Secondary Light Scattering in the Atmosphere and on Polarization Anomalies during Twilight." *C. R. (Doklady) Acad. Sci. URSS,* 40:223–225 (1943).
25. Gockel, A., "Beiträge zur Kenntnis von Farbe und Polarisation des Himmelslichtes." *Ann. Phys., Lpz.,* (4) 62: 283–292 (1920).
26. Hulburt, E. O., "The Polarization of Light at Sea." *J. opt. Soc. Amer.*, 24:35–42 (1934).
27. —— "The Brightness of the Twilight Sky and the Density and Temperature of the Atmosphere." *J. opt. Soc. Amer.*, 28:227–236 (1938).
28. —— "Optics of Atmospheric Haze." *J. opt. Soc. Amer.*, 31:467–476 (1941).
29. —— "Optics of Searchlight Illumination." *J. opt. Soc. Amer.*, 36:483–491 (1946).
30. Jensen, C., "Normale, gestörte und pseudonormale Polarisations-erscheinungen der Atmosphäre." *Meteor. Z.*, 49:419–430 (1932).
31. —— "Beiträge zur Photometrie des Himmels." *Metcor. Z.*, 14:488–499 (1899).
32. —— "Atmosphärisch-optische Messungen in Ilmenau." *Beitr. Geophys.*, 35:166–188 (1932).
33. Johnson, E. A., and others, "The Measurement of Light Scattered by the Upper Atmosphere from a Search-Light Beam." *J. opt. Soc. Amer.*, 29:512–517 (1939).
34. Kalitin, N. N., "Zum Studium spektraler Polarisation des Himmelslichtes." *Meteor. Z.*, 43:132–140 (1926).
35. Khvostikov, I. A., "The Optical Piercing of the Atmosphere with Projector Rays." *Izvestiia Akad. Nauk SSSR,* Ser. fiz., 10:403–414 (1946).
36. —— and Sevčenko, A. N., "Applications de la méthode polarimétrique à l'étude de la structure des couches supérieures de l'atmosphère." *C. R. (Doklady) Acad. Sci. URSS,* 13:359–363 (1936).
37. Khvostikov, I. A., and others, "Sur la liaison des anomalies de la polarisation du demi-jour avec l'état de l'ionosphère." *C. R. (Doklady) Acad. Sci. URSS,* 26:900–903 (1940).
38. Kimball, H. H., "Observations of Solar Radiation with the Ångström Pyrheliometer at Asheville and Black Mountain, N. C." *Mon. Wea. Rev. Wash.*, 31:320–334 (1903).
39. Link, F., "Die Dämmerungshelligkeit im Zenit und die Luftdichte in der Ionosphäre." *Meteor. Z.*, 59:7–12 (1942).
40. —— und Sekera, Z., "Diotropische Tafeln der Erdatmosphäre." *Publ. Obs. nat., Prague,* 14:1–28 (1940).
41. Linke, F., "Die Verwertung von Sonnenstrahlungsmessungen in Luftfahrzeugen." *Z. Geophys.*, 1:55–59 (1924–25).
42. Martens, F. F., "Über ein neues Polarisationsphotometer für weisses Licht." *Phys. Z.*, 1:299–303 (1900).
43. Milch, W., "Über den Einfluss grösserer Teilchen in der Atmosphäre auf das Polarisationsverhältnis des Himmelslichtes." *Z. Geophys.*, 1:109–117 (1924–25).
44. Neuberger, H., "Beiträge zur Untersuchung des atmosphärischen Reinheitsgrades." *Aus d. Arch. dtsch. Seew.,* Vol. 56, No. 6, 53 pp. (1936).
45. —— "Studies in Atmospheric Turbidity in Central Pennsylvania." *Penn. St. Coll. Studies,* No. 9, 36 pp., 1940.
46. —— "Arago's Neutral Point: A Neglected Tool in Meteorologieal Research." *Bull. Amer. meteor. Soc.*, 31:119–125 (1950).
47. Pernter, J., "Zur Theorie des Bishop'schen Ringes." *Meteor. Z.*, 6:401–418 (1889).
48. Piltschikoff, N., "Sur la polarization spectrale du ciel." *C. R. Acad. Sci., Paris,* 115:555–558 (1892).
49. Ponizovsky, Z. L., and Rosenberg, G. V., "On Polarization Anomalies in Scattered Light of Twilight Sky as Connected with the Condition of the Ionosphere." *C. R. (Doklady) Acad. Sci. URSS,* 37:218–220 (1942).
50. Rayleigh, Lord (J. W. Strutt), "On the Light from the Sky, Its Polarization and Colour." *Phil. Mag.*, 41:107–120, 274–279 (1871).
51. Reeger, E., und Siedentopf, H., "Die Streufunktion des atmosphärischen Dunstes nach Scheinwerfermessungen." *Optik,* 1:15–41 (1946).
52. Rosenberg, G. V., "On a New Phenomenon in the Scattered Light of a Twilight Sky." *C. R. (Daklady) Acad. Sci. URSS,* 36:270–274 (1942).
53. —— "Polarization of Secondary Scattered Light in the Case of Molecular Scattering." *Izvestiia Akad. Nauk SSSR,* Ser. geograf. i geofiz., 3:154–161 (1949).
54. Rubenson, R., "Mémoire sur la polarisation de la lumière atmosphérique." *Nova Acta Soc. Sci. upsal.*, 4:5 (1864).
55. Schirmann, M. A., "Dispersion und Polychromismus des polarisierten Lichtes, das von Einzelteilchen von der Grössenordnung der Wellenlänge des Lichtes abgebeugt wird." *Ann. Phys., Lpz.*, 59:493–537 (1919).
56. —— "Neue theoretische Untersuchungen über die Polarisation des Lichtes an trüben Medien und deren Konsequenzen für die Probleme der atmosphärischen Polarisation." *Meteor. Z.*, 37:12–22 (1920).
57. Sekera, Z., "Lichtelektrische Registrierung der Himmelspolarisation." *Beitr. Geophys.*, 44:157–175 (1935).
58. —— "Über die Vergleichsmöglichkeit der visuellen und lichtelektrischen Messung der Himmelspolarisation." *Čas. Pěst. Math.*, 67:278–287 (1938).
59. —— "Dispersion of the Atmospheric Polarization in Relation to the Size and Nature of Larger Scattering Particles." (Prepared for publication.)
60. Siedentopf, H., "Zur Optik des atmosphärischen Dunstes." *Z. Meteor.*, 1:417–422 (1947).
61. Smosarski, W., "Zur Theorie des Cornu-Photopolarimeters." *Beitr. Geophys.*, 54:235–244 (1939).
62. —— "Polarisation des Himmelslichtes im Weltpol und andere Beobachtungen." *Beitr. Geophys.*, 48:213–224 (1936).
63. Soret, J. L., "Influence des surfaces d'eau sur la polarisation atmosphérique et observation de deux points neutres à droite et à gauche du Soleil." *C. R. Acad. Sci., Paris.*, 107:867–870 (1888).
64. —— "Sur la polarisation atmosphérique." *Arch. Sci. phys. nat.*, 20:429–471 (1888).

65. Tichanowski, J. J., "Resultate der Messungen der Himmelspolarisation in verschiedenen Spektrumabschnitten." *Meteor. Z.*, 43:288–292 (1926).
66. —— "Die Bestimmung des optischen Anisotropiekoeffizienten der Luftmolekülen durch Messungen der Himmelspolarisation." *Phys. Z.*, 28:252–260 (1927).
67. Tousey, R., and Hulburt, E. O., "Brightness and Polarization of the Daylight Sky at Various Altitudes above Sea Level." *J. opt. Soc. Amer.*, 37:78–92 (1947).
68. van de Hulst, H. C., "Scattering in the Atmospheres of the Earth and the Planets" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed. Chicago, University of Chicago Press, 1949. (See pp. 49–112)
69. Voss, W., "Eine Abart des Savartschen Polariskops." *Z. InstrumKde.*, 52:332–333 (1932).
70. Wörner, H., "Messungen der Sonnenstrahlung, der Himmelspolarisation und der Blaufärbung des Himmelslichtes." *Z. Meteor.*, 3:166–169 (1949).

VISIBILITY IN METEOROLOGY

By W. E. KNOWLES MIDDLETON

National Research Council, Canada

Introduction

This discussion will be devoted to a brief study of those interrelations between the optical properties of the atmosphere and the characteristics of human vision which determine how far a given object can be seen at such and such a moment. This important distance, when referred to the ordinary objects which lie around the horizon of a meteorological observer, is called by meteorologists "the visibility"; it is not with any hope of changing this terminology, but only in the interests of logical discussion, that the same quantity, applied now to *any* object, will here be given the name *visual range*. The term is self-explanatory.

The theory of the visual range is by this time in a fairly satisfactory condition, largely because of very extensive researches conducted during World War II, especially in the United States. As in many other divisions of meteorology, however, the extreme complexity of atmospheric conditions makes it difficult to apply the available theory to actual instances, especially in the important case of visual range along inclined paths. Even the instrumentation necessary to measure the appropriate optical constants of the atmosphere has not been developed to the point where it is in general use at meteorological stations. Theory is well in front of practice.

We are unable to refer the reader to any very up-to-date summaries of the whole field; in fact to nothing later than 1941 [33, 35]. In view of our restrictions on space we must assume that at least one of these two monographs is available to the reader. It is to be hoped that before very long a more up-to-date general account will be published.

The Behavior of Light in the Atmosphere

Light, by its interaction with the atmosphere, produces many beautiful phenomena which are dealt with elsewhere in this book. Our concern here is only with the way in which it is attenuated in its passage through the air and with the manner in which it is diffused by scattering.

As far as any practical interest is concerned, we may neglect those rare occasions when the air is nearly free from particles larger than the molecules of gases, in view of the fact that the visual range in such a pure atmosphere would be several hundred kilometers at sea level. The reader may be referred to Cabannes [10] for a masterly discussion of such molecular scattering. On all occasions when the visual range is of any practical importance, by far the greater part of the effect of the atmosphere on light is produced by particles much larger than molecules, which may be thought of as the disperse phase of an atmospheric colloid or aerosol.

These particles are of many kinds, but from our standpoint the most interesting of them are the liquid droplets, generally aqueous solutions of hygroscopic substances, which in various radii from about 10^{-6} to 10^{-1} cm form the obscuring matter in haze, fog, and mist. The actual nature of the hygroscopic nuclei involved is one of the great unsolved problems of meteorology, and has become the subject of a controversy, for the details of which the reader should consult Wright [52, 53, 54, 55, 56], Simpson [43, 44] and Findeisen [21]. Whatever their nature, they increase in size with increasing relative humidity, as more and more water condenses on them. Up to a radius of about 0.5 micron (5×10^{-5} cm) they show selective scattering in visible light which makes them appear bluish by reflection, and we call them *haze*. With further increase in size, this selectivity practically disappears, and we have *fog*, which is typically colorless. We must now consider in a little more detail the scattering of light by such spherical particles.

For particles of radius a in the range $0.1\lambda < a < 10\lambda$ (λ = wave length of light), which includes most kinds of haze and some fogs, the theory of Mie [38][1] has been found entirely adequate. Starting with the electromagnetic theory, Mie was able to calculate the intensity I (lumens per steradian per particle per lumen per m^2 illuminance) in a direction making an angle ϕ with that of the incident light. This is a function of $2\pi a/\lambda$ and of m, the index of refraction of the particles (1.33 for water). To show the large variation of the polar diagram of the scattered light with the radius of the particle, Fig. 1 is presented, the limiting value as $a \to 0$ being shown by a dotted curve (Rayleigh atmosphere).

By integrating $I(\phi)$ over the sphere, the total amount of light scattered may be calculated, and it is found that this is generally greater than that incident on the droplet. For large droplets the ratio K of these quantities approaches 2, and the explanation is to be found in diffraction. Since K has been calculated [30, 47] as a function of the parameter $2\pi a/\lambda$, the coefficient of attenuation by scattering for an atmosphere containing N droplets per m^3, each of radius a, is

$$b = NK\pi a^2 \quad \text{meter}^{-1}. \tag{1}$$

In such droplets, as has been shown by Zanotelli [57], absorption is negligible, so that the extinction coefficient σ is also given by equation (1).

The correctness of these ideas is now acknowledged, and they have been remarkably well verified in natural haze by Dessens [15, 16], who caught haze particles on minute spiders' webs.

For the larger droplets of fog, Bricard [8, 9] has

1. Concisely set forth by Stratton [46, p. 563].

produced an adequate geometrical theory of scattering, serving as an extrapolation of the Mie theory. Many workers, notably Houghton and Radford [26] and Bricard [6, 7], have measured the size distribution in natural fogs, obtaining unimodal curves with maxima in the region between 4 and 10 microns radius. Such fogs should be nearly nonselective. All these researches may be criticised on the basis of sampling, especially since Driving and his co-workers [18] have presented indirect evidence for the presence of large numbers of very small droplets. Such evidence is hard to obtain because, as Dessens [15] points out, the total optical effect of these small particles is not very important.

The researches of Dessens should be repeated and an attempt made to extend them to natural fogs in order to decide this point. It is also possible that the use of the electron microscope in such work might settle the controversy about the nature of the nuclei. Another line of research which might be undertaken in some sparsely inhabited region concerns the explanation of

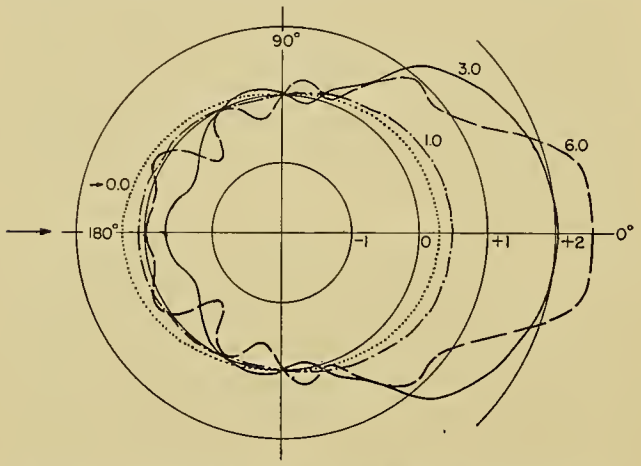


FIG. 1.—Polar diagram of scattering by water droplets, normalized at 90°. The radii are in log units. The numbers on the curves refer to values of $2\pi a/\lambda$.

the surprisingly low selectivity shown by very clear air such as occasionally permits a visual range of 150 or 200 km. Observations, chiefly in Europe, never seem to show anything like the inverse fourth power of the wave length demanded by the theory for pure gases. As a hypothesis, one may assume the effect of a comparatively small number of relatively large particles.

The Reduction of Contrast by the Atmosphere

The common observation that the more distant an object, seen in daylight, the greater its luminance, was first reduced to mathematical form in 1924 by Koschmieder [29]. The simplest case is that of a black object of intrinsic luminance zero seen against a horizon sky of luminance B_h. On the assumption of a uniform atmosphere having a scattering coefficient b_0, and illumination by the sun and a uniform (overcast or cloudless) sky, Koschmieder showed that such an object seen at a distance r will have an apparent luminance

$$B_b = B_h(1 - e^{-b_0 r}). \tag{2}$$

He further showed that an object which is not black, but has an intrinsic luminance B_0, will at a distance r have an apparent luminance

$$B = B_0 e^{-b_0 r} + B_h(1 - e^{-b_0 r}). \tag{3}$$

We cannot calculate B_0 without knowledge of the distribution of light, even if we know the properties of the object.

In these equations no mention is made of absorption, and while it is customary to write similar equations using σ instead of b, it is not immediately obvious how such an extension can be justified. Actually a more detailed analysis, using the concept of the space light B_a, which is a function of the scattering properties of the air and of its illumination, does lead to an equation

$$B = B_0 e^{-\sigma_0 r} + B_h(1 - e^{-\sigma_0 r}). \tag{4}$$

Such an analysis has been carried out by Duntley [19], who also threw off the restriction of horizontal vision and introduced a quantity $\bar{R}$ which he called the *optical slant range*, which "represents the horizontal distance in a homogeneous atmosphere for which the attenuation is the same as that actually encountered along the true path of length R" [19, p. 182]. The equation for downward oblique vision becomes

$$B = \frac{B_a(0)}{\sigma_0}(1 - e^{-\sigma_0 \bar{R}}) + B_0 e^{-\sigma_0 \bar{R}}, \tag{5}$$

in which $B_a(0)$ and σ_0 are the values of B_a and σ corresponding to the air near the ground. In the horizontal case it turns out that B_h is to be identified with $B_a(0)/\sigma_0$.

A more useful statement of the law may be made in terms of *contrast*. If B_0 and B_0' are the inherent luminances of two objects adjacent in the field of view and B_R and B_R' their apparent luminances, then defining contrast in the usual way,

$$C_0 = \frac{B_0 - B_0'}{B_0'}, \qquad C_R = \frac{B_R - B_R'}{B_R'}, \tag{6}$$

we may write two equations such as (5) and simplify to obtain

$$C_R = C_0(B_0'/B_R')e^{-\sigma_0 \bar{R}}. \tag{7}$$

This is completely general. If we confine ourselves to objects seen against the horizon sky, $B'(= B_h)$ is independent of distance, and (7) reduces to

$$C_r = C_0 e^{-\sigma_0 r}. \tag{8}$$

We have not space to expound the theory further in its applications to oblique vision, but it should perhaps be pointed out that practical situations generally require information which, at best, has to be estimated.

Equation (8) has been tested more or less thoroughly by many workers, and there is no longer any doubt about its adequacy. A matter about which there is some disagreement, however, is the exact nature of the extinction coefficient σ. In deriving equations similar to (5), Duntley [19] has made use of a theory originally developed by Schuster [41] which dealt with the distribution of diffuse radiation in stellar atmospheres. The

extinction coefficient used, which was introduced by Duntley, is that pertaining to diffused radiation. The present writer believes that this is incorrect. The only portion of the light from an object which goes to form an image of it is that which reaches the eye from the direction of the object and it would seem logical to use the extinction coefficient for directed light in such a discussion. It seems probable that the choice of the other quantity was conditioned by a desire to explain some results of Douglas and Young [17] which suggested that the value of σ determined by the photometry of a projector is slightly less than that calculated from (8). The author has recently shown [37] that the discrepancy may arise from other causes. It would nevertheless be a comfort to have a long series of simultaneous measurements of σ by the two methods, especially in fog.

The Relevant Properties of the Eye

If we are to use equations such as (7) to determine the visual range of objects, it is obvious that we need to know the least value of contrast that the eye can appreciate. Similarly, if we are interested in the visual range of light signals, we need data on the threshold illuminance E_t at the eye. This can easily be transformed into a contrast limen, so that one set of data is really sufficient for both problems.

The eye can exist in two states of adaptation, known as the dark-adapted and light-adapted conditions, the transition taking place at a field luminance of about 2×10^{-3} candles m^{-2}. The reader may be referred to Stiles, Bennett, and Green [45] for a discussion of the properties of the eye in the two states and we shall only note here that color vision is restricted to the light-adapted state.

It has been known for a long time that the threshold of contrast increases at low values of luminance and for objects of small angular subtense. In meteorology it has become a sort of convention to adopt a value $\epsilon = \pm 0.02$ for ordinary objects in the daytime; but there is no doubt that this is frequently far from the truth. During World War II a very extensive investigation was made at the Louis Comfort Tiffany Foundation, and reported by Blackwell [4]. This report covers 450,000 observations of circular objects ranging from 0.6 to 360 minutes of arc in angular diameter, and over a very wide range of background luminance (5×10^{-4} to 4×10^{2} candles m^{-2}); Fig. 2 shows one set of curves interpolated from some of the results. Note the straight portions of the curves; over this range of visual angle the product of the area and the luminance of a stimulus (*i.e.*, its candlepower) is a constant; this is the range in which the signal can be considered a "point source," and values of threshold illuminance may easily be derived from the curves.

These curves refer to the contrast required for 50 per cent probability of detection by an observer using both eyes under natural conditions. For almost certain detection the values of contrast should be multiplied by 2 or at most by 3.

A recent field investigation by Blackwell [5] makes it highly probable that these laboratory values will also apply under field conditions, provided that the factors of attention and search do not enter. Other experiments [2, 11] indicate that vision through binoculars follows the same laws, provided that allowance is made for the reduction of contrast by the optical system.

The factors of attention and search remain to be investigated, as does the enormously complicated phenomenon of *recognition*, which also brings in the question of visual acuity, and is a matter worthy of the interest of any number of extremely able psychologists.

There have been many investigations of the threshold illuminance of lights and its dependence on field luminance (see [45]). While satisfactory absolute values can be derived from the Tiffany data, "practical" values have been sought, with the general result of about 0.2 lumens km^{-2} for fixed, achromatic sources on a moonless night. If the light is flashing, a somewhat greater illuminance is required for equal conspicuity, depending

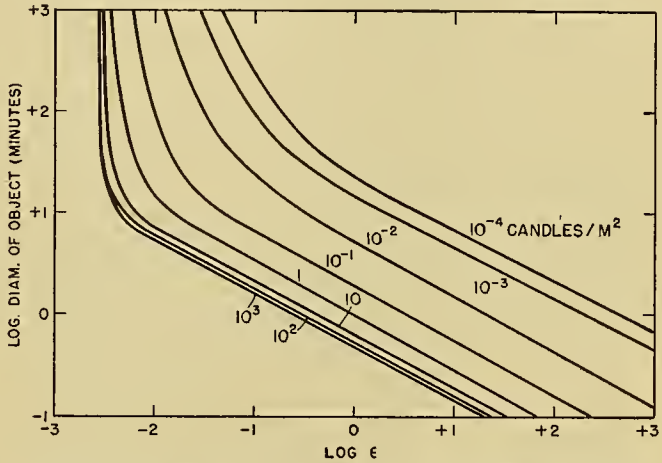


Fig. 2.—Threshold of contrast of circular objects, from the Tiffany data. Each curve refers to the background luminance marked.

on the time of the flash; this has been investigated by various authors (see [45]). Threshold illuminance varies with the color of the point source, and there are some papers [24, 25, 34] dealing with the recognition of colored lights. The general result of these investigations is to show that only red and blue-green are "safe" colors near their threshold illuminance.

The Calculation of the Visual Range

We are now in a position to combine the results of the last three sections and calculate the visual range of objects or of lights. For either computation we must know the extinction coefficient σ, and in the case of an inclined path of sight we must also know or estimate its variation in the vertical. We must know the contrast threshold appropriate to the angular size (and probably the shape) of the object and to the field luminance; or the threshold illuminance for a light of the characteristics concerned, against the existing background. There is really no fundamental difference between the

two cases, but it is simpler not to think of contrast when dealing with a point source of light.

Dealing first with the visual range of an object seen against the horizon sky, the problem is simply to put the proper values of C_0 and ϵ in equation (8) and solve for r. For a given object, which looks smaller as it becomes more distant, ϵ is unfortunately an empirical function of r.

The standard procedure, used by all writers until very recently, is to restrict the problem to "large" objects in full daylight; that is, to the approximately vertical portions of the curves at the extreme left of Fig. 2, so that ϵ may be considered constant. Meteorological writers have adopted a quantity variously called the *standard visibility* [53], *Luftlichtweite* [32] and, more recently, the *meteorological range* [20], calculated on the assumption that $\epsilon = \pm 0.02$. Let us call this V_m, and note that if it is referred to a black object ($C_0 = -1$) against the horizon sky, it is just a convenient substitute for σ_0; because if we write (8)

$$-0.02 = -e^{-\sigma_0 V_m} \tag{9}$$

and take natural logarithms, we obtain

$$V_m = 3.912/\sigma_0 . \tag{10}$$

The "meteorological range" is defined as "that distance for which the contrast transmittance of the atmosphere is two per cent" [20, p. 238].

If now we break away from the restrictions of a "large object" and full daylight, we immediately run into the difficulty that equation (8) can be solved only by a process of successive approximations. The awkwardness of this led the workers at the Tiffany Foundation to devise a remarkable series of nomograms [20], prepared for various levels of field luminance from overcast starlight to full daylight, from which the visual range of an object of any area may be read directly if its inherent contrast and the meteorological range are known. These nomograms were based on the actual Tiffany data referred to previously, and therefore correspond to a 50 per cent probability of detection. It is hoped that further nomograms will be forthcoming, based on a probability of detection much nearer unity, though the existing ones may be used for many purposes by dividing the inherent contrast of the object by 2 before entering the nomogram [20, p. 249].

The estimation of the inherent contrast of the object remains a stumbling block, except for the ideal black object. If we had a grey object of luminance factor β standing vertically under a sky of *uniform* luminance B_h, its inherent luminance would be $B_0 = B_h\beta/2$ and its contrast $C_0 = \beta/2 - 1$. A white object, for example, would have a contrast of $-\frac{1}{2}$, and this quantity has been used in a great deal of theory under the mistaken assumption that a densely overcast sky is uniform. Unfortunately its luminance is about three times as great at the zenith as at the horizon [39] and it has been shown [36] that, for a vertical white object, this results in values of C_0 between 0 and $+1$, depending on the reflectance of the ground. The complications naturally increase when the sun is shining.

Turning now to oblique paths of sight, we have the remarkably ingenious theory of Duntley [19] and the nomograms based on it [20]. The reader is asked to consult the two papers concerned, especially the first, which is far too long to summarize here, and to make up his own mind as to the practical utility of the theory. It may be that some "operational" research is indicated.

The visual range of lights at night presents a simpler problem. The illuminance from a source of I candlepower at a distance r in an atmosphere of extinction coefficient σ_0 is

$$E_r = Ir^{-2}e^{-\sigma_0 r}, \tag{11}$$

and the visual range is then given by introducing the threshold illuminance E_t:

$$E_t = IV^{-2}e^{-\sigma_0 V}. \tag{12}$$

A suitable nomogram may easily be constructed for the solution of (12). If oblique vision is involved, the problem of measuring or estimating σ remains.

Instruments for the Measurement of the Visual Range

Before we can calculate the visual range of any particular object or light signal, we must have a knowledge of σ or of V_m. A large number of different instruments have been designed for the measurement of these quantities, some requiring photometric settings on the part of the observer, others completely objective or even automatic, but none of them seem to be in use at any considerable number of meteorological stations.

Nearly all these instruments fall into four classes: (1) devices which measure the transmittance of a more-or-less extended sample of the atmosphere; (2) instruments which measure the reduction of known contrasts; (3) instruments which measure the light scattered from a small sample of air at one or many angles; and (4) empirical meters of various types which do not measure σ directly.

Transmittance meters form a fairly numerous class, which may be divided into two subclasses depending on whether they are usable only at night or at all times. Those for use only at night are generally visual telephotometers and may make use of a distant light, as for instance that of Collier and Taylor [13], or of a beam of light projected from the instrument and returned by a distant mirror, as that of Foitzik [22, 23]. For a very excellent discussion of visual telephotometers and their limitations the reader may be referred to a paper by Collier [12].

A similar duality of optical systems is found in photoelectric transmittance meters, which are often usable throughout the day and night. Those with a distant mirror, represented by that of Bergmann [3], can easily make use of a modulated light beam and a tuned amplifier to make them insensitive to daylight, and a null method of measurement to reduce the effect of fluctuations in the supply voltage. They have the disadvantage of complexity and very high cost. Simple photometry of a distant projector [17] is cheaper.

All telephotometers suffer from a serious contraction of their scales as the visual range becomes greater. It has also been shown by Middleton [37] that the beam must be made extremely narrow if very large errors are to be avoided in some kinds of weather.

Telephotometers have been devised which measure the ratio of the luminances of a distant object and the horizon sky just above it, as for example that of Löhle [31]. Difficulties of eliminating stray light, at least in any simple routine of observation, limit the precision of such instruments. Somewhat more practical, particularly with modern photoelectric circuits, is the measurement of the apparent luminance of a nearby deep black box, specially constructed for the purpose.

Scattering meters are not numerous. The "polar nephelometer" of Waldram [48, 49] which measures the light scattered by a sample of air in almost any direction is an excellent example. The "Loofah" [1] measures the scattering at 150°, which has been found to approximate a mean value under many conditions, especially at sea. All such meters fail under urban conditions because they take no account of absorption.

Finally there are any number of empirical instruments (Jones [28], Wigand [51] and others) which operate by the addition of "veiling glare" to the distant scene. Of these, the type most sound in theory was devised independently by Shallenberger and Little [42] in the United States, and by Waldram [50] in England. It is called the "disappearance range gauge" by Waldram, and presents the observer with two images of the horizon one above the other, of which the fainter may be made to disappear by turning a control. It is at least a valuable aid to visual estimation.

In spite of the very large number of instruments which have been devised, the problem remains open for solution; and perhaps it is largely an economic problem. The writer hesitates to predict the advent of a useful instrument at such a cost that meteorological services will give it wide distribution.

The Visual Range in Practice

In the absence of instruments, meteorological observers estimate "the visibility" by observing whatever objects (or lights at night) may be available around their stations. It is officially recommended [27] that dark-colored marks should be used in the daytime, and that they should be visible against the horizon sky whenever possible; also that they should be of reasonable angular dimensions. The Conference of Directors at Washington in 1947 also recommended a table [27, p. 118] which relates the visibility of lights to the daytime visibility code. This table was prepared on the basis of equations (10) and (12), using three different and highly arbitrary values of E_t, "pending the results of further investigation." We shall return to this matter in a moment, but first we wish to refer to resolution 169 of the same conference [27, p. 220], which sets forth the "Table for VV (Visibility)." This table proceeds in steps of 20 meters from 20 to 200 m, which is probably useful; but then it continues in steps of 200 m up to 16 km! It is unfortunate that the framers of the code omitted to provide instructions which would tell the unhappy observer how he is to distinguish between a visibility of 15.6 km and one of 15.8 km. Among the workers in this field it would be difficult to find one optimist who would expect even the most elaborate instrument to perform such a task.

In the absence of absurd requirements such as the above, the estimation of "visibility" in the daytime is a fairly satisfactory procedure if a reasonable series of suitable marks is available—not necessarily one for every distance in the code. It is otherwise at night. The central difficulty of night observations lies in the unknown and variable adaptation level of the observer's eye, accentuated by the fact that the observer has recently come from a brightly lighted room. In the absence of an instrument the observation is often the merest guess, and a suitable objective meter is greatly to be desired for use at night.

We shall close with a brief remark on the treatment of visibility observations as climatological data. The usual procedure is to count the number of occurrences of each code number, but it has been pointed out by Poulter [40] and by Wright [53] that this gives an undue prominence to the greater visual ranges. The expedients adopted by these two workers to improve the situation are very different. Poulter multiplied each frequency by the reciprocal of the range of distances embraced by the corresponding code number; Wright calculated what amounts to the mean extinction coefficient. Neither of these procedures involves much extra work and they should be given consideration by the climatologists.

Possibilities for Progress

It will be evident from what has been set down above that any further progress is likely to be made in the direction of experiment rather than theory. Our most hampering uncertainty concerns the value, or range of values, of the threshold of contrast actually entering into meteorological observations of "visibility," and some serious effort should be made to clear this matter up. The questions of attention and search are also in this domain of psychophysics, and could be the subjects of an immense amount of work. However, they are of military rather than meteorological interest.

On the physical side, the elegant researches of Dessens should be extended, and the question of the origin of the nuclei needs further expert attention.

There is also a need for new instruments: a simple, inexpensive one for use at many stations mainly, or even exclusively, at night; and a more elaborate telephotometer or other such "visibility meter" for important stations like aircraft carriers and large airports. In order to be of any use at all, an instrument of the latter sort would have to be very accurate, sensitive, and stable.

Finally, the applause of all meteorologists and aviators is certain to greet anyone who devises a really practical method of measuring the transparency of the atmosphere as a function of height for at least a few

thousand feet above the ground. This is certainly our leading practical problem in this subject.

REFERENCES

1. Admiralty Research Laboratory, *A Hazemeter for Direct Measurement of Atmospheric Scattering Coefficient.* A.R.L. Report A.R.L./R 1/K.904, London, March 1949.
2. Bartley, S. H., and Chute, Eloise, *Final Report on the Effects of Binocular Magnification on the Visibility of Targets at Low Levels of Illumination.* O.S.R.D. Rep. No. 4433, Nov. 30, 1944.
3. Bergmann, L., "Ein objektiver Sichtmesser." *Phys. Z.*, 35:177–179 (1934).
4. Blackwell, H. R., "Contrast Thresholds of the Human Eye." *J. opt. Soc. Amer.*, 36:624–643 (1946).
5. —— *Report of Progress of the Roscommon Visibility Tests, June 1947–Dec. 1948.* Paper read to the Aviation Lighting Comm. of the I.E.S., Washington, Apr. 21, 1949.
6. Bricard, J , "Étude de la constitution des nuages au sommet du Puy-de-Dôme." *La Météor.*, ser. 3, pp. 83–92 (1939).
7. —— "Contribution à l'étude des brouillards naturels." *Ann. Phys., Paris*, ser. 11, 14:148–236 (1940).
8. —— "Lumière diffusée en avant par une goutte d'eau sphérique " *J. Phys. Radium*, ser. 8, 4:57–66 (1943).
9. —— "Reflexion, refraction et diffraction de la lumière par une goutte d'eau sphérique." *Ann. Géophys.*, 2:231–248 (1946).
10. Cabannes, J., *La diffusion moléculaire de la lumière.* (Avec la collaboration de Yves Rocard.) Paris, Les Presses Universitaires de France, 1929.
11. Coleman, H. S., and Verplanck, W. S., "A Comparison of Computed and Experimental Detection Ranges of Objects Viewed with Telescopic Systems from Aboard Ship." *J. opt. Soc. Amer.*, 38:250–253 (1948).
12. Collier, L. J., "Visual Telephotometry." *Trans. Illum. Engng. Soc. London*, 3:141–154 (1938).
13. —— and Taylor, W. G. A., "A Telephotometer Employing the Maxwellian View Principle and Its Use in Measuring Atmospheric Transmission." *J. sci. Instrum.*, 15:5–17 (1938).
14. Dessens, H., "Les noyaux de condensation de l'atmosphère." *La Météor.*, ser. 4, No. 8, pp. 321–327 (1947).
15. —— "Brume et noyaux de condensation." *Ann. Géophys.*, 3:68–86 (1947).
16. —— "Un compteur de particules atmosphériques." *La Météor.*, ser. 4, No. 8, pp. 328–340 (1947).
17. Douglas, C. A., and Young, L. L., *Development of a Transmissometer for Determining Visual Range.* U. S. Dept. of Commerce, C.A.A. Tech. Div. Rep. No. 47, Feb. 1945.
18. Driving, A. J., Mironov, A. V., Morozov, V. M., and Khvostikov, I. A., "The Study of Optical and Physical Properties of Natural Fogs." *Izvestiia Akad. Nauk (SSSR)*, Ser. Geogr. i Geofiz., No. 2, pp. 70–82 (1943) (in Russian).
19. Duntley, S. Q., "The Reduction of Apparent Contrast by the Atmosphere." *J. opt. Soc. Amer.*, 38:179–191 (1948).
20. —— "The Visibility of Distant Objects." *J. opt. Soc. Amer.*, 38:237–249 (1948).
21. Findeisen, W., "Entstehen die Kondensationskerne an der Meeresoberfläche?" *Meteor. Z.*, 54:377–379 (1937).
22. Foitzik, L., "Ein neuer Sichtmesser." *Meteor. Z.*, 50:473–474 (1933).
23. —— "Über ein Gerät und eine Methode zur Messung der Tages- und Nachtsicht (Sichtmesser)." *Z. Meteor.*, 1:330–337 (1947).
24. Hill, N. E. G., "The Recognition of Coloured Light Signals Which Are Near the Limit of Visibility." *Proc. phys. Soc. Lond.*, 59:560–574 (1947).
25. Holmes, J. G., "The Recognition of Coloured Light Signals." *Trans. Illum. Engng. Soc. London*, 6:71–97 (1941).
26. Houghton, H. G., and Radford, W. H., "On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 6, No. 4 (1938).
27. International Meteorological Organization, Conference of Directors, Washington, 1947, *List of Resolutions.* Lausanne, O.M.I., 1948.
28. Jones, L. A., "A Method and Instrument for the Measurement of the Visibility of Objects." *Phil. Mag.*, Ser. 6, 39:96–134 (1920).
29. Koschmieder, H., "Theorie der horizontalen Sichtweite." *Beitr. Phys. frei. Atmos.*, 12:33–55 and 171–181 (1924).
30. La Mer, V. K., and Sinclair, D., *Verification of Mie Theory.* O.S.R.D. No. 1857. Dept. of Commerce Office of Publication Board, Rep. No. 944.
31. Löhle, F., "Über ein Sichtphotometer zur Messung der optischen Trübung der bodennahen Luftschicht." *Z. tech. Phys.*, 16:73–76 (1935).
32. —— "Sichtschätzung und Luftlichtmessung." *Z. angew. Meteor.*, 53:71–82 (1936).
33. —— *Sichtbeobachtungen vom meteorologischen Standpunkt.* Berlin, J. Springer, 1941.
34. McNicholas, H. J., "Selection of Colors for Signal Lights." *J. Res. nat. Bur. Stand.*, 17:955–980 (1936).
35. Middleton, W. E. K., *Visibility in Meteorology*, 2nd ed. Toronto, Univ. of Toronto Press, 1941.
36. —— "Note on the Visual Range of White and Grey Objects." *Quart. J. R. meteor. Soc.*, 73:456–459 (1947).
37. —— "The Effect of the Angular Aperture of a Telephotometer on the Telephotometry of Collimated and Noncollimated Beams." *J. opt. Soc. Amer.*, 39:576–581 (1949).
38. Mie, G., "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen." *Ann. Phys., Lpz.*, 25:377–445 (1908).
39. Moon, P., and Spencer, D. E., "Illumination from a Nonuniform Sky." *Illum. Engr. N.Y.*, 37:707–726 (1942).
40. Poulter, R. M., "The Presentation of Visibility Observations." *Quart. J. R. meteor. Soc.*, 63:31–45 (1937).
41. Schuster, A., "Radiation Through a Foggy Atmosphere." *Astrophys. J.*, 21:1–22 (1905).
42. Shallenberger, G. D., and Little, E. M., "Visibility Through Haze and Smoke and a Visibility Meter." *J. opt. Soc. Amer.*, 30:168–176 (1940).
43. Simpson, G. C., "Sea-Salt and Condensation Nuclei." *Quart. J. R. meteor. Soc.*, 65:553–554 (1939).
44. —— "Sea-Salt and Condensation Nuclei." *Quart. J. R. meteor. Soc.*, 67:163–169 (1941).
45. Stiles, W. S., Bennett, M. G., and Green, H. N., *Visibility of Light Signals with Special Reference to Aviation Lights.* Aer. Res. Comm. Reports and Mem., No. 1793, London, 1937.
46. Stratton, J. A., *Electromagnetic Theory.* New York, McGraw, 1941. (See p. 563)
47. —— and Houghton, H. G., "A Theoretical Investigation of the Transmission of Light Through Fog." *Phys. Rev.*, 38:159–165 (1931).
48. Waldram, J. M., "Measurement of the Photometric Properties of the Upper Atmosphere." *Trans. Illum. Engng. Soc. London*, 10:147–188 (1945).
49. —— "Measurement of the Photometric Properties of the

Upper Atmosphere." *Quart. J. R. meteor. Soc.*, 71:319–336 (1945).

50. —— *Disappearance Range Gauge.* General Electric Co. Res. Lab. Rep. No. 8672, Wembley, Eng., May 23, 1945.
51. Wigand, A., "Ein Methode zur Messung der Sicht." *Phys. Z.*, 20:151–160 (1919).
52. Wright, H. L., "The Size of Atmospheric Nuclei; Some Deductions From Measurements of the Number of Charged and Uncharged Nuclei at Kew Observatory." *Proc. phys. Soc. Lond.*, 48:675–689 (1936).
53. —— "Atmospheric Opacity: a Study of Visibility Observations in the British Isles." *Quart. J. R. meteor. Soc.*, 65:411–439 (1939).
54. —— "Sea-Salt Nuclei." *Quart. J. R. meteor. Soc.*, 66:3–11 (1940).
55. —— "The Origin of Sea-Salt Nuclei." *Quart. J. R. meteor. Soc.*, 66:11–12 (1940).
56. —— "Atmospheric Opacity at Valentia." *Quart. J. R. meteor. Soc.*, 66:66–77 (1940).
57. Zanotelli, G., "Assorbimento elementare della luce nel passagio attraverso alle nubi." *Atti Accad. ital., Rend Cl. Sci. fis. mat. nat.*, 2:42–50 (1940).

ATMOSPHERIC ELECTRICITY

Universal Aspects of Atmospheric Electricity *by O. H. Gish* 101

Ions in the Atmosphere *by G. R. Wait and W. D. Parkinson* 120

Precipitation Electricity *by Ross Gunn* 128

The Lightning Discharge *by J. H. Hagenguth* 136

Instruments and Methods for the Measurement of Atmospheric Electricity *by H. Israël* 144

Radioactivity of the Atmosphere *by H. Israël* 155

UNIVERSAL ASPECTS OF ATMOSPHERIC ELECTRICITY

By O. H. GISH[1]

Lincoln, Nebraska

One becomes aware of weather through sensibility to heat or cold, calm or storm, brightness or gloom, but one's unaided senses do not reveal the ever-present universal aspects of atmospheric electricity. Of course one sees lightning and hears thunder, but these are heralds of an impetuous, stormy aspect of atmospheric electricity which is not universal. Lightning is seldom seen in the polar regions and even in temperate latitudes it is seen only a small fraction of the time. The study by Brooks [8] indicates that on one-fourth of the earth's surface thunder may be heard on less than one day in a hundred.

But everywhere on the earth, electric forces in the open atmosphere are readily detectible with instruments. Many measurements made in most representative regions of the earth show that (1) during fair weather the average electric field strength or potential gradient is usually more than 100 volts per meter, (2) the electric potential increases with elevation above the surface but the field strength, or the potential gradient, decreases, and (3) the direction of the electric field in fair weather is such that positive ions in the air drift towards the earth and negative ions drift away. One must infer from these observations of the electric field that the electric charge on the surface is always negative everywhere on the earth, except in the vicinity of thunderstorms or where drifting dust or some other local charge-generating process temporarily disturbs the normal aspect.

Another fact of fundamental importance is that air, in the open, is not a perfect insulator: Although the electrical conductivity of air is so small that for most practical affairs it need not be considered, this property does play an important role in determining the electric state of the atmosphere. For example, because of this property and the existence of an electric field, an electric current flows from air to earth and the average magnitude of this current, based on numerous measurements, is such that 90 per cent of the negative charge on the earth would be neutralized in thirty minutes. Despite this, the charge at the present time doubtless is about the same as it was one hundred years ago when Sir William Thomson (later Lord Kelvin) made the first reliable quantitative measurements of electric field strength.

How is the negative charge of the earth or the corresponding electric field maintained? or, in other words, How and where is negative electricity supplied to the earth at the rate required to compensate, on the average, for the loss by electric conduction? Many attempts to answer this question have been made since the problem was first recognized, about the beginning of this century, but most of the answers have been definitely confuted. The challenge presented by this problem seemed to some to call for radical measures. Several eminent physicists thought that the answer might require some modification of basic laws of physics or might depend on some physical entity or property that had not yet been discovered [1]. These nonclassical suggestions were made at a time when all other proposals seemed inadmissible, but since then some evidence has been found which lends support to the suggestion that thunderstorms supply negative electricity to the earth at a rate which is adequate to maintain the negative electric charge of the earth and the electric field of the atmosphere.

According to this view, each thunderstorm which has developed sufficiently acts as a generator of electricity and all the thunderstorms of the earth, acting as generators connected in parallel between the earth and the high atmosphere, provide the supply current which maintains the high atmosphere at a potential of several hundred kilovolts positive with respect to the earth. The requirements which this supply current must satisfy will be seen better after the review, which follows, of the chief universal aspects of atmospheric electricity, but from what has already been said it will be evident that the more general requirements are (1) the supply current generated in the thunderstorms must flow upward into the high atmosphere and then spread out and return to earth as a more or less uniformly distributed air-earth conduction current, (2) the magnitude of the supply current must equal the total current from air to earth in all fair-weather areas of the earth, and (3) the variations of the total supply current should correspond to the variations in the total air-earth current. Whether or not the net electric current which flows between the earth and thunderstorms meets these requirements has not been definitely ascertained because the electrical circumstances under thunderstorms are so complex that it has not been feasible to make the measurements which are needed.

The foregoing requirements are clearly indicated by extensive measurements of air-earth current density which have been made in representative areas of the earth, and it is from these measurements that the magnitude and characteristics of the supply current may now be inferred.

Values of the air-earth electric current density i, although very small, have been obtained satisfactorily at several places for a number of years by automatic registration. The value of i may be obtained either (1) by a "direct" method [19] in which the current is "collected" on an insulated plate set flush with the earth's

1. Retired from Department of Terrestrial Magnetism, Carnegie Institution of Washington.

surface, or (2) by an indirect method[2] which involves the measurement of three factors: (*a*) potential gradient or electric field strength E, (*b*) the electrical conductivity of the air attributable to the positive ions λ_1, and (*c*) the conductivity attributable to negative ions λ_2. The electric current density then is obtained from the relation

$$i = (\lambda_1 + \lambda_2)E. \qquad (1)$$

The sum $(\lambda_1 + \lambda_2) = \lambda$ will here be called the *total conductivity*, or simply *conductivity* when the latter entails no ambiguity. The technique for making automatic registrations of these three factors is now more satisfactory than that for registration by the direct method. This, and the advantage for analytical purposes of knowing how the several factors vary, is the reason that the indirect method has been used in most long series of registrations.

ELECTRICAL CONDUCTIVITY OF AIR

The conduction of electricity in air and other gases became a concrete conception in the last years of the nineteenth century. Coulomb found in 1785 that an electrically charged body when exposed in air loses charge at a rate given by the law which bears his name. However, his discovery received little attention until 1887 when W. Linss made measurements, two times each day for two years, of the proportional rate of dissipation, or the coefficient of dissipation a, of electricity from a charged body exposed in the open air. The coefficient a is defined by Coulomb's law, namely, $dQ/dt = -aQ$, where Q represents the quantity of electricity on the body and dQ/dt represents the rate at which charge is lost. These measurements showed that, in present-day terminology, (1) the electrical conductivity of air, which is roughly proportional to a, varies considerably from time to time, (2) it is greater in summer than in winter, and (3) during the year the conductivity on the average varies inversely as the potential gradient. This inverse relationship is frequently found. It implies that the electric conduction-current in fair weather tends to vary less than at least one of the two component factors, namely, potential gradient and conductivity. But at some places where the air conductivity is small the air-earth current density is considerably less than normal.

The conductivity of air was earlier thought to arise from the presence of impurities, such as particles of dust, smoke, fog, or water in its various forms, until J. Elster and H. Geitel, about 1895, from their numerous measurements of electrical dissipation, showed the reverse, namely, that air generally conducts electricity best when pure and relatively dry. These observations were clarified when the conception of gaseous ions was introduced near the end of the nineteenth century.

But how are ions formed in the open air? Elster and Geitel, in search for an answer to that question, discovered that the air over land generally contains radioactive matter, that most of the important constituents of the earth's crust contain measurable amounts of radioactive matter and that the former is doubtless derived from the latter. Since it had recently been found that radiations from radioactive substances form ions in the surrounding air, the conductivity of the air over land, but not of that over the oceans, seemed to be largely accounted for.

The first clue of the ionizing agent which is active at sea appeared when Elster and Geitel and C. T. R. Wilson in the first years of this century found that air from which radioactive matter had been carefully removed continued to be somewhat conductive. These observations apparently stimulated a series of investigations by other physicists, which eventually led to the discovery of what are now commonly called cosmic rays. That these ionizing rays are of extraterrestrial origin was first clearly indicated by observations made by V. Hess (1911) during ten balloon flights. These observations were verified and extended by W. Kolhörster in 1913.

Beginning in 1915 many measurements of the cosmic radiation were made on all oceans during cruises of the *Carnegie*. These showed that the intensity is about the same at sea as at sea level on land. Furthermore, measurements over the open seas showed that there the radioactive content of air is not more than two per cent of that found on the average over land. Other measurements made on these cruises showed that the amount of radium in sea water, far from land, is less than one per cent of the amount found in the soil. The results of these observations strongly corroborated those of Hess and Kolhörster and showed that this radiation is doubtless of universal distribution and constitutes the preponderant ionizing agent over the oceans.

Furthermore, balloon observations of the several factors involved indicate that everywhere in the troposphere and stratosphere, at altitudes greater than one or two kilometers, the air is ionized almost exclusively by the cosmic radiation. This radiation is, accordingly, an all-important factor in determining the character of the universal aspects of atmospheric electricity.

But this statement does not apply in the region above the stratosphere, namely, the ionosphere extending upward from an altitude of 60 km. There the electrical conductivity is much greater than would prevail if cosmic radiation were the only ionizing agent. The intense ionization of the ionosphere is attributed to ultraviolet light and corpuscular radiation from the sun. Perhaps the comparatively great electrical conductivity in the lower part of the ionosphere plays a part in promptly distributing the supply current from the thunderstorms to remote areas over the earth, but this has not yet been proven. This distribution may occur at a lower level.

Since, aside from the possibility just mentioned, the ionosphere is presumably not involved in the phenomena which are usually regarded as belonging in the category of atmospheric electricity, the interesting elec-

2. Methods of measuring the elements of atmospheric electricity are described in the article in this Compendium by H. Israël entitled "Methods and Instruments for the Measurement of Atmospheric Electricity."

trical properties of that region will not be discussed here. This subject is treated in references [1] and [5].

Without the conductivity of the troposphere and stratosphere, which depends chiefly upon the cosmic radiation, atmospheric electrical phenomena would undoubtedly be very different, and the universal distribution of the fair-weather electrical phenomena that are now observed would doubtless not occur.

At sea level the cosmic radiation forms ions in pairs, one positively the other negatively charged, at a rate, depending upon magnetic latitude, of 1.5 to 2.0 ion-pairs per cubic centimeter per second. This is practically the complete rate of ion formation over much of the ocean area and doubtless also over land in the polar regions, but in the lower atmosphere over most land areas the birth rate of ions is several fold greater than this on account of the additional ionization there by radioactive matter. The magnitude of the ionization rate near the earth over land is not readily determined and estimates vary between wide limits, but 10 ion-pairs per cubic centimeter per second may perhaps be taken as an approximate representative average value.

Despite the greater rate of production of ions, the electrical conductivity of the air over land generally does not exceed that at sea, and at some places, especially near large cities, it is much less. For example, in the outskirts of Washington, D. C., it is about one-seventh the value at sea. This apparent paradox was resolved when it was found that in air which contains certain impurities, the normal small ions are transformed into large ions which drift more slowly in the electric field and thus contribute less to the conductivity. These large ions are indeed so very sluggish that if all the small ions were transformed in this way the conductivity would be reduced to a very small fraction of the normal value.

The electrical conductivity of a gas which contains ions of various types may be expressed as

$$\lambda = e \sum k_i n_i, \tag{2}$$

where e is the electronic charge, k is the ionic mobility, and n is the concentration of ions of each type. All ions are here assumed to carry a single electronic charge. The mobility varies as the inverse of the density of the gaseous medium and depends upon several factors including the character of the ion species, the sign of the ionic charge, and the "size" of the ion. Values ranging from 0.0003 to 0.0007 $cm^2\ v^{-1}\ sec^{-1}$ for the mobility of large ions have been reported, whereas an average value for the small ions in the atmosphere at sea level is about 1.4 $cm^2\ v^{-1}\ sec^{-1}$, but the standard deviation of these values is rather large. Measurements made in the laboratory indicate that the mobility of the negative ion in air is about 1.3 times that of the positive ion.

Since in the atmosphere the large ions appear to be formed chiefly at the expense of small ions, the air conductivity is reduced when air is polluted with substances such as some products of combustion, which occur as molecular aggregates with a diameter of the order of 10^{-6} cm. These, upon capturing small ions, become large ions with such low mobility that they play an insignificant role in the conduction of electricity. The result of this is that the conductivity is reduced because the terms kn (equation (2)) which apply for small ions are decreased more than the corresponding terms for large ions are increased.

The concentration of small ions n, when equilibrium is established between the rate of production and the rate of destruction, is given approximately by the relation

$$q = \alpha n^2 + \beta n N, \tag{3}$$

where q is the rate of production of small ions (ion-pairs per cubic centimeter per second); n is the concentration of small ions, either those positively charged or those negatively charged; N is the concentration of the positively charged, the negatively charged, and the electrically neutral large-ion constituents; α is the coefficient of combination for small ions, with a value, at standard temperature and pressure, of about 1.6×10^{-6}; β is a parameter whose value is of the same order as that for α, but this value apparently depends upon some factors which are not yet identified. An average of values for β determined from the data of Cruise VII of the *Carnegie* is about 2×10^{-6} [20].

In deriving this form of the ionic equilibrium relation, which is a simplification of more general relations [15],[3] assumptions have been made which restrict its application. But in many cases where the data used for N are the values measured with an Aitken nuclei counter, and those for n are the measures of small ion concentration, this equation seems to be satisfied. The term containing N in (3) usually is dominant in the lower atmosphere over land and, especially in the vicinity of large cities, the term in n^2 is negligible, but at altitudes greater than 1 or 2 km in the free atmosphere, the latter term apparently is dominant during fair weather. These two terms are of about equal importance for the average conditions which prevail in the air near the surface over the oceans (N about 2000). Thus the equilibrium value of n is given by $n = \sqrt{q/\alpha}$ for clean air and by $n = q/(\beta N)$ for air that is polluted with many Aitken nuclei.

The magnitude of air conductivity λ not only varies from place to place at sea level but it also varies from time to time. The average value of λ measured over the oceans on Cruise VII of the *Carnegie* was 2×10^{-4} stat mho. Values over land are sometimes greater than those over the oceans, but smaller values are found at most places where measurements have been made. These values for land are so variable that an average has little significance. At Kew Observatory in the vicinity of London, λ appears to be about one-twelfth the value at sea and the variations are dependent to a large extent upon the varying pollution of the air. At the Huancayo Observatory near Huancayo, Peru, located at an altitude of 11,000 ft in a valley between ranges of the Andes mountains, the average λ is large—about three times the average for the oceans—but it is less than should be expected for a station at that alti-

3. Consult "Ions in the Atmosphere" by G. R. Wait and W. D. Parkinson, pp. 120–127 in this Compendium.

tude even if cosmic radiation were the only ionizing agent. At Huancayo, λ undergoes an 80 per cent decrease in about one hour between 6 and 8 A.M. during the dry season (Fig. 5). The low value continues during the daylight hours and is followed by a gradual increase which begins at about sunset. The abrupt decrease of λ is accompanied by a corresponding increase in the count of Aitken nuclei.

Apparently on the days when this occurs a shallow stratum of very stable air is established at night. The nuclei, initially present in the air near the surface, coalesce and settle out during the night and any radioactive matter exhaled from the ground is entrapped. Thus there are two factors which tend to gradually increase the conductivity at night. In the morning when the stable air layer breaks up and mixing sets in, nuclei presumably come down from the higher air and such radioactive matter as may have been entrapped is dispersed throughout a greater depth of air. Measurements indicate that the rate of ionization in daytime is about 80 per cent of that for nighttime at this station. Accordingly, the greater part, about 80 per cent of the abrupt decrease of λ in the morning, is attributable to a corresponding increase in the pollution of the air by substances which can form ions which drift very slowly in the earth's electric field.

These are examples of the kind of information now available which seems to justify the statement that most of the large changes of air conductivity, not only changes with time but also with position, are associated with changes in the purity of the air.

Temperature and pressure, although important factors where change of altitude is involved, effect only minor changes in air conductivity at sea level. The fact that the conductivity is greater in summer than in winter at a number of places may be attributable partly to a temperature effect. Calculations indicate, however, that for an annual temperature range of 30C, the corresponding range of conductivity for pure air would be 18 per cent of the mean, but the actual range is much greater than this—other factors apparently are involved.

Some observations indicate that the content of radioactive matter in the air over land is greater in summer than in winter and the greater conductivity in summer is probably in part a consequence of this, but the information about the annual variation of the radioactive content of air, or about the rate of ion formation q, is at present insufficient for making an appraisal of the quantitative importance of this factor. One would expect an effect of this kind only in regions where the exhalation of radioactive matter from the soil is hindered more during the winter season than during summer, owing to the prevalence of such conditions as a snow cover, frozen damp soil, or unfrozen waterlogged soil. The chief part of this annual variation of λ in the vicinity of large cities is attributable to a corresponding variation in the pollution of the air and there is some evidence that this factor plays a dominant part in effecting the annual variation of conductivity at most places on land where observations have been made.

The air conductivity near the earth is also affected by the electric field. During normal weather the concentration of negative ions near the surface decreases as the field strength increases. This occurs because negative ions drift away from the earth when the potential gradient is positive (field strength negative) and few, if any, such ions are supplied to the air from the earth. This effect of the electric field upon air conductivity is very pronounced when, as during storms, the field strength is large. Examples are given in Figs. 4 and 6. In Fig. 6, beginning at about 14^h the negative conductivity (λ_2) is negligible during most of the following hour while the positive conductivity (λ_1) is about normal. But shortly after 15^h15^m, λ_2 abruptly returns to a normal value and at the same time λ_1 almost vanishes. This condition continues for about fifteen minutes, then, at 15^h30^m, λ_2 again vanishes and λ_1 returns to a normal value. Six other such alternations occur before the storm ends at about 17^h20^m.

Most of the changes of the electric field, which cause these marked changes in conductivity, are not clearly seen on the electrogram for potential gradient in Fig. 6. That correspondence is better illustrated in Fig. 4 during the interval 0^h to 3^h when the electric field, being less intense, was clearly recorded most of the time.

Many of the changes of air conductivity are not accompanied by noticeable changes in air-earth current density. It is only when the change in conductivity extends throughout a considerable range of altitude that a marked correlation is seen. These cases must be taken into account when one is examining data for the broader universal aspects of atmospheric electricity. For example, in estimating the magnitude and the character of the variations of the supply current from measurements of i, allowance must be made for abnormalities which depend wholly upon local circumstances. The most significant of these abnormalities, for areas of fair weather, depend upon local modifications of the distribution of air conductivity with altitude.

Air conductivity depends upon altitude in a complicated way. The value at an altitude of 18 km (60,000 ft), during the notable flight of the balloon *Explorer II*, was 100 times the average for sea level. The chief factors involved here are (1) the intensity of cosmic radiation increases with altitude, (2) the rate of ion formation for a given ionizing radiation decreases directly as the air density, (3) the mobility of the ions varies inversely as the air density, (4) the rate of ion destruction in pure air of a given ion concentration decreases with altitude, varying directly with the $\frac{1}{3}$ power of pressure (approximately) and inversely as the $\frac{7}{3}$ power of absolute temperature, (5) the concentration of radioactive matter exhaled from the earth over land decreases with altitude, (6) the pollution of the atmosphere usually decreases with altitude, and (7) the dependence of β (equation (3)) upon temperature and pressure doubtless plays a part of unknown magnitude in determining the variation of λ with altitude in the lowest kilometer or so.

The last three factors apparently are relatively insignificant at altitudes greater than one or two kil-

ometers during normal weather but, in the vicinity of storms, effects which may be attibuted to these factors (particularly the last two) were observed by O. H. Gish and G. R. Wait at considerably higher altitudes.

The rate of ion formation q, when cosmic radiation is the only ionizing agent, depends on factors (1) and (2). Values for q as a function of altitude are shown in Fig. 1. These computed values are based on observations of cosmic radiation reported by Bowen, Millikan, and Neher and on average data for temperature and pressure for each of the two latitudes. The value of q increases from a low value at sea level to a maximum at an altitude of 12 to 13 km. The maximum for the higher latitude is more than two times that for the lower latitude.

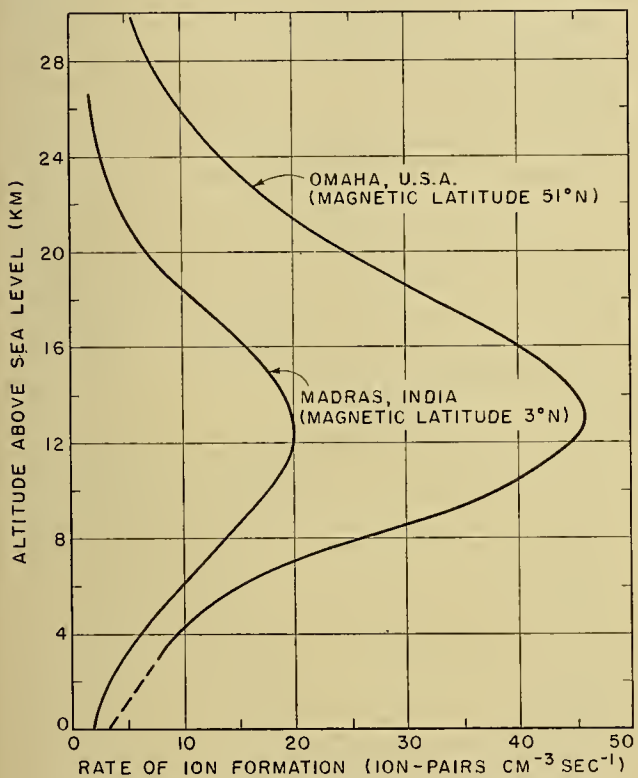


Fig. 1.—Rate of ion formation by cosmic radiation for middle and low latitudes. (*From data of Bowen, Millikan, and Neher.*)

An increase of conductivity with altitude was first shown by measurements of λ made on twelve balloon flights during the period 1905–20. Eleven of these started in Germany and one in Russia. The maximum altitude at which measurements were made was less than 6 km except for one in which it was nearly 9 km.

Continuous registration of λ up to a maximum altitude of 22 km was made during the flight of the stratosphere balloon *Explorer II* [14]. The results for this flight are shown in Fig. 2. The crosses on graph *A* represent direct measurements of λ_1, and the circles represent values derived from direct measurements of λ_2 by multiplying the latter by 0.78. This factor is the ratio of the mobility of positive ions to the mobility of negative ions. The smooth graph *B* represents values calculated from measured values of cosmic radiation and of temperature and pressure, the latter two observed during the flight. During this flight the conductivity increased from the surface up to an altitude of 18 km (60,000 ft) where it was about 100 times the value at the surface. In the altitude range 18–22 km, it varied irregularly but in general decreased with altitude. This

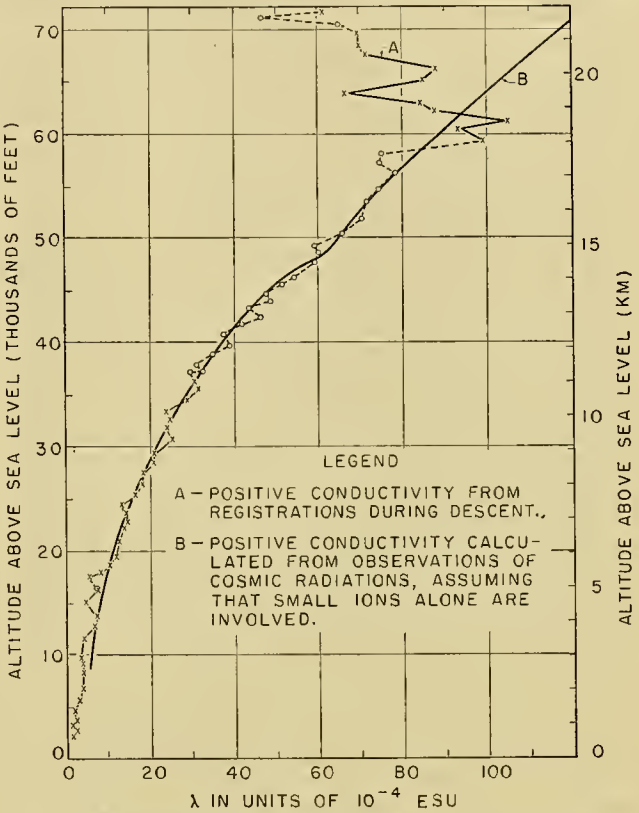


Fig. 2.—Air conductivity, flight of *Explorer II* near Rapid City, South Dakota, November 11, 1935.

last feature may be attributed to the presence there of substances (perhaps Aitken nuclei) which served for the formation of large ions. This decrease of conductivity is in the same region where ozone was simultaneously found to be especially abundant. Whether this feature is universal, whether it is generally associated with ozone, and whether there are factors in the atmosphere at yet higher levels which effect a similar diminution of the conductivity are questions which have important bearing on other than electrical aspects of the atmosphere. Therefore, further investigation of the conductivity of the high atmosphere is called for. At present, there is indirect evidence which indicates that this diminution of conductivity at high levels is either not universal or else is rather limited in vertical extent. The unexplored region to which this statement applies extends from 22- to about 60-km altitude. Above 60 km the concentration of ions has been determined at a number of places on the earth by quantitative studies of the "reflections" of radio waves. Such observations demonstrate that the air above that level is extraordinarily conductive.

Other measurements of λ up to an altitude of over 14

km (48,000 ft) were made in 1948 by Gish and Wait during ascent or descent for electric surveys over thunderstorms (a joint project of the U. S. Air Force and the Department of Terrestrial Magnetism of the Carnegie Institution of Washington). These unpublished values are on the whole consistent with those shown here except that they tend to be smaller.[4]

The values of λ versus altitude, shown in graph B (Fig. 2) were calculated for the case in which the air contains no large ions and the cosmic radiation is the only ionizing agent. The agreement between these and the measured values is one of the considerations which leads to the conclusion that throughout much of the high atmosphere pollution is negligible and cosmic radiation is the chief ion-producing agent.

The relatively large air conductivity at high altitudes and the increase with altitude are features upon which several other aspects of atmospheric electricity depend. For example, (1) the suggestion that thunderstorms are the source of the supply current could not be seriously considered if λ were not a rapidly increasing function of altitude, and (2) the decrease of electric field strength with altitude and the character of the diurnal variation and that of some of the other variations of field strength depend on these aspects of λ.

For cases in which the electric field strength or the air-earth current, or both, depend upon the over-all effect of the conductivity throughout a vertical column of the atmosphere, it is convenient to use the concept of electric resistance, rather than electric conductance, of the column.

The resistance of a vertical column of the atmosphere of 1 cm^2 cross section and extending from the earth's surface up to a certain height is obtained by integrating the reciprocal of λ from the surface up to that height. A value of "columnar resistance" r, obtained in this way from the values of λ shown in Fig. 2, is 10^{21} ohms for a column extending from sea level to an altitude of 18 km. If this is regarded as representative for the whole earth, the total effective resistance R from the surface to the 18-km level would be 200 ohms. Most of this resistance is offered by the lower part of the column. The highest eight kilometers contribute only five per cent of the total, whereas the lowest two kilometers contribute 50 per cent.

Estimates of r up to altitudes considerably greater than 18 km, based on observed values of cosmic-radiation intensity and the indicated trend of that radiation at altitudes not yet explored, indicate that the value of r for a column extending up to the ionosphere would not exceed by more than 10 per cent that for 18 km, provided the small ions at those higher levels are not transformed into larger, less mobile types, in appreciable number. Such estimates may at least set a lower limit for r between the earth and the ionosphere. An upper limit, not exceeding by as much as 50 per cent the value up to an altitude of 18 km, seems to be indicated by the extent to which air pollution of local origin—say from a city like Washington, D. C.—diminishes the air-earth conduction current through diminution of air conductivity in the lower atmospheric strata.

The columnar resistance is apparently 15 to 20 per cent greater near the equator than in middle latitudes, owing chiefly to a dependence of the vertical distribution of cosmic-radiation intensity upon latitude (Fig. 1). This is of interest here because it provides an explanation of the tendency, found by S. J. Mauchly, of the potential gradient measured at sea to be about 17 per cent less in the equatorial belt than in belts of higher latitude, whereas the air conductivity at the surface was found to be practically independent of latitude. This variation of r with latitude is not, however, as large as the variation from place to place on land. Over an urban area the value of r may be several times that in the relatively pure air over open country.

When all these circumstances are considered, it seems likely that an average value of r is about 10^{21} ohms and that the effective resistance R over all fair-weather areas of the earth is not far from 200 ohms. Here R is equivalent to r/S where S is the area of the earth. It is assumed here that the total area over which electrical storms are in progress at a given instant is negligible—the data for thunderstorms of the earth collected by C. E. P. Brooks indicate that the total storm area is usually less than $0.008S$.

The electrical conductivity of air in the ionosphere is apparently so great that electric fields there must be very small and can be disregarded in this discussion. No direct measurements of λ have been made in the ionosphere, but a reliable estimate of the order of magnitude is obtained by using the measurements, made by "radio" methods, of the equivalent ion density, together with estimates of air density at the corresponding altitudes. The value for λ, estimated in this way for an altitude of 70 km, near the lower boundary of the ionosphere, is about 10^7 stat mho cm^{-1}, or the resistivity, the reciprocal of conductivity, is less than 10^5 ohm cm. According to this estimate, the relaxation time, namely $1/(4\pi\lambda)$, at this altitude is less than 10^{-8} sec. This means that a local concentration of free electric charge cannot persist here for an appreciable time; it would, indeed, be diminished to 0.01 per cent of the initial value in 0.1 μsec.

Similar circumstances occur in the earth: The resistivity of most of the material near the earth's surface is less than that for air in the lower ionosphere—that of ocean water is less than 100 ohm cm. Even for geological structures where the highest values of earth-resistivity (10^6 to 10^7 ohm cm) have been measured, the relaxation time is of the order of microseconds. In contrast to this, the relaxation time for air near sea level is generally greater than 400 sec while for air at 18-km altitude it is about 4 sec. Because of these circumstances it seems permissible to describe the world-wide aspects of atmospheric electricity with the aid of a spherical condenser model as follows.

4. Subsequent to the preparation of this article, results of the measurements described above have been published and will be found in "Thunderstorms and the Earth's General Electrification," by O. H. Gish and G. R. Wait, *J. geophys. Res.*, 55:473–484 (1950).

The earth and the ionosphere, or possibly the upper stratosphere, serve as the inner element and the outer element, respectively, of a spherical electrical condenser. Because the air between the inner and outer elements is conductive this condenser has a "leakage" resistance R. A difference of potential V between the elements is maintained by the supply current. The leakage current I is V/R, and for steady conditions this is also the magnitude of the supply current. For $I = 1800$ amp and $R = 200$ ohms (values based on observations), $V = 360{,}000$ v. A somewhat lower value of V is obtained from measurements of potential gradient made on balloon flights at altitudes ranging from sea level to about 10 km (see equation (4)).

Apparently I and V are the chief variables in this relation and of these V is regarded as the independent variable. No appreciable variation in R is indicated by the data now available. Although the average conductivity measured on Cruise VII of the *Carnegie* (mean epoch, 1929) was only 74 per cent of that for Cruise VI (mean epoch, 1921), the average air-earth current density did not differ significantly, that for Cruise VII being 107 per cent of the value for Cruise VI. This indicates that I, and consequently V/R, was essentially the same for these two epochs. This fact does not necessarily exclude the possibility that V and R were related as dependent and independent variables, respectively, because if R had increased but the supply current had remained constant, V would have varied directly as R. But there are no grounds for thinking that R varies appreciably owing to variation throughout the atmosphere in the rate of ionization, the chief factor. The chief ionizer, the cosmic radiation, seldom varies by more than a few per cent of the mean; variations in amplitude as great as 3 to 4 per cent occur infrequently, usually during magnetic storms, and last only from a few hours to a day or two. On only four occasions in more than a decade have increases of more than 10 per cent of normal (at sea level) been observed [12]. Perhaps on these occasions a detectible decrease in R occurred. This should be revealed by a simultaneous increase of λ at widely distributed stations and a corresponding decrease of E. No such world-wide changes of λ and E have yet been definitely detected, but a special examination should be made of the four cases just mentioned.

Apparently the only possible source of world-wide changes in λ, amounting to more than a few per cent and continuing for long periods, is corresponding changes in the pollution of air with nuclei which serve for the formation of large ions. Even if such extensive changes do occur near the earth's surface, there would be no comparable change in R unless the change in λ occurs throughout most of the troposphere. A decrease of λ from the surface to a considerable altitude apparently does occur over limited areas especially in the vicinity of cities or industrial centers where there is notable pollution of the atmosphere. The extent of this is such that the columnar resistance r appears to be increased severalfold but the total area involved is doubtless such a small part of the earth's surface that only immeasurable effects in R may be expected. Periodic changes in r, such as diurnal variations, are also in evidence at some places, but these depend mainly upon local circumstances and apparently do not affect R appreciably.

These statements about variations in r are based on information obtained by an indirect method which depends upon the assumption that $ir = V$ is the same everywhere on the earth at a given instant. If i is measured simultaneously at two places, then according to this condition the ratio of the value of r at one place to that at the other place is equal to the inverse ratio of corresponding values of i. The efficacy of this method of analyzing such data also depends upon the circumstance that at some places, notably over the oceans, r seems to be much more nearly constant than at some places on land. This device has been used chiefly for estimating the average diurnal variation of r [21, 22]. The results are at least plausible.

That the more prominent variations of λ occur chiefly in the lower part of the atmosphere is indicated by the reciprocal relation, frequently found, between λ and E (electric field strength), namely, $\lambda E = i$ where i is approximately constant. This indicates that in these cases r is not modified much by the changes in λ, and that accordingly the vertical extent of the changes in λ and E is relatively small. The height of the air stratum involved has been estimated in several ways. At Paris a height of 200 m or less was indicated by the observation that variations of E near the top of the Eiffel Tower were chiefly of the universal type, whereas at a nearby ground station the variations were much more complex. Heights for the region of abnormal conductivity of the order of one to two kilometers have been estimated for other situations, in which r is appreciably modified.

Of the features of air conductivity discussed here, those of chief importance for the broader aspects of atmospheric electricity are (1) electrical conductivity of air is a universal property, (2) this property increases with altitude and at some altitude, probably less than 60 km above sea level, is so great that at a given instant the electric potential at that level is essentially the same everywhere over the earth, and (3) the electrical conductance between the earth and the high atmosphere, or the reciprocal of this, the resistance R, probably is not subject to appreciable variation.

THE ELECTRIC FIELD OF THE ATMOSPHERE IN FAIR WEATHER

Many observations of the electric field in the atmosphere have been made during the nearly two hundred years since Franklin made his famous kite experiment in 1752 and Lemonier, later in that year, first observed an electric field in the atmosphere during clear weather. Before the end of the eighteenth century a number of the characteristics of the electric field were correctly inferred from qualitative observations made chiefly in Europe. Quantitative measurements of the electric field strength came into vogue after Sir William Thomson in 1860 stressed the need of measurements which could

be compared even though made by different observers and at different places.

During this latter period, measurements have been made on most of the representative areas of the earth, including a number of series for the polar regions, many widely distributed measurements on the oceans, and measurements made during balloon flights (all in Europe). Continuous registration has been used at most land stations since about 1910 and was also successfully employed at sea during Cruise VII of the *Carnegie*. The features of the electric field which are clearly shown by these data are as follows for fair-weather conditions.

The electric field strength E is negative, or the potential gradient is positive, wherever and whenever fair weather prevails. The exceptions to this are rare and transitory. For example, a positive field (negative gradient) is registered while a dust whirl passes near the station. An example of such an effect appears on Fig. 5 between 14^h40^m and 14^h50^m. Since at the surface of a conductor $E = -\partial V/\partial Z = 4\pi\sigma$, where $\partial V/\partial Z$ is the potential gradient along the outwardly directed normal and σ is the surface charge density, one concludes that the charge of the earth in fair-weather areas is negative and that the potential of the air increases with altitude.

The potential gradient (or $-E$) at the surface varies considerably from place to place. The average potential gradient at sea during Cruise VII of the *Carnegie* was 130 v m^{-1}. At some places on land an average considerably less than 100 v m^{-1} has been reported, but values larger than that for the oceans are prevalent in densely populated areas: at the Kew Observatory near London the average value exceeds 300 v m^{-1}. In the polar regions and in open country, far from sources of atmospheric pollution, the average is about the same as that for the oceans. Apparently a value of about 130 v m^{-1} is a fairly representative average for the preponderant part of the earth's surface.

The potential gradient decreases with altitude. The measurements of gradient made on 57 balloon flights, mostly over central Europe, as summarized by Schweidler [17], satisfy the following empirical equation:

$$\partial V/\partial Z = 90 \exp(-3.5z) + 40 \exp(-0.23z), \quad (4)$$

where $\partial V/\partial Z$ is expressed in volts per meter for z, the altitude in kilometers. There is a sharp decrease from the surface up to about 1-km altitude followed by a slower decrease. The value at $z = 0.5$ km is less than half the value at the surface, and at $z = 9$ km it is less than 4 per cent of the surface value. The data for altitudes less than 1 km are scattered much more about the general trend than are those for higher altitudes. This may be due to variable pollution in the air near the earth, a circumstance that is indicated by other types of atmospheric electric observations such as the contrast between measurements of E made on the Eiffel Tower and of those made at a neighboring ground station.

A decrease of E with altitude, if widespread and of the magnitude and character indicated by these data, doubtless is attributable chiefly to a corresponding increase of air conductivity. For these conditions $\lambda E = i$ is independent of altitude if the electric space-charge density ρ does not vary with time and provided convection plays a negligible part in the transport of electricity in the atmosphere. The latter conditions doubtless are satisfied during most fair weather. The drifting of snow, or of dust, however, is accompanied by the generation and convection of electric charge and although this is a source of conspicuous modifications of the electric field near the earth on some occasions, for average fair-weather conditions the magnitude of electric convection-current density calculated on the basis of available data is negligibly small.

Measurements of both λ and E were made on only a few balloon flights. For these the product λE was not invariable with altitude, but this result is probably attributable to a variation of the air-earth conduction current with either, or both, time and position [10].

The potential gradient of fair weather also decreases with altitude in the lowermost few meters owing to a depletion of the negative ions in this region, the so-called *electrode effect*. Under the action of the electric field, negative ions drift away from the earth and since such ions are not emitted from the earth at an appreciable rate, the flux of negative ions, and accordingly the concentration, must vanish at the surface. But for steady conditions, the flux of positive ions from air to earth is nearly continuous in the lowermost few meters, and at the air-earth boundary is equal to the total electric flux at higher levels. Two attending conditions are (1) positive ions predominate or a positive space charge prevails in air that is contiguous with the earth, and (2) the potential gradient decreases with altitude. These general aspects of the electrode effect may be modified or obscured when either or both the electric convection-current component and the displacement-current component are appreciable, or when the vertical distribution of air conductivity near the surface is abnormal. The conclusion that the electrode effect is a universal characteristic of the electric field of the atmosphere depends chiefly on the observed fact that λ_1/λ_2 is usually greater than unity in fair weather and that it increases as the gradient increases.

The temporal variations of potential gradient are usually of complex origin, however there are types which may be classified on the basis of origin, as follows. From the several equivalent expressions for air-earth current density, namely $+\lambda E$ and $-V/r$, one obtains the relation

$$E = -V/(\lambda r), \quad (5)$$

which is valid except for the rather unusual condition that i is dependent upon altitude, or that the displacement current is appreciable. On this basis the following four types of change in E may be adopted: type (a) in which E is independent of V, λ, and r, but λ depends upon E; type (b) in which $E \propto 1/\lambda$, with V and r constant; type (c) for which $E \propto 1/r$, while both λ and V are constant; type (d) in which $E \propto V$, while λ and r are constant. In this discussion E and λ are generally used to denote values observed in air practically contiguous with the earth, V denotes the potential of the ionosphere (or upper stratosphere) relative to the earth, and r denotes the columnar resistance—the elec-

trical resistance of a vertical column of air 1 cm² cross section extending from the observation point up to the ionosphere.

Variations of the first three types generally are marked by local characteristics; many occur at irregular intervals but some follow a diurnal or other periodic routine, which varies much from place to place. The only thing in common among these variations is that usually factors of local origin tend to increase the gradient during daylight hours more than at night, but the contrasts in the amplitude and character of these variations at different places are remarkable. An example of the contrast between periodic local variations of this sort may be seen by comparing an electrogram for potential gradient (Fig. 3) obtained at the Watheroo Magnetic Observatory, Western Australia, with a corresponding electrogram (Fig. 5) made at the Huancayo Magnetic Observatory near Huancayo, Peru (observatories established and operated by the Department of Terrestrial Magnetism of the Carnegie Institution of Washington). In the latter electrogram, the gradient at

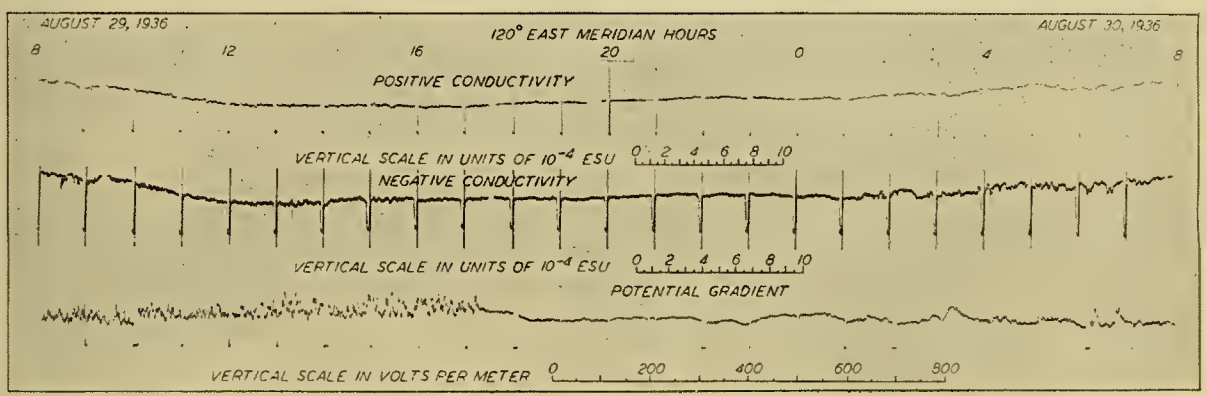


FIG. 3.—Electrograms, Watheroo Magnetic Observatory, "quiet" day.

8^h30^m (local time) is about five times the value registered two hours earlier. In the former, there is a relatively small and gradual increase of gradient, beginning at about 8^h local time and continuing throughout the daylight hours. The universal diurnal variation, which will be discussed later, is, of course, present in both these cases, but allowance for that would not change the order of contrast seen here. Analyses of such records from these two stations indicate to the author that the local aspect of the diurnal variation of potential gradient at Watheroo is largely of type (*c*) while that for Huancayo is chiefly of type (*b*).

At the Kew Observatory, where the influence of local air pollution is prominent, a semidiurnal variation of gradient of local origin appears with maxima at about 8^h and 21^h local time. These apparently are a combination of types (*b*) and (*c*) with type (*b*) dominant in the morning and type (*c*) coming into evidence later in the day [22]. The latter is an interesting example of a local influence which may be explained as follows.

The introduction into the air in the general vicinity of the observatory of substances which serve as nuclei for the formation of large ions proceeds at an enhanced rate during the period 4^h to 20^h. Near the surface the concentration of these nuclei increases rapidly between 6^h and 8^h in winter (3^h to 9^h in summer) and as a consequence λ decreases; then as convection and turbulence increase during the day, the local contamination decreases somewhat, and λ increases; but when the air becomes more stable in the evening, local contamination increases and λ decreases again until checked when, after 20^h, both the rate of introduction of the nuclei and their concentration decrease. Throughout the daylight hours the columnar resistance r increases to a maximum at about 19^h. This doubtless indicates that the content of nuclei in the air column has been increasing during this period. A decrease of r which occurs during the night is doubtless the combined effect of (1) decreased rate of supply, and (2) scattering and other modes of dissipation of the nuclei.

Potential gradient variations of type (*a*) are most prominent during stormy weather. Examples of such effects are exhibited in Figs. 4 and 6. These figures are reproductions of unretouched electrograms obtained at the Watheroo Magnetic Observatory and at the Huancayo Magnetic Observatory, respectively. Figures 3–6 inclusive are samples selected from the thousands of such registrations made during the years 1922–46 at Watheroo and 1925–46 at Huancayo. Points for zero gradient are recorded each hour. Ordinates above this line of "zeros" correspond to positive values of gradient. During the 24-hr period of record shown in Fig. 4, three periods of rainfall were indicated by the station rain gauge. These are indicated at the top of the figure. The corresponding storm effects in potential gradient are exhibited in the lowest electrogram. Most of these were of moderate intensity on that day, and varied from positive to negative values of gradient.

During the ten-minute period beginning at 18^h10^m the gradient was entirely negative and at its greatest intensity exceeded 400 v m^{-1}, the limit of registration. During the period 0^h to 0^h50^m, several changes from negative to positive gradient occurred. The trace of these may not appear in the reproduction—the changes were rapid and the extreme values exceeded the limits of registration, namely, −400 and 700 v m^{-1}. But fortunately, in cases where the trace for potential gradient is not clearly shown, the sign of the gradient can usually be ascertained by an examination of the traces for con-

ductivity, because of the electrode effect previously discussed. A characteristic of that effect is that a low value of negative conductivity and a normal value of positive conductivity should accompany a large positive potential gradient, while for a large negative gradient the roles of positive and negative conductivity should be interchanged. There is abundant evidence as well as a rational basis for this rule, but such evidence is not clearly shown in Figs. 4 and 6 during periods of greater activity because the original gradient trace was very dim when rapid changes of gradient occurred and was "off scale" for considerable periods. However, the reader doubtless can see evidence of trends of this character in these figures.

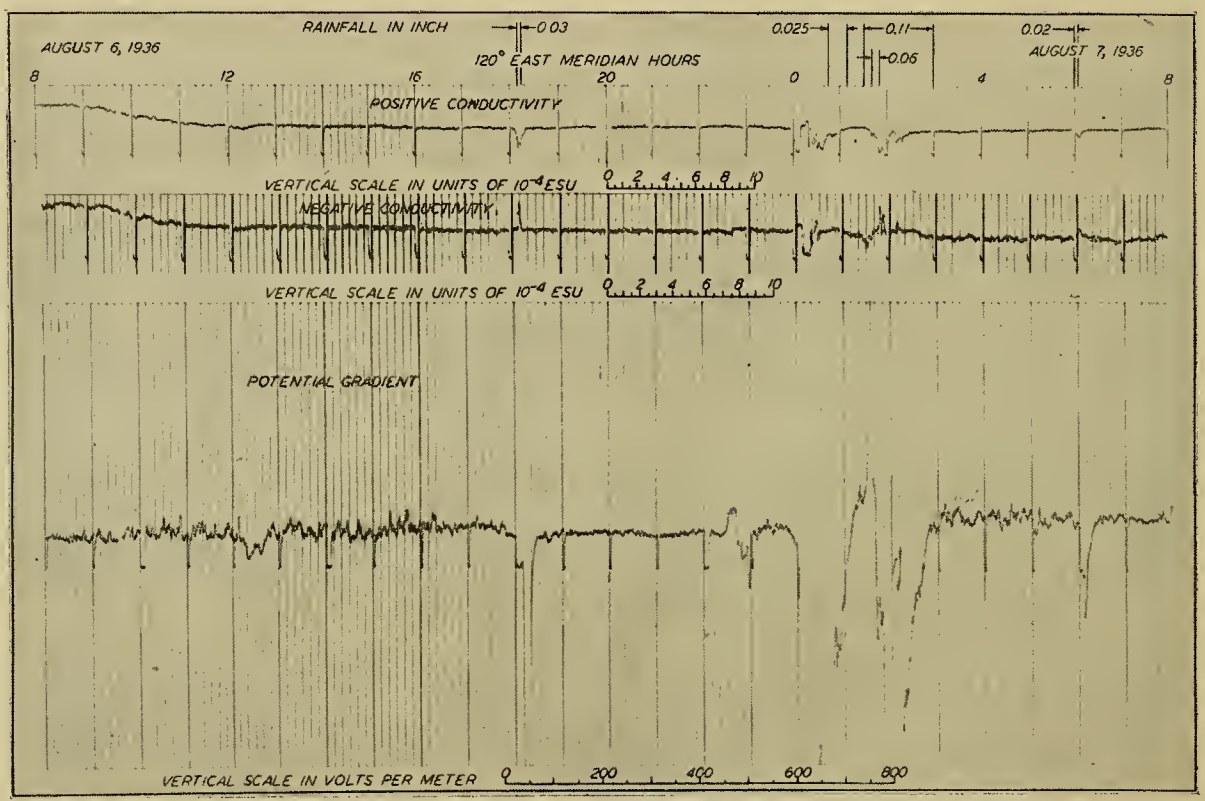


Fig. 4.—Electrograms, Watheroo Magnetic Observatory, with periods of light rain.

With the aid of this rule one can readily see that during the periods when intense electric fields prevail the sign of the gradient varies. The record shown in Fig. 6 covers two storms in a 24-hr period: one in the interval 20^h to 22^h, approximately, on March 17, the other beginning about 14^h March 18 and lasting more than three hours. During the former period, a negative gradient was recorded for more than one hour but an intense positive gradient for only 10 min is indicated. During the latter storm an intense positive gradient prevailed during the first 75 min. This was followed by an intense negative gradient which continued about 15 min. In the next hour and three-quarters, eight abrupt reversals of sign occurred.

Such characteristics of potential gradient must be attributed to considerable masses of electric charge or "charge-clouds" in the vicinity of the observatory—a charge-cloud does not necessarily coincide with a visible cloud. Ten such charge-clouds are indicated by the record for the latter storm. So many alternations of potential gradient in one series (barring changes which accompany lightning discharges and which are of too short duration to register with the apparatus used here) is somewhat unusual. They appear more frequently in records for the Huancayo station than in any other records available to the author.

The suggestion that such a series of changes in gradient may be attributed to a number of charge-clouds drifting by in tandem array may be an oversimplification. The only model of this sort which could account for the abrupt change of sign—within about one minute—and the intense field, continuing for 10–15 min or more between reversals, is one in which the charge-clouds are at an elevation which is small compared with their lateral dimension, and the array of clouds, charged alternately positive and negative, is closely packed. It is doubtful whether such a model is compatible with several physical conditions which seem to be required to maintain the charge of such clouds. Any further discussion of phenomena of which this case is representative should come under the category of the thunderstorm electric field, a topic which is discussed very briefly in this article.

Another example of a field change of type (*a*) is shown on the trace for potential gradient in Fig. 5, just before 14^h50^m. This is a characteristic "dust whirl" or "dust devil" effect. The sign of the field change indi-

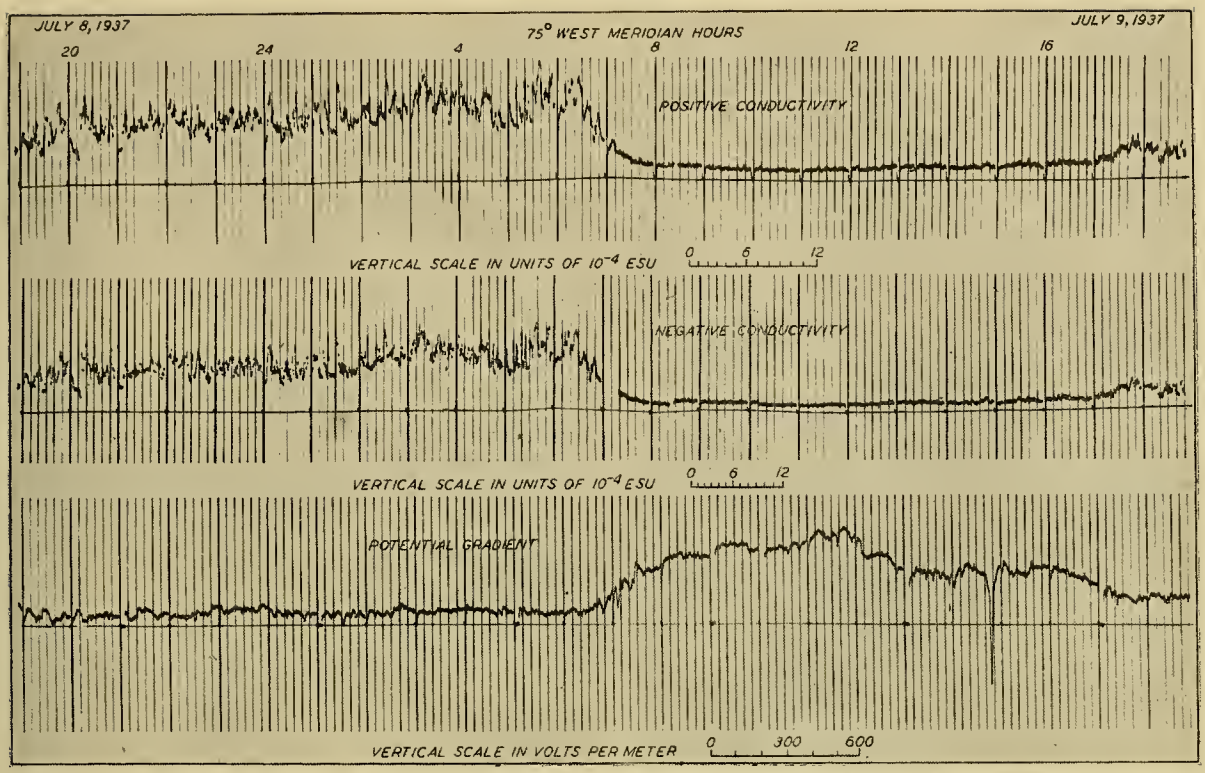


FIG. 5.—Electrograms, Huancayo Magnetic Observatory, "quiet" day during dry season.

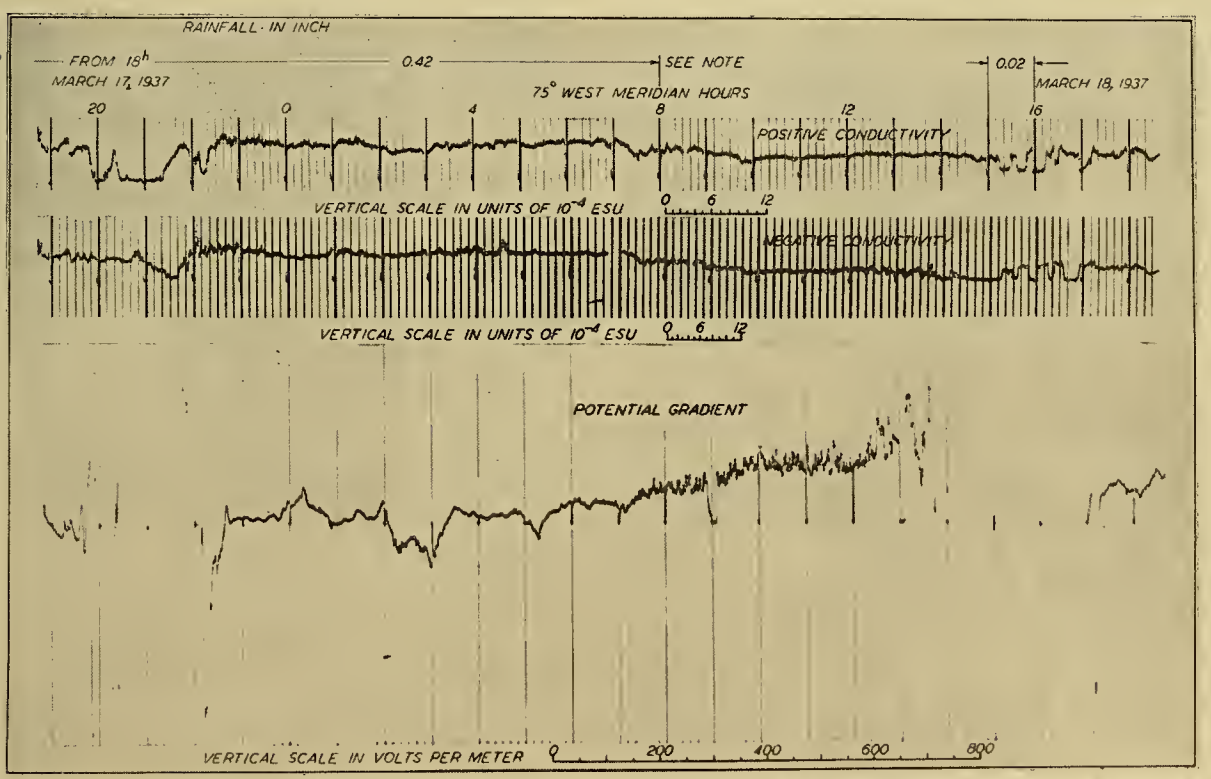


FIG. 6.—Electrograms, Huancayo Magnetic Observatory, rainy day. (Note: Clearing in morning after 10.5h; early afternoon March 18, increasing cloudiness, distant rain, and thunder to southwest at 14h.)

cates that the electric charge of the dust in this whirlwind is preponderantly negative.

The relatively small and rapid variations of electric field, which are common during daytime at the Watheroo Magnetic Observatory, are illustrated in Fig. 3. This effect, apparently of type (*a*), is clearly correlated with wind strength and wind variability. Presumably the fine sand, prevalent in this semiarid region

of Western Australia, is agitated enough by the wind to introduce puffs of electric charge into the air near the surface. There is some evidence that small variations of conductivity are associated with the field changes and the former are doubtless dependent on the latter. The drifting of snow, like the drifting of sand or dust, also modifies the electric field, but during the former, positive charge is introduced into the air.

A description of field changes of types (*a*), (*b*), and (*c*) is relevant here chiefly because these phenomena must be recognized for what they are and must be distinguished from the universal aspects of the electric field in order that the latter may be identified.

Changes of type (*a*), which are prominent during storms, have importance for the clues they provide

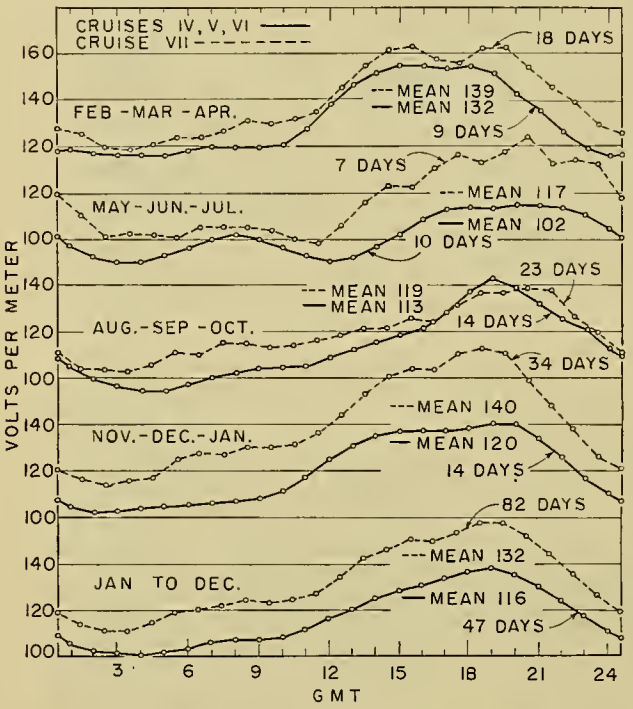


FIG. 7.—Diurnal variation of potential gradient on the oceans, *Carnegie* Cruises IV, V, and VI, 1915–21 and Cruise VII, 1928–29.

concerning the electrical constitution of storms, but changes of types (*b*) and (*c*) are of minor interest as geophysical phenomena, especially when they depend upon the activities of man.

The discovery of the *universal diurnal variation of potential gradient* was delayed because most of the atmospheric electric observations, prior to about the year 1915, were made near cities or at other places where field changes of types (*a*), (*b*), and (*c*) are prevalent. This feature, first noted by Mauchly in 1921 [6], has great importance for the subject. His analysis of the data for potential gradient over the oceans, obtained on Cruises IV, V, and VI of the magnetic survey yacht, *Carnegie*, showed that the diurnal variation of the gradient at sea proceeds according to universal time, not according to local time.

This is illustrated by the results which are exhibited in Fig. 7. Average values of potential gradient, in volts per meter, are there plotted as ordinates against the hours of the Greenwich day, counting from midnight to midnight. The combined results for Cruises IV, V, and VI are shown separately from those for Cruise VII. The latter, although generally of greater amplitude, vary in nearly the same manner as the former. If attention is fixed on either of the two lowest graphs, which represent hourly averages for the year, it is to be seen that the gradient is smallest about four hours after Greenwich midnight and greatest at about 19^h. That this feature varies somewhat throughout the year is shown by the other graphs which represent different quarters of the year. When these same data are averaged for hours of the local day, no significant diurnal variation is disclosed. Thus the change in gradient during the day does not depend upon daylight and darkness or other factors which vary in a manner directly related to the elevation of the sun at a given place. Another way of viewing this universal aspect is as follows. The negative charge of the earth as a whole in fair weather is greatest at about 19^h GMT and least at about 4^h GMT, and the total air-earth current I varies in a similar manner. The latter depends partly upon the fact that over the oceans the diurnal variation of λ is negligible.

This universal variation in potential gradient is also manifested in the diurnal variation of potential gradient at most places on land. But factors of local origin there often introduce component variations which may preponderate, especially in or near centers of population where the conductivity of the air is affected by atmospheric pollution. The concentration of these impurities is, of course, dependent not only on the rate at which they are supplied to the atmosphere but on the strength and direction of the wind, on rain, or on other processes which scatter these substances or carry them to or from the place at which the electrical features are considered. Variations of the content of radioactive matter in the air over land also result in variations of the concentration of the small ions, upon which the conductivity of the air chiefly depends. Over land the air conductivity and the gradient may, therefore, be expected to vary during the day in a manner and to an extent which is often dependent upon a complex and variable set of local circumstances.

The average potential gradient of fair weather, the universal diurnal variation of gradient, and doubtless some other temporal variations are of type (*d*), that is, they depend directly upon V, which, in turn, is presumably dependent upon the supply current.

The universal aspects of the electric field of the atmosphere which are of chief importance may be summarized briefly as follows: (1) the electric potential gradient in the atmosphere during fair weather is positive everywhere on the earth and the average value is about 130 v m^{-1}; (2) the value near the equator is about 80 per cent of the value at higher latitudes; (3) the gradient decreases with altitude in free air, rapidly in the first kilometer, then more slowly, till at an altitude of 10 km it is about 5 per cent of its value at the sur-

face; (4) the gradient at the surface varies during the day according to a universal routine with a minimum at 4^h GMT and a maximum at 19^h GMT and a diurnal range equal to 35 per cent of the daily mean; (5) the character and amplitude of this universal diurnal variation and also the daily mean apparently vary during the year. Exceptions to the last two statements are conspicuous at some times and places, but on a world-wide view the exceptions are presumably insignificant.

THE AIR-EARTH CURRENT

The electric current which flows from air to earth in fair-weather areas is predominantly an electric conduction current. Although it is conceivable that electric convection, as in the case of charged rain falling to earth, or electric displacement which depends upon the rate of change of electric field strength, may sometimes play a part in determining the electric current density between air and earth, yet estimates of the magnitude of the effects of these factors indicate that in fair weather they are usually small and of such character that they need not be considered in this brief review of the phenomena of atmospheric electricity. This is corroborated by several tests in which values of i determined by the direct method were found to be about the same as those determined at the same place and time by the indirect method. The air-earth current density i is, therefore, essentially equal to λE, and the description of it is implied in the foregoing discussions of λ and E.

The average values of i derived from measurements over the oceans on cruises of the *Carnegie* are, in the opinion of the author, fairly representative for the earth as a whole. The bases for this opinion are largely as follows: (1) the geographical distribution of these measurements is much more extensive than that for all other data, and (2) the data are much freer from large and persistent local effects than those for most land stations. It should be noted, however, that for some land stations far from centers of human activity, and notably for those in the polar regions, the data are relatively free from such local effects and have about the same characteristics as those for the oceans. It is on this basis that the following statements rest. The average value of i derived from measurements of λ and E at sea is about 10^{-6} stat amp cm^{-2}. There is no clear evidence of any trend toward either a higher or a lower value in the 15-yr period, 1915–29, to which these data apply. But there is an indication that near the equator i is somewhat less than at higher latitudes [20].

The annual and the diurnal variation of i at sea have about the same character and relative magnitude as the potential gradient because only a small diurnal variation of λ is indicated by the observations and the latter is probably a local time effect. Also there is no definite evidence of an annual variation of λ at sea.

The total electric current I from air to earth for all fair-weather areas is i multiplied by the area of the earth. That the error incurred by neglecting here the area of thunderstorms is less than one per cent may be inferred from the data assembled by Brooks [8].

The mean value of I obtained in this way is about 1800 amp. The error in this estimate is probably not greater than 10 per cent. The diurnal and possible annual variation of I are of the same character and relative magnitude as the corresponding characteristics of the average for i. This total current from air to earth must have a counterpart which "completes the circuit." The term "supply current" is used here to denote this counterpart, an elemental universal feature of atmospheric electricity. Where and how is the supply current generated? The answer to that question has been sought during the last fifty years. The status of that search at the present time is discussed in the following section.

THE SUPPLY CURRENT AND THUNDERSTORMS

Many proposals have been made to account for the fair-weather aspects of atmospheric electricity. Most of these were soon found to be untenable. The one surviving proposal which may give the answer as to *where* the supply current is generated is credited to C. T. R. Wilson. This is the suggestion that the supply current is generated in thunderstorms. If this is the case, the answer as to *how* it is generated doubtless must await the development of an acceptable theory for the generation of electric charge in thunderstorms. This section is devoted to an appraisal of Wilson's suggestion.

The universal aspects of atmospheric electricity which should be recalled in this connection are, expressed in terms of the more comprehensive element I, as follows: (1) the yearly average of the total current I from air to earth in fair-weather areas seems to be nearly constant at a value of about 1800 amp; (2) the daily mean of I probably varies during the year, being greater in the six months from November to April than in the rest of the year; (3) the diurnal variation of I is such that on the average during the year the maximum occurs at about 19^h GMT and the minimum at about 4^h GMT; (4) the character and range of this diurnal variation vary during the year.

The magnitude of the supply current and the character of its annual and diurnal variations must be the same as those listed here for I. Near the earth the supply current must flow upward, from earth to air, (opposite to that for I) but at some undetermined altitude, probably in the high stratosphere and below the ionosphere, the vertical component vanishes and the current is dispersed laterally. Then it returns to earth as an air-earth conduction current which is nearly uniformly distributed over the earth. The circuit is completed through the earth to the source, or sources, of the supply current, which may be located in thunderstorms. The fact that the air conductivity increases with altitude and has relatively large values at high altitudes seems to be of vital importance for the existence of such an electric circuit.

The first evidence in favor of Wilson's suggestion that the supply current is generated in thunderstorms was obtained in an analysis, made by Whipple [22], of the thunderstorm data for the world, assembled by Brooks [8]. That investigation indicates that thunderstorm activity, for the earth as a whole, varies during the Green-

wich day in the same manner as does the air-earth current of fair weather. This evidence is exhibited in Fig. 8, where the ordinates represent the area over which land thunderstorms are in progress at the time of the Greenwich day denoted by the abscissae. By comparing this graph with either of the lowest graphs for potential gradient in Fig. 7, a remarkable similarity will be noted. Whipple also concludes that the character of the diurnal variation changes during the year in about the same manner for the two phenomena. This result indicates that if the net current generated in the representative thunderstorm is directed upward from the earth to the high atmosphere, and if the total current from all storms is great enough (average 1800 amp), the supply current would satisfy the requirements listed above. It is surprising, in view of the character of the data and their scantiness for large areas of the earth, that this result could be obtained. Whipple apparently used satisfactory methods in his analysis. Maybe this result indicates that thunderstorms over the oceans and in sparsely settled regions of the earth do not contribute much to the supply current.

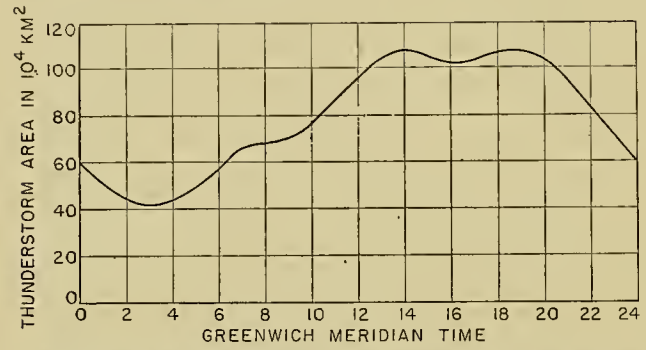


FIG. 8.—Diurnal variation in prevalence of thunderstorms. The ordinate is the estimate of the average area of land-thunderstorms in progress on the whole earth. (*After F. J. W. Whipple and F. J. Scrase.*)

Other methods should be used to check Whipple's results. It seems that it may be practicable now by the use of suitably designed atmospheric recorders at a comparatively small number of well-distributed stations to make a complete "sweep" of the earth and with satisfactory registration for one year to obtain the required data. From such data one would hope to derive a record of lightning frequency as a function of Greenwich time which may be more suitable information for this purpose than records of thunderstorm occurrence.

Although Whipple's results seem to show that *if* thunderstorms are the seat of the supply current, the diurnal and annual variation of that current would each be of the right type, it is yet to be ascertained whether the thunderstorms do contribute a current of the right sign and magnitude to the circuit previously described. In principle this may be ascertained either by making measurements of the required elements at the earth's surface beneath storms or by making the proper survey over the top.

No adequate survey beneath a thunderstorm has ever been made. However, T. W. Wormel made part of the required measurements at Cambridge, England, which he used together with data obtained elsewhere by other observers to draw up a balance sheet of the quantity of electricity lost and gained by 1 km^2 in a year. Wormel's balance sheet is as follows:

Negative charge gained	Coulombs km^{-2} yr^{-1}
1. By natural point discharge	100
2. By lightning discharge	20
Total negative charge gained	120
Positive charge gained	
3. By atmospheric conduction	60
4. By precipitation	20
Total positive charge gained	80
Net gain, negative charge	40

Although it is interesting to see what comes from an attempt to strike such a balance it may not have much significance for the problem at hand. Item 3 is the best-determined one, but Wormel uses the small value obtained at a few places in England whereas the electric surveys of the Department of Terrestrial Magnetism, Carnegie Institution of Washington, indicate a value almost twice this for representative areas of the earth. The interpretation of the data from which Item 1 was obtained may be questioned and, furthermore, it certainly is much too large for vast areas of the earth—especially the oceans and the polar regions. In addition to questioning whether Items 2 and 4 are representative it should be noted that uncertain elements enter into their estimation. This approach to the problem seemed so difficult that another way was sought.

Surveys of the electric current density over the tops of thunderheads seemed to the author to be feasible when in 1946 pressurized aircraft became available for use in scientific projects. Owing to the absence of precipitation and the rarity of lightning in the clear air above a thunderhead, electric circumstances there are simpler than beneath the storm. This was the basis for expecting that the transfer of electricity in the air above the storm occurs chiefly by electric conduction and that, as a consequence, measurements of the vertical component of the electric current density made at short intervals (2 sec) on a number of traverses over a storm would constitute an adequate basis for estimating the magnitude and direction of the total current from a storm. Surveys of this sort were made in 1947 and 1948, as a joint project of the Department of Terrestrial Magnetism, Carnegie Institution, and the U. S. Air Force. A technical report is being prepared by O. H. Gish and G. R. Wait.[5]

Successful surveys of twenty-four storms were made. In these storms the current was directed upward, that is, positive charge was transferred upward. This is favorable to Wilson's suggestion. The magnitude of the current ranged from zero to 6.5 amp. The average current for all storms was 0.6 amp, and that for all except the one unusually large value, was 0.3 amp.

Is an average current per thunderstorm cell of 0.3 to 0.6 amp adequate to satisfy the requirement that the total supply current be 1800 amp? It would be, if the

5. See "Thunderstorms and the Earth's General Electrification," by O. H. Gish and G. R. Wait, *J. geophys. Res.* 55: 473–484 (1950).

average number of thunderstorm cells or centers of electric activity in progress on the earth were between 3000 to 6000. This is several times Brooks' estimate of 1800. The latter estimate is, however, not applicable here for the following reasons. The data used by Brooks consisted of reports of thundery days—days on which thunder was heard—and there is no indication that he took into account the fact that at some places two, and occasionally more, visually distinct thunderstorms occur within hearing distance of a station on the same day. Neither was it considered that in a given thunderstorm there are sometimes several well-separated centers of electrical activity in progress simultaneously. These considerations indicate that a world population of between 3000 and 6000 centers of electrical activity may be admissible, and, that the answer to the foregoing question is "Probably, yes."

This provisional answer is, however, subject to another condition, namely, that the values for total current obtained from these surveys may be too large because part of the current, represented by the measurements, may flow in a local circuit either to the lower pole of the cloud or to bound charge on the earth in the vicinity of the storm. An adequate evaluation of this effect has not yet been completed. Until this is done and until a more reliable estimate can be made of the world population of centers of electric activity, one can say only that the available evidence is favorable for the view that the supply current is generated in the thunderstorms, but much more investigation is required in order to reach a definite and reliable conclusion. This theory has at the present time no serious rival. No other theory known to the author has the potentiality of accounting for the universal, diurnal, and annual variation of I.

An interesting corollary of this theory is that if the supply current originates in thunderstorms, the universal, diurnal, and other variations may be regarded as a measure of the thunderstorm activity of the whole earth. It accordingly seems likely that the record of i at a station where local disturbances are small would constitute an approximate record of the thunderstorm activity of the earth, showing how it varies from day to day and perhaps even from hour to hour. Small local effects are more likely to obscure the latter than the former. Such effects could, however, doubtless be largely eliminated from the data if i were registered at a moderate number of suitably selected stations. Such records might be valuable in the study of some problems of world meteorology. For example, they might provide answers to questions such as, Does world thunderstorm activity vary from year to year and is it correlated with sunspot activity? The data now available for i apparently point toward a negative answer to these questions.

How is the supply current generated? How are the electric charge-clouds of thunderstorms developed? These are two questions which doubtless have a single answer, provided that the supply current is generated in thunderstorms. In the author's opinion, no adequate answer to these questions has yet been published. Because of this circumstance the discussion which follows is limited to an outline of important aspects of the theories which have been proposed, the chief electrical features which must be accounted for, and some other conditions which must be satisfied.

The electrical cycle of a thunderstorm may be regarded as consisting of the following parts: (1) an initial separation of charge, a small-scale phenomenon in which some particles of precipitation become charged with electricity of one sign and the adjacent air, or some of the smaller particles, becomes charged with the other sign, (2) a large-scale separation of charge, possibly occurring in several steps, by which large charge-clouds of several cubic kilometers in volume are formed at more or less definite levels in the thundercloud, (3) the initiation of discharge from a cloud, namely, the mobilization of the charges residing on particles widely distributed throughout the charge-cloud, thus preparing for the succeeding steps which constitute (4) the lightning discharge.

The last of these, the lightning discharge, has been comparatively well elucidated in recent years. It is the subject of a separate article in this Compendium[6] and will not be discussed here.

Knowledge regarding the other steps, however, is in an unsatisfactory state. More factual evidence is required before theories of these aspects of the electric cycle of the thunderstorm can be securely established.

The initiation of a lightning discharge—the getting together of the charges which are widely distributed on ice particles or raindrops or both—although one of the unexplained aspects of the thunderstorm has little relevance to the development of the charge-cloud, and will be dismissed with the statement that the glow discharge, which starts at the surface of charged particles, is presumably involved. When the glow discharge spreads throughout a sufficient volume of the charge-cloud, the discrete charges on the particles are mobilized for the lightning discharge. However, there is little direct evidence to support this surmise.

The electric structure of the thunderstorm is often so complex that exact descriptions, such as could be made with the aid of general harmonic analysis when adequate data are available, have not been undertaken. One might say, however, that in the classical work of C. T. R. Wilson a fair estimate of the principal moment was obtained. An advance beyond this was realized in the work of Workman, Holzer, and associates [23]. The results of these and other investigators are in fair agreement on the gross aspects of the electric charge-clouds. These and some other features which bear on the later discussion are listed here.

1. There are two principal charge-clouds in the typical thunderstorm. The positive cloud is usually located at an altitude greater than that of the negative cloud. An altitude of 6 to 7 km for the former and of 3 to 4 km for the latter has been estimated.

2. Charge-clouds tend to be cumuliform rather than stratiform.

6. Consult "The Lightning Discharge" by J. H. Hagenguth, pp. 136–143.

3. The size of a charge-cloud which has developed to the stage where lightning of average intensity may occur is comparable to that of a sphere of at least 1-km diameter and probably is several times that dimension. For a cloud of smaller size, having a charge of 20 coulombs, the dielectric strength would be exceeded in air containing water drops, ice pellets, or crystals.

4. The region of primary electrical activity (location uncertain) where the initial charge separation and the large-scale separation occur jointly, doubtless has a cross section less than that of the charge-clouds.

5. The average cloud charge developed between lightning flashes is 20 to 30 coulombs.

6. The average rate of regeneration for charge-clouds, after lightning discharges, is about 4 amp but values as great as 20 amp are sometimes indicated. In regeneration the charge approaches an equilibrium value in approximately an exponential manner. One may accordingly speak of a relaxation time. An average value of the latter is about 5 sec. The relatively small departure of individual values from this average seems to have some significance.

7. No convincing evidence has been found of abnormally large electrical conductivity of the air above thunderclouds or under thunderclouds at the earth's surface.

These items will be referred to by number in later paragraphs.

The initial separation of charge has been attributed to various processes: The disruption of drops of pure water will effect a separation of electric charge with positive charge on the drops and negative charge in the air (a few thousandths of one per cent of some salts as an impurity annuls this effect). G. C. Simpson developed a theory which depended on this process. This theory was in vogue for more than a decade, but it became untenable when it was found that the positive charge-cloud is usually above the negative cloud.

Ion-capture is the fundamental process in a theory developed by C. T. R. Wilson. This process may be illustrated as follows. A water drop located in the normal electric field will have a positive charge induced on its lower surface and a negative charge on its upper surface. If the atmosphere contains ions (large ions assumed by Wilson), the positive ions will drift downward and the negative ions upward both with a velocity much less than that of the falling drop. The negative ions will be attracted by the positive charge on the bottom of the drop and those located in or near the path of the drop will be captured. The positive ions, however, will be repelled by the charge on the bottom of the drop and escape capture because when such an ion is in position to be attracted by the negative charge on the top of the drop, that attraction is not great enough to effect a capture. In this way larger drops may acquire a net negative charge. The larger drops with their negative charge will accumulate in the lower part of the thundercloud while positive charge, without involving the ion-capture process, will accumulate at a higher level. This is a favorable aspect of Wilson's theory, since it is in accord with the observed orientation of the principal charge-clouds. Although there are reasons for thinking that pellets or crystals of ice may also capture ions in this way, this requires more investigation if the ion-capture process is to be considered active in the region of sub-zero (centigrade) temperature where the principal charge-clouds are found. This ion-capture process requires that a relatively large concentration of ions, preferably of low mobility, obtains in the electrically active part of a storm cloud. Such a condition has not yet been observed, except as a local phenomenon of relatively rare occurrence, namely at times when glow or brush discharge (St. Elmo's fire) occurs. Unless there is some potent source of such ions other than the glow or spark discharge, Wilson's ion-capture process can act only in a secondary role after fields capable of initiating glow discharge have been developed by another mechanism.

The collision of ice particles has been suggested as a process for the initial charge separation. Charge developed by drifting snow (positive in the air) is more conspicuous than that for splashing rain but it is doubtful whether this process is effective in the atmosphere remote from the earth's surface.

Evaporation, condensation, and sublimation acting singly or in combination have been assumed as primary processes [11, 16]. Findeisen's experiments [11] indicated that these processes are much less effective than is the process which is involved in the formation of sleet (*Vergraupelung*). Dinger and Gunn [9] found an effect associated with the freezing of water and the melting of ice. Gunn [16] also assumed that a raindrop is essentially a concentration cell, and he indicated how it may acquire a net negative charge during condensation and a net positive charge during evaporation. Frenkel [13] developed a theory in which the electrokinetic potential of a raindrop, or cloud droplet, was proposed as the basis for the initial separation of charge. The fundamental element of this theory is similar to that of Gunn's theory.

A very active process of charge separation was discovered by Workman and Reynolds [24]. This occurs during orderly freezing of water in which very small quantities of certain salts are dissolved. The effectiveness and the direction of the process depend upon the concentration and nature of the solute. In most solutions for which a pronounced effect was reported, negative ions are captured by the ice. A prominent exception to this was found in solutions of the ammonium salts for which the solid phase acquires a positive charge.

The charge developed by this process in the freezing of one cubic centimeter of water is extraordinarily large for a number of solutions which were examined. For solutions of *NaCl* the greatest effect was for an 0.83×10^{-4} normal concentration. This gave a charge of 9.2×10^{4} stat coulombs from the freezing of one cubic centimeter of solution. The solid phase acquired a negative and the liquid phase a positive charge. The largest value reported was for a *CsF* solution of 11.1×10^{-4} normal concentration. This is 44×10^{4} stat coulombs cm^{-3}. These values are of a much greater order of magnitude than the largest value reported by Lenard and

associates for the violent disruption of a water drop, the latter being about 2 stat coulombs for each cubic centimeter of water that is involved. Perhaps the results reported by Findeisen and by Dinger and Gunn are in part attributable to the orderly freezing of water.

Other investigations of Workman and Reynolds indicate that cloud droplets may contain solutes of the right type and in suitable concentration for this process to occur in the atmosphere. Favorable results were also obtained in an experiment designed to imitate the growth of hail.

The advantage which derives from the relatively large amount of electric charge separated in the orderly freezing of suitable solutions may be illustrated by the following simple calculation. If the amount of charge separated in the formation of one gram of hail or sleet is 10^4 stat coulombs (about 2 per cent of the largest value reported), then in order that the rate of charge regeneration of a charge-cloud be 4 amp, or 1.2×10^{10} stat amp (Item 6, of the foregoing list), it is necessary that at least 1.2×10^6 g (about one short ton) of hail be formed each second in the region of primary electrical activity. In contrast to this, the mass of water drops that would have to be violently disrupted each second, if the breaking-drop process were the basic factor, would be at least 6×10^9 g or more than 6500 short tons. This apparently shows that the breaking-drop process is not adequately active, but that generation of charge by orderly freezing may be sufficiently active provided that, among other conditions already indicated, hail is always a large constituent of the hydrometeors in a typical thunderstorm.

None of the other primary processes so far proposed has yet been shown to have the generating capacity required. Acceptance or rejection of either Gunn's or Frenkel's theory depends on whether or not the relaxation time of the process is adequate, that is whether $1/(4\pi\lambda)$ in the region of primary activity, is, on the average, about 5 sec (Item 6). This is equivalent to saying that these theories are not acceptable unless the air conductivity in the region of primary activity is at least ten times that for normal air at an altitude of 5 km. At present it seems unlikely to the author that this condition is satisfied, but since this opinion is based chiefly on indirect evidence, more direct exploration in the future may bring forth evidence which contradicts this view.

The ion-capture process which is elemental in C. T. R. Wilson's theory also postulates that a relatively high concentration of ions prevails in the region of primary activity, where the initial separation of charge occurs. Since these ions are assumed to have a very small mobility, it seems likely that the air conductivity would be abnormally small in some parts of the cloud. This postulate cannot be definitely refuted by evidence now available.

The large-scale separation of charge which follows the initial step in charge generation must also proceed at the rate of 4 amp for each typical center of electrical activity. In all theories mentioned in this article it is postulated (1) that after the initial step the larger drops, or particles of precipitation, tend to have an electric charge of sign opposite to that of very small particles or of air ions, and (2) that, principally under the action of gravity, large particles fall away from the small ones at a velocity v which is equal to the difference in the terminal velocities.

Although there are no obvious alternatives to these postulates, there seems to be some ground for doubting whether the second is acceptable. This is illustrated in the following paragraph.

Let ρ denote the total net charge on the large particles in a cubic centimeter of air; v, the average velocity of these particles relative to the smaller ones; and A, the cross-sectional area, normal to the direction of v, of the region of charge separation. The total current I' from large-scale separation then is $I' = \rho v A$. Now in order that the value of I' may be 4 amp, $\rho v A$ must equal 1.2×10^{10} stat amp. The space charge ρ is limited by several circumstances. One which is amenable to simple treatment is that for no considerable proportion of the drops or particles shall the charge q of each drop of radius r be greater than $100r^2$. Larger values lead to electrical discharge. If there are n drops in each cubic centimeter, all of the same size and same charge, the maximum admissible space charge is $\rho_m \leqq 100nr^2$.

The mass of drops m in 1 cm^3 of air is also limited, and in terms of this m, the foregoing expression for the upper limit of ρ may be written $\rho_m \leqq 300m/(4\pi r)$. Now v is an increasing function of r, but in such a way that $\rho_m v$ *decreases* with an increase of r if m is constant. For a very large value of m, namely, 5×10^{-6} g cm^{-3}, $\rho_m v$ is 0.25 for hailstones having a diameter of 3 cm, and is 1.1 for droplets of 0.4-mm diameter. If the intermediate value 0.5 for $\rho_m v$ is used, one finds that $A \geqq 2.4$ km^2. Such a value for A is of satisfactory magnitude, but in view of the assumptions made here this estimate is doubtless much smaller than is actually required. It seems unlikely that in nature a large proportion of the precipitation particles are highly charged at a given time, and it is also doubtful whether such a large concentration of water, in either the liquid or the solid phase, occurs in a typical thunderstorm. The effect of the electric field, which tends to reduce v, especially if the particles are small, is also not considered here. Because of these and other considerations it seems evident that the value required for A is much larger than 2 or 3 km^2. But if this conclusion is correct, that would entail the difficulty of accounting for the size, structure, and orientation of the charge-clouds—features which are, at least roughly, indicated by measurements of the electric field above thunderclouds, within them, and below them at the earth's surface.

The object of the foregoing statement is to indicate how unsatisfactory is the present status of theories regarding the large-scale separation of charge in thunderstorms. No theory is yet secure if settling under the action of gravity is assumed to be essential in the large-scale separation of charge. If observations eventually show that the concentration of water in the typical thunderstorm is considerably greater than 5 g m^{-3}, at least in the region of primary electrical activity, this

difficulty might be removed. But it may be that some force other than gravity effects the separation of the larger from the smaller charged particles. A combination of centrifugal action and straight wind has been suggested, but it is not yet evident that the wind structure of the thundercloud is suitable for this.

This discussion of selected electric features of thunderstorms and of theories which have been proposed to account for those features was designed to show that at the present time there are only a few clues as to how the charge-clouds are developed. One of these is the recent discovery that the orderly freezing of very dilute solutions of certain salts is a very effective process for the initial step in the generation of electric charge. The potency of this process is such that if conditions in the thunderstorm are favorable, the charge on the precipitation particles may be maintained at a value so near the maximum, limited by electric breakdown, that the prospects of accounting for the large-scale separation of charge will be improved. More observations of electrical and other conditions, especially within thunderstorms, and more carefully controlled speculation are required to find the answer to the question, How are the charge-clouds in thunderstorms developed?

Until the foregoing question is answered, the added question, How is the supply current generated? will probably also remain unanswered. This statement depends upon the fact that with the evidence now at hand one may entertain the view that the supply current is generated in the thunderstorms of the earth. Before this view was advocated, the fair-weather aspects of atmospheric electricity and thunderstorm electricity were regarded as unrelated geophysical phenomena. Now, since it seems likely that the universal aspects of atmospheric electricity derive from thunderstorms, a more unified exposition of the subject is feasible.

From a remote position in space an ideal observer, whose acute vision could encompass the whole earth, might see in the broad prospect of atmospheric electric phenomena the following features: first, the numerous thunderstorms in progress on the earth—usually several thousand of them. He would notice that they are very scarce in the polar regions and especially abundant in the afternoon on land areas in middle and low latitudes. If each hour throughout a year he should count the total number of electric storm centers, he would probably note that, on the average, the count tends to be greatest for the hour when it is mid-afternoon at about longitude 75°W and least for the hour when it is mid-afternoon at about longitude 150°E. Other variations in the count would doubtless also be found.

If this observer had a special sense with which he could "see" a flux of electricity, he would not only notice the lightning flashes in and about the electric storm centers, but would also see a complicated pattern of electric flux in, about, and beneath each electric storm center. But the most alluring feature would be a narrow stream of the electric effluvium which emerges from the top of the upper charge-cloud, flows upward and along its course, widens and becomes less dense until it merges at a high level with similar streams from the other storms to form a world-wide ocean of electric effluvium. From this ocean a much more tenuous but nearly uniform electric flux could be seen to proceed downward to the earth everywhere except in and about electric storms. The circuit continues through the earth and back to the storm centers. The density of this universal flux from air to earth would be seen to vary during the day, from day to day, and during the year, in about the same manner as does the corresponding count of storms. But apparently it varies little, if at all, from year to year.

This is an impressionistic sketch of the broad prospect of atmospheric electricity as it is seen by the author in the light of evidence now available.

OUTSTANDING PROBLEMS

In the broadest sense there are two main problems in the field of atmospheric electricity:

1. To locate the source of those universal aspects which are epitomized in the concept, supply current.

2. To elucidate the mechanism which generates the supply current. If, as now seems likely, the supply current is generated in thunderstorms, the last-mentioned problem and that posed by the electric aspects of the thunderstorm have much in common.

It is of course necessary to have an adequate quantitative description of the phenomena before the rationale of the subject is developed. In this respect there are many minor and some major inadequacies. Some of the more important observations required in order to fill this need are:

1. Measurements of air conductivity in the high atmosphere, especially in the altitude range 18 km to 60 km. New techniques would be required to make the measurements in the upper part of this region.

2. Measurements of air conductivity, and counts of Aitken nuclei in thunderstorms. These items are suggested, rather than measurements of small-ion and large-ion concentrations, because measurement of the latter elements is more difficult. The technical difficulties of making reliable measurements of air conductivity from aircraft, during flight through clouds, have not yet been overcome.

3. More extensive, or more comprehensive, information about the population of the electric storm centers that are in progress on the earth. Until this information is available, no reliable estimate can be made of the average current which a typical electrical storm center must contribute to the make-up of the supply current. It is desirable that data be collected to verify, or to refute, the diurnal variation in storm population reported by Whipple. Perhaps this could be done by radio methods used for measuring atmospherics and "background noise," with some modification to adapt them for use in a world survey of electric storm activity. Apparently it would be necessary to make such measurements at a relatively small number of well-distributed stations.

4. More data for the air-earth current density i or for the elements λ and E, in order to ascertain whether, and in what manner, the supply current varies during

the year. The evidence now at hand is inadequate for a reliable determination of even the qualitative aspects of this feature. Except for measurements made on the cruises of the *Carnegie*, data for i have been obtained at only a few places.

5. Information which will help better to ascertain the electric structure of the thunderstorm. More measurements of electric field in the vicinity of thunderstorms (made simultaneously at a number of stations) are required for more types of storms. More measurements, made within storms, of the electric charge distribution are desirable.

6. More determinations of the rate of regeneration of charge, after a lightning discharge. Although the data now at hand appear to be comparatively reliable, they should be checked because they set a requirement which may not be satisfied by any theory which depends upon the force of gravity for the large-scale separation of charge.

7. Determination of which of the various known processes of initial charge separation occur in a thunderstorm. It will doubtless be difficult to find a definite answer to this, but the answer should be sought.

8. Information on how the widely scattered discrete charges in a charge-cloud are mobilized for the lightning discharge. This is another question for which a more definite answer is awaited.

These are some of the matters in atmospheric electricity which deserve attention. Others have been indicated in the body of this article.

REFERENCES

This brief list of treatises and papers is designed chiefly to serve as a guide for the reader or investigator who desires to read more comprehensive discussions of the subject. Technical articles which contain good bibliographies have been given preference. These citations, however, should not be used as a basis for assigning priority of or credit for discovery.

I. Treatises.

1. Fleming, J. A., ed., *Terrestrial Magnetism and Electricity.* New York, McGraw, 1939. Reprinted with corrections: New York, Dover Publications, 1949. The following chapters bear on atmospheric electricity:
   Gish, O. H., "Atmospheric Electricity," Chap. IV.
   Torreson, O. W., "Instruments Used in Observations of Atmospheric Electricity," Chap. V.
   Berkner, L. V., "Radio Exploration of the Earth's Outer Atmosphere," Chap. IX.
   Schonland, B. F. J., "Thunder-Clouds, Shower-Clouds and Their Electrical Effects," Chap. XII.
   Harradon, H. D., "Bibliographical Notes and Selected References," Chap. XIII. This bibliography contains references to other general treatises and to many technical papers published prior to 1939.

II. Treatises Published since 1939.

2. Chalmers, J. A., *Atmospheric Electricity.* New York, Oxford, 1949.
3. Israël, H., *Das Gewitter.* Leipzig, Akad. Verlagsges., 1950.
4. Maurain, C., *La Foudre.* Paris, A. Colin, 1948.
5. Mitra, S. K., *The Upper Atmosphere.* Calcutta, The Royal Asiatic Society of Bengal, 1948.

III. Technical Papers and Reviews.

6. Ault, J. P., and Mauchly, S. J., "Ocean Magnetic and Electric Observations Obtained aboard the *Carnegie*, 1915–21." Res. Dept. Terr. Magn., *Carneg. Inst. Wash. Publ.* No. 175, 5: 197–286 (1926).
7. Bricard, J., "L'Équilibre ionique de la basse atmosphère." *J. geophys. Res.*, 54: 39–52 (1949).
8. Brooks, C. E. P., "The Distribution of Thunderstorms over the Globe." *Geophys. Mem.*, 3: 145–164 (1925).
9. Dinger J. E. and Gunn, R., "Electrical Effects Associated with a Change of State of Water. *Terr. Magn. atmos. Elect.*, 51: 477–494 (1946).
10. Everling, E., und Wigand, A., "Spannungsgefälle und vertikaler Leitungsstrom in der freien Atmosphäre, nach Messungen bei Hochfahrten im Freiballon." *Ann. Physik.*, 66: 261–282 (1921).
11. Findeisen, W., "Über die Entstehung der Gewitterelektrizität." *Meteor. Z.*, 57: 201–215 (1940).
12. Forbush, S. E., Stinchcomb, T. B., and Schein, M., "The Extraordinary Increase of Cosmic-Ray Intensity on November 19, 1949." *Phys. Rev.*, 79: 501–504 (1950).
13. Frenkel, J., "A Theory of the Fundamental Phenomena of Atmospheric Electricity." *J. Phys.* (*U.S.S.R.*), 8: 285–304 (1944).
14. Gish, O. H., and Sherman, K. L., "Electrical Conductivity of Air to an Altitude of 22 km." *Nat. Geogr. Soc. Contrib. Tech. Papers, Stratosphere Ser.*, No. 2, pp. 94–116 (1936).
15. —— *Ionic Equilibrium in the Troposphere and Lower Stratosphere.* Internat. Assoc. Terr. Magn. Elect., Washington Assembly, Sept., 1939.
16. Gunn, R., "The Electricity of Rain and Thunderstorms." *Terr. Magn. atmos. Elect.*, 40: 79–106 (1935).
17. Schweidler, E., *Luftelektrizität. Einführung in die Geophysik*, Bd. II. Berlin, J. Springer, 1929. (See pp. 291–375)
18. —— *Die Aufrechterhaltung der elektrischen Ladung der Erde.* Hamburg, H. Grand, 1932.
19. Scrase, F. J., "The Air-Earth Current at Kew Observatory." *Geophys. Mem.*, Vol. 7, No. 58 (1933).
20. Torreson, O. W., and others, "Scientific Results of Cruise VII of the *Carnegie* during 1928–1929. Oceanography, III, Ocean Atmospheric-Electric Results." *Carneg. Instn. Wash. Publ.*, No. 568 (1946).
21. Wait, G. R., "Atmospheric-Electric Results from Simultaneous Observations over the Ocean and at Watheroo, Western Australia." *Trans. Amer. geophys. Un.*, 23: 304–308 (1942).
22. Whipple, F. J. W., "Modern Views on Atmospheric Electricity." *Quart. J. R. meteor. Soc.*, 64: 199–213 (1938).
23. Workman, E. J., Holzer, R. E., and Pelsor, G. T., "The Electrical Structure of Thunderstorms." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 864 (1942).
24. Workman, E. J., and Reynolds, S. E., *Thunderstorm Electricity.* Final Rep., Signal Corps Res. Contract No. W-36-039sc-32286, Sept. 30, 1948. (See p. 26)

IONS IN THE ATMOSPHERE

By G. R. WAIT

Carnegie Institution of Washington

and W. D. PARKINSON

Johns Hopkins University

In considering the subject of atmospheric ions, it has been possible, because of space limitations, to consider only those conditions of the atmosphere which are regarded as normal and to include in the discussions, which are of necessity brief, only those items which are most important. References to investigations have also been restricted to those regarded as most pertinent to the subject at hand.

It has been known since late in the eighteenth century that an insulated charged body left in the air will slowly lose its charge by a process other than leakage across the supporting insulators. This process of gaseous conduction of electricity was interpreted by J. J. Thomson [42] as due to the presence in the gas of charged particles which, by analogy with the conduction in electrolytes, were called "ions."

The concentration of ions in the atmosphere is measured by an instrument called an *ion counter*. In principle the instrument is simply a charged cylindrical condenser through which air is drawn at a known velocity. The rate at which the charge on the central electrode changes provides a measure of the number of ions in the air stream and consequently the ion concentration of the air. If the voltage between the electrodes is sufficiently high, all ions from the air stream will be collected. Normally this voltage difference is selected so as to collect all ions of a particular mobility group and as few as possible of the less mobile ions. If a measure of the concentration of a lower-mobility group is desired, interference from ions of higher mobility may be avoided by first passing the air through another cylindrical condenser operated with sufficient voltage to remove the more mobile ions but only a small proportion of those to be measured. The theory and operating details of ion counters are discussed in special articles and texts on the subject [6, pp. 252–258; 10, 28, 41].

Ionic Mobilities

An electron is released in the formation of a pair of small ions, but remains free for only a short time. The fact that no high-mobility group of negative ions is found shows that very few free electrons are present in the lower atmosphere. Small ions of the atmosphere have an average life of the order of a minute, so they are thoroughly aged during most of their lives, and are subject to the influences of many atmospheric constituents existing in minute quantities. According to the picture recently given by Overhauser [34], an atmospheric small ion can be imagined as a charged molecule which is continually associating with, and dissociating from, one or more other molecules. The time a certain type of molecule remains associated with an ion depends on its electrical properties.

The average velocity with which an ion drifts through a gas under the influence of an electric field is proportional to the strength of the field. The ratio of the velocity to the field strength is called the *mobility* of the ion. The unit of measurement (referred to hereafter as centimeters) is centimeters per second per volt per centimeter.

Numerous mobility measurements on gaseous ions have been made in the laboratory [22]. The value of the mobility is found to depend upon both the ion and the gas through which it moves. It also is affected by the presence of very slight traces of water vapor and other impurities. The ion can apparently become attached to, or lose its charge to, an impurity molecule. Some molecular impurities are known to associate readily with a positive ion, some with a negative ion, and some with either. In most laboratory measurements, an effort is made to remove all traces of water vapor or other impurities so the values may be representative of the gas. For freshly formed positive and negative ions in air, Erikson [5] obtained a value for the mobility of 1.87 cm. In a few hundredths of a second the positive-ion mobility had decreased to 1.36 cm, due, he believed, to the ion becoming attached to a neutral air molecule. Bradbury [3] obtained what he considered a more representative value for each sign after taking extreme precautions to remove practically all traces of impurities. Values obtained by him were 2.21 and 1.60 cm as the negative and positive ion mobilities, respectively.

This requirement for such a high degree of purity for the air in mobility determinations leads one to question the extent to which laboratory-determined values can be accepted as the values of the small-ion mobilities in atmospheric electric work.

Mobility determinations of the small ions in the atmosphere have been made by various methods. Probably the method most frequently used is the ratio of air conductivity to small-ion content. This will not result in high precision unless proper precautions are taken to eliminate various errors which can easily enter and unless a correction is made for the lower-mobility ions caught by the ion counter.

The published values [14] of mobilities of small ions in the atmosphere, from measurements made many years ago, show a great deal of scatter. There are reasons

for believing that in many of these determinations the precision was not especially high, and in some cases it seems probable that considerable error resulted from the collection of lower-mobility ions by the ion counter (resulting in too low a value for the mobility). The latter was probable because at one time it was customary to apply to the ion counter a rather high potential, frequently as much as several times that required for small-ion saturation. A list of a few mobility values in air, from measurements where it is known that precautions were taken to avoid various possible errors, is given in Table I. The several values are in quite good

Table I. Some Mobility Values for Small Ions

Reference	Station	Period	Mobility in cm sec^{-1}/volt cm^{-1}	
			$k+$	$k-$
[44]	*Carnegie*, Cruise VII	1928–1929	1.30	1.39
[39]	College, Alaska	1932–1933	1.45	1.75
[14]	Canberra, Australia	1934	1.29	1.40
[29]	Glencree, Ireland	1937	1.28	1.40

agreement, especially in view of the very large difference in meteorological conditions at the various stations and the possibility that such conditions might affect the mobility of the atmospheric ion.

Since the mobility of an ion is affected by impurities in the gas, it seems likely that atmospheric ions may be affected by impurities in the air. A sudden change in mobility, believed to be due to such causes, was found by Parkinson [35] at Huancayo, Peru (altitude 3300 m, mean pressure 518 mm). From carefully controlled mobility determinations and from simultaneous measurements of air conductivity and ion concentrations which were corrected for the lower-mobility ions caught by the ion counter, it was found that a sudden drop in mobility took place each morning at about seven o'clock local time. This drop occurred at the precise time when there was a large influx of molecular impurities into the atmosphere. The average values of $k+$ and $k-$ for the 7-hr period before the influx of impurities were 2.40 and 3.46, respectively. The mean values for the 7-hr period immediately following the influx were 2.00 and 2.35, respectively. The change in negative-ion mobility was nearly twice as great as the change in positive-ion mobility. Thus it appears that the molecular impurities at Huancayo associate more readily with the negative than with the positive ion.

Impurities in the atmosphere tend to be graded as to size, so that ions of a mobility lower than that of the small ion are often, although not always, found within rather narrow mobility ranges. Pollock [37] in Sydney, Australia, observed a group (intermediate ion) with a mobility between 0.1 and 0.02 cm. A similar mobility group has been found in other localities [14, 16, 47]. In some localities, on the other hand, there appears to be no such group [52, 56]. A lower-mobility group (around 10^{-4} cm) first detected by Langevin [21] has since been observed in a number of localities, and the ions in this group are now called the "Langevin," or large, ions of the atmosphere. Ions of still lower mobility have been observed in some localities [17].

Pollock found that the mobility of the intermediate ion was a function of the vapor pressure of the atmosphere, diminishing from around 0.1 to about 0.02 cm when the vapor pressure increased to about 17 mm, whereupon the ion was suddenly transformed into the large ion. Both Hogg [14] and Wait [47] examined this matter and neither found any tendency for the intermediate ion to be transformed into the large ion at any vapor pressure. Wait, however, found that the mobility of the intermediate ion was a function of the vapor pressure. At a pressure corresponding to 0 mm the mobility was about 0.5 cm while at 30 mm pressure it was only 0.05 cm.

Yunker [56] found a continuous mobility distribution in California. He presented a distribution in which the number of ions per mobility interval increases with increasing mobility. His results on artificially ionized air indicate that the concentration of ions of mobility down to 7×10^{-4} cm varies inversely with the nuclei concentration, which shows that the ions formed by charging of condensation nuclei have a still lower mobility.

Nature of Slow Ions

It is usually considered that a large ion is a charged condensation nucleus of the type discovered by Aitken [1]. This ion is usually singly charged, but multiply charged ions can exist [5, 15]. In general, nuclei appear to consist of some hygroscopic core around which a stable agglomeration of water molecules can form. Landsberg [20] has presented a detailed discussion of what is known of their properties. Sulphuric acid (a common product of combustion) and oxides of nitrogen probably play important parts in the formation of nuclei. Wright [55] discusses the matter of salt spray forming the core material for a nucleus which appears to be somewhat larger and is important in the atmosphere over the oceans and at some coastal places, but less important in most inland localities.

Pollock [37], in his announcement of the existence of intermediate ions, considered them to consist of a "rigid nucleus enveloped by a dense atmosphere of water vapour." Since the intermediate ion appears to exist only in certain localities, its nature probably varies greatly. If, as found by Wait [47], its mobility varies with water-vapor pressure in accordance with Blanc's law, then the size of the ion is probably constant. This argues against a hygroscopic ion. Hogg [16], working in London, detected intermediate ions the sizes of which were multiples of an aggregate of some 2000 molecules. He assumed them to be composed of sulphuric acid.

Rate of Ionization

The term *ionization* as used here refers to the production of ions, and not to ion concentration (as is sometimes the case). Small ions may be produced in air by a variety of methods, such as by chemical and mechanical means, by ultraviolet light of sufficiently short wave

length, by breaking drops, and by drifting snow and dust. Under particular circumstances one or more of the above-mentioned processes may add appreciably to the ion population of the air. None of these, however, are normally important as ionizers of the atmosphere. The chief ionizer in the lower stratosphere and the troposphere, except in the lower atmosphere over land, is the cosmic rays. In the lowest kilometer or so over land, ionization of the air is due chiefly to radiations from radioactive matter in the earth and in the air.

The average rate of ionization of the atmosphere q, over land and near the earth's surface, has been estimated by Hess [10, pp. 167–171] at about 10 pairs per cc per sec. This value is based on the average amount of radioactive matter in the earth and in the air and is summarized in Table II. According to this estimate, approximately half of the total ionization is due to radioactive matter in the air, one third is due to radioactive matter in the soil, and one sixth due to cosmic rays. The ionization due to cosmic rays and radioactive matter in the soil is probably more or less constant with time at a given station. That due to radioactive matter in the air, however, is subject to variations, since the quantity of radioactive matter in the air varies with time. The amount of radioactive matter in the air depends upon two factors: (1) the rate at which it is dissipated in the atmosphere, and (2) the rate of exhalation from the soil.

Table II. Ionization of the Air Near the Earth's Surface Over Land in Per Cent of Total

Ionizer	Ionizing ray				Total	
	α	β	γ	Cosmic rays		
Radium } in air	30	1	1		32	51
Thorium } in air	18	1	—		19	
Radium } in soil	—	1	32		33	
Thorium } in soil						
Cosmic rays				16	16	
Total	48	3	33	16	100	

The rate of exhalation of radioactive gas from the soil is subject to considerable variation, being affected by such factors as temperature of the soil, wind force, dryness of soil, and covering on the ground [4, 57, 58]. The rate of dissipation in the atmosphere depends upon several factors, but particularly upon air turbulence. Zeilinger found a diurnal variation in the rate of exhalation from the soil, a maximum occurring in the morning hours and a minimum in the evening. A diurnal variation curve with similar characteristics has also been found for the radon content of the atmosphere [57].

A systematic diurnal variation in the rate of ionization of air near the earth over land would be expected, with a maximum occurring in the morning and a minimum in the evening. Continuous observations on the rate of ionization have been reported at three stations, Canberra, Australia [13], Washington, D. C. [48], and Huancayo, Peru [35], in which diurnal variation curves were obtained. A curve for each of the three stations is shown in Fig. 1. The curves are of the type expected when each is plotted on local time. Slightly different methods were employed in measuring the ionization.

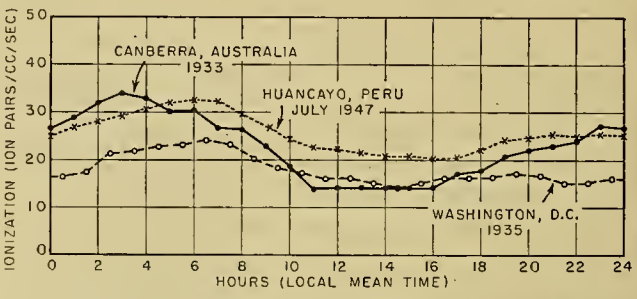


Fig. 1.—Daily variation in ionization of the lower atmosphere.

Hogg [13] measured the ionization of the air after it was drawn into a thick-walled sealed chamber. Wait [48] and Parkinson [35] measured the ionization in a thin-walled ionization chamber (the stopping power of the wall for alpha particles was equivalent to 1.5 cm of air). The values plotted for Huancayo represent the rate of ionization inside the chamber and must be multiplied by about 1.6 to correspond to the ionization of the atmosphere.

Small-Ion Balance

The processes of ion formation just described are balanced by the processes of small-ion destruction. One such process is the recombination between small ions of opposite signs. Ions of other mobility groups will also be active in neutralizing small ions. This is particularly true of the large ions. A third process is one not of neutralization but of the conversion of a small ion into a large ion by coalition of a small ion and a neutral condensation nucleus. The equation of ion balance for positive small ions, taking into account the processes mentioned so far, is

$$q = \alpha n_1 n_2 + \eta_{12} N_2 n_1 + \eta_{10} N_0 n_1. \qquad (1)$$

There is an analogous equation for the negative small ions. The symbols have the following meanings: q is the rate of small-ion formation (expressed in ion-pairs per cc per sec); α is the recombination coefficient for small ions; n_1 and n_2 are the concentrations of positive and negative small ions, respectively (in ions per cc); N_2 is the concentration of negative large ions; and N_0 is the concentration of neutral condensation nuclei. The constants η_{12} and η_{10} are known as *combination coefficients*.

In general the small-ion concentration is accounted for by such an equation. In some places it is necessary to take into account the effect of intermediate ions, but their effect is always small compared to that of the large ions and uncharged condensation nuclei. Over land, the term due to recombination is generally small compared to the other terms, and equation (1) can be written:

$$q = n_1(\eta_{12} N_2 + \eta_{10} N_0). \qquad (2)$$

Now the balance of large ions, shown later, requires that

$$\eta_{10}N_0n_1 = \eta_{12}N_1n_2, \tag{3}$$

and similarly for the negative ions:

$$\eta_{20}N_0n_2 = \eta_{12}N_2n_1. \tag{4}$$

If one further assumes that (3) is equal to (4), then (2) can be rewritten as

$$q = 2\eta_{12}n_1N_2, \tag{5}$$

so that, if changes in q are neglected, the value of n_1 is approximately inversely proportional to N_2.

Processes which cause the destruction of small ions other than those involving combination with small or large ions or condensation nuclei can play an important part under some conditions. Large particles such as smoke and dust can remove small ions [2, 48]. In the atmosphere over the oceans the most important process of destruction of small ions is recombination, but here the influence of nuclei is perceptible in the balance of small ions [44]. In this connection it is interesting to mention the results of a long series of observations of atmospheric conductivity over the oceans. A steady decrease has been noted since 1912, indicating that even here there is a gradual accumulation of atmospheric pollution, due apparently to the increasing industrialization over the world [48].

Recombination of Small Ions

This subject has been extensively studied, both theoretically and experimentally. The reader is referred to the exhaustive treatment given by Loeb [22], and by Jaffé [18]. The subject cannot be pursued here; however, a word about columnar recombination is in order. In the high ion-density track or column left by an alpha particle there will be a high rate of columnar recombination. Thus the effective rate of ion formation is less than that which actually occurs. Only those ions which escape from the column have a life of any appreciable length. The number escaping will be increased by winds and eddy diffusion. In atmospheric electricity this matter has received but little consideration. It has been discussed by Nolan [31].

The value of the effective recombination of small ions is usually taken as 1.6×10^{-6} cc per sec for air at atmospheric pressure [6, p. 178]. However, Luhr and Bradbury [23] give a value of 1.23×10^{-6}, Sayers [38] gives 2.4×10^{-6} and Nolan [31] gives a value of 1.4×10^{-6}.

Combination of Small Ions with Condensation Nuclei

The theory of small-ion combination with condensation nuclei or large ions has received less attention than has the theory of recombination. If all variables in the equations of small-ion balance were measured, it would be possible to determine the value of combination coefficients. Usually this is not done, but relations between parameters are assumed. If it is asumed that equation (3) is equal to equation (4), it follows that:

$$N_0/N_1 = \eta_{21}/\eta_{20},$$
$$N_0/N_2 = \eta_{12}/\eta_{10},$$
$$N_1/N_2 = \eta_{12}n_1/\eta_{21}n_2.$$

If we assume that $N_1 = N_2 = N$, we obtain the relationships:

$$N_0/N = \eta_{21}/\eta_{20} = \eta_{12}/\eta_{10},$$
$$n_1/n_2 = \eta_{20}/\eta_{10} = \eta_{21}/\eta_{12},$$

which were assumed by Nolan and de Sachy [26]. Whipple [53] has deduced the following equations, which are sometimes used as an auxiliary relation between parameters:

$$\eta_{12} = \eta_{10} + 4\pi ek_1,$$
$$\eta_{21} = \eta_{20} + 4\pi ek_2, \tag{6}$$

where k_1 and k_2 represent the mobility of the positive and negative small ion, respectively.

Almost every observer has used a different method of determining these combination coefficients. Gish and Sherman [9] gave a thorough discussion of these methods together with a table of values prior to 1940. More recent values have been summarized by Parkinson [36], and Table III shows representative values.

Table III. Combination Coefficients (in units of 10^{-6})

	η_{10}	η_{20}	η_{12}	η_{21}
Least value	0.4	0.6	2.2	2.5
Greatest value	6.8	7.6	8.7	9.7
Median value	0.6	1.1	2.4	4.5

It is not surprising that these values vary so widely when it is remembered that various methods of determination have been used. Also it is not reasonable to expect that the combination coefficients would be precise universal constants since they depend upon the nature (*i.e.*, the mobility) of slow ions, which will vary with conditions and hence from place to place. Also, no notice is usually taken of the distribution of mobilities involved in the ions responsible for the destruction of small ions, nor is an allowance made for intermediate ions as distinct from large ions.

Several investigators [32, 33] have found that the age of the large ion has an influence on the value of the combination coefficient, the coefficient becoming higher the older the nuclei. This is probably due to an increase in the size of the nuclei with age, owing to coagulation.

Large-Ion Balance

If we assume that a distinct group of large ions exists, then we can speak of their balance just as we do in the case of small ions. One method of formation of large ions is the process which has been mentioned above as one which destroys small ions, namely the fusion of a small ion with a neutral condensation-nucleus. The other process of destruction is the neutralization of charge by a small ion of opposite sign. There will also be a recombination term, similar to that for small ions, and a linear term ζN to account for diffusion [30]. We can write Q for the formation of large ions by other

methods; then the equation of positive large-ion balance becomes

$$dN_1/dt = Q + \eta_{10}n_1N_0 - \eta_{21}n_2N_1 - \gamma N_1N_2 - \zeta N_1. \quad (7)$$

There will be a similar equation for the negative large ions. In this case the value of Q will not necessarily be equal for the positive and negative large ions, for the situation is not as simple as for small ions where the formation is principally due to a process wherein an ion pair is produced.

Kennedy [19] and Nolan [30] found values of γ of the order of 10^{-9}. Hogg [12] reported a value about ten times this, but failed to take into account coalition of neutral nuclei and diffusion or falling out of larger nuclei as a means of dimunition of nuclei. Wait and Torreson [51] found that the value of γ varies with the age of the ion. In a later study, Wait [48] found that for singly charged ions the value of γ could be expressed as a function of the mobility k of the ion. A provisional value of this relationship is given by the equation

$$\gamma = 1.25 \times 10^{-6} k^{3/4},$$

which appears to hold for all mobilities between that of the large and that of the small ion. The measured value of γ for $k = 3.2 \times 10^{-4}$ was 3.1×10^{-9}. In the free atmosphere away from sources of large ions the terms Q, γN_1N_2, and ζN_1 are usually so small they can be neglected, and (7) reduces to

$$dN_1/dt = \eta_{10}n_1N_0 - \eta_{12}n_2N_1,$$

and for large-ion equilibrium conditions,

$$\eta_{10}n_1N_0 = \eta_{12}n_2N_1,$$

which is equation (3). Equation (4) represents the analogous equation in the case of the negative large ions.

Ionic Concentrations in the Lower Atmosphere and Their Variations

Both large- and small-ion concentrations vary considerably from place to place and from time to time at a given place. Over land, the two usually vary systematically but in opposite directions, both during the day and throughout the year.

The large-ion content (in ions per cc) of the air over the oceans averages only a few hundred of each sign. More extensive measurements have been made of the condensation-nuclei content of the air over the oceans from which estimates may be made of the large-ion content. Landsberg [20], making use of all available measurements at the time, estimated that the average nuclei content of the air over the oceans was 940. Hess [11] more recently found an average over the Atlantic, between America and Europe, of 527 on one leg of a cruise and 659 on another leg. On Cruise VII of the *Carnegie* [44], the average nuclei concentration was 1370 over the Atlantic and 2350 over the Pacific. From all of these results one would estimate between 100 and 200 large ions of each sign for the air over the Atlantic and two to three times this number over the Pacific. The large-ion content over the oceans, as estimated on the basis of the number of small ions found during Cruise VII, is around 180 of each sign [44], which is in reasonably good agreement with the estimates above. No diurnal variation was found in the condensation-nuclei content of the air over the oceans during Cruise VII, from which one might surmise that the large-ion content would likewise be constant through the day.

Over land, Landsberg [20] estimated that the average condensation-nuclei content of the air in country districts was around 10,000 per cc, in towns around 30,000, and in cities around 150,000. The corresponding large-ion content might accordingly be estimated at between 1000 and 2000 for a country area, between 5000 and 6000 for a town, and between 20,000 and 30,000 for a city. These must be regarded as rough estimates only.

Most observations on the condensation and large-ion content of the air over land have been taken in the vicinity of human habitations. Both the annual and diurnal variations as well as the absolute values are greatly affected by human activities. When observed sufficiently far from industrial activities of man, the large-ion content, like the condensation-nuclei content of the air, shows a maximum during the winter when heating of homes is greatest, and a minimum in the summer. For a similar reason, the condensation-nuclei content of the air generally shows a maximum during the middle of the day [15, 20, 26, 46, 56]. The diurnal variation in large-ion concentration obtained by Yunker [56], on the other hand, was more or less opposite to this. The curve for large ions found at Washington [49] is likewise quite opposite and similar to Yunker's curve, both being plotted on local time. Torreson [43] first pointed out the discrepancy between the condensation-nuclei curves and the large-ion curves at Washington, and Wright [54] suggested an explanation based upon a variation in size of the condensation nuclei with relative humidity. Sherman [40] could find no evidence of a diurnal variation in the ratio of uncharged to charged nuclei in the air, as required by Wright's hypothesis. This apparent paradox has not yet been explained, but it raises the question whether condensation-nuclei counts tend to include relatively large particles while the large-ion counts include smaller particles. Additional experiments will be required to find an answer to the question as to why the two curves are of such different character and to ascertain if a similar difference is to be found at other places.

Over the oceans, even though the rate of small-ion production is less than that over land (only 10 to 15 per cent as great), the small-ion concentration in the two areas usually differs but little. This is accounted for by a smaller number of large ions over the oceans and consequently a slower destruction rate of the small ions. Over the oceans, the average small-ion content of the air was about 500 and 400, respectively, for the positive and negative ions during Cruise VII of the *Carnegie* [44]. Over land the concentration of each sign varies from around 100 over polluted areas to about 1000 for areas unpolluted with industrial smokes and gases.

The small-ion content of the lower atmosphere is generally lowest during the winter months and highest during the summer, just opposite to that found for the large-ion content. The values usually reach a maximum during the morning hours and a minimum during the afternoon. This type of variation, while largely controlled by the variation in large-ion content of the air, also depends to some extent upon the variations in the rate at which ions are produced, which, according to the curves of Fig. 1, is greatest during the early morning hours.

Variation in Ion Concentration of the Atmosphere with Height above Ground

The value of the large-ion concentration of the atmosphere at various heights above the ground can only be inferred from the values of condensation nuclei obtained at various altitudes. Wigand [20] made fourteen balloon flights on which condensation nuclei were measured. A rapid decrease in concentration with altitude was found; at a height of 3 km the concentration was only 3 per cent of that at ground level.

Much more information is available concerning the value of the ratio n_1/n_2, how it varies and why, than there is concerning the ratio N_1/N_2. Moussiegt [25], from observations in the Alps, found mean values of 2214 and 2335, respectively, for N_1 and N_2, from which one would deduce a value of 0.95 as the ratio of N_1/N_2. Hogg [15], in Canberra, Australia, found a mean value for this ratio of 1.22. In Washington, D. C., continuous records (unpublished results) of large, small, and intermediate ion concentration of air, alternating each week between the positive and negative ions, were made during September and December, 1935. Since these data are not strictly simultaneous, the ratios can be regarded as only approximate. In view of the paucity of published data from which a comparison of positive- and negative-ion content can be made, a summary of these data seems worth while and is given in Table IV. The large-ion concentrations in ions per 10^{-2} cc are represented by N_1 and N_2, while M_1, M_2, and n_1, n_2 represent the positive and negative intermediate-ion and the positive and negative small-ion concentrations, respectively, in ions per cc. The ratios M_1/M_2 and n_1/n_2

TABLE IV. MEANS OF VARIOUS HOURLY VALUES OF VARIOUS ELEMENTS IN WASHINGTON

Means of	Elements									Month (1935)
	N_1	N_2	M_1	M_2	n_1	n_2	N_1/N_2	M_1/M_2	n_1/n_2	
5 largest	77.2	72.4	121	126	284	266	1.49	1.05	1.19	
5 smallest	35.8	27.3	94	91	180	145	1.00	0.95	1.01	September
All values	51.1	42.8	107	108	229	212	1.10	1.00	1.10	
5 largest	79.3	74.0	167	195	169	159	1.27	1.01	1.22	
5 smallest	46.8	38.5	139	139	128	108	0.98	0.90	1.01	December
All values	60.8	48.8	153	170	188	173	1.12	0.96	1.12	

The values of the small-ion concentration made on thirteen balloon flights by various observers and that deduced from *Explorer II* air-conductivity measurements have been summarized by Gish [6, p. 194]. He concludes that the concentration increases roughly by 1000 ions per cc (one sign) for each 2-km increase in altitude. This rate of increase is true up to an altitude of about 6 km, after which it gradually diminishes, according to the *Explorer II* data, to less than half this value at 14-km altitude.

Ratio of Positive-Ion to Negative-Ion Concentration

A relationship between the ratios N_1/N_2 and n_1/n_2 has been deduced by Gish [6, p. 183] from the equilibrium equations and is given by the equation

$$(n_1/n_2)^2 = bN_1/aN_2,$$

where $a = \eta_{10}/\eta_{21}$ and $b = \eta_{20}/\eta_{12}$. If one assumes that $\eta_{10}/\eta_{20} = \eta_{12}/\eta_{21}$, then from the equation above it follows that

$$(n_1/n_2)^2 = (\eta_{20}/\eta_{10})^2 N_1/N_2,$$

which is the relationship assumed by Nolan and de Sachy [26]

show no consistent diurnal variation, while the ratio N_1/N_2 varies more or less in a manner opposite to that of the large-ion content [49].

During fair weather the average concentration of positive small ions exceeds that of the negative small ions by 10 or 20 per cent. Particularly at certain stations, the ratio n_1/n_2 varies in a manner similar to the variation of the earth's field (electrode effect). This is due to the repulsion of negative ions by the negatively charged ground. During a thunderstorm, because of the intense electric field with frequent changes in sign, the ratio undergoes frequent changes in value, say from a very high to a very low value and vice versa, usually several times during a storm.

Mean Life of an Ion

The mean life of an ion is the time interval between formation and destruction of the average ion. For small ions the mean life τ_n in seconds, for the case when large-ion concentration is sufficiently small, is given by the equation

$$\tau_n = n/q, \qquad (8)$$

in which n represents the small-ion concentration and q, the rate of ionization (rate of small-ion formation).

When the large ions are numerous, then in accordance with the assumptions made in deriving (5),

$$\tau_n = \frac{1}{2\eta_{12}N_2} . \qquad (9)$$

Over the oceans the value of τ_n will generally be around 5 or 6 min. Over land, in areas where the large ions are few, the mean life τ_n may be expected to be only 10 or 20 per cent of the above. In localities where large ions are numerous, the value of τ_n will be still smaller, probably only 5 per cent or so of the ocean value.

The mean life τ_N of the large ion is given approximately by the equation:

$$\tau_N = \frac{1}{\eta_{12}n_1} .$$

Over the oceans and over land where the large ions are not very numerous, the value of τ_N may be between 15 and 20 min. In areas where the large ions are numerous, the value of τ_N may easily exceed an hour.

Because of the relatively short life of the small ion, the small-ion concentration may be expected to follow with little lag those factors which tend to produce a change in the concentration. Much greater lag may be expected in the changes of large-ion concentrations.

Outstanding Problems in the Field of Atmospheric Ions

A number of investigations have been carried out for the purpose of evaluating the factors which control or regulate the small-ion content of the lower atmosphere. There is growing evidence that some of the factors, for example the value of the combination coefficients between the small ions and the charged and uncharged condensation nuclei, may vary from place to place. This is probably due to a difference in character of the nuclei which results in a difference in nuclei size. A close correlation would therefore probably be found between the values of the combination coefficients and the mobility of the large ion. In future work, a recognition of this possibility may assist in harmonizing results which otherwise might appear to be inconsistent.

Another small-ion regulating factor which is difficult to evaluate is q, the rate of ionization of the atmosphere. Probably the most promising method of evaluating this factor is through the use of a thin-walled ionization chamber. This cannot be accomplished, however, without certain difficulties. Cognizance must be taken of the low radioactive content of the air within the chamber compared to that of the outside air and of the fact that the ionization within the chamber will depend upon the wall thickness and upon the particular voltage applied to the central electrode. The latter arises from the fact that complete saturation is never achieved, due probably to columnar ionization inside the chamber. A full discussion of this problem is not possible within the limits of this article.

It is quite probable that meteorological conditions play an important role in altering the efficiency of some or all of the small-ion regulating factors. Temperature, humidity, and pressure of the atmosphere, for example, are likely to exert an influence on such factors as combination coefficients, recombination coefficients, and the mobility of ions. There is urgent need for careful experiments designed to secure much-needed information along such lines.

Multiply charged large ions should be examined as to regularity of, and conditions of, occurrence and as to their effect on large-ion mobility and on establishment of small-ion equilibrium conditions.

REFERENCES

1. Aitken, J., *Collected Scientific Papers of John Aitken*, edited by C. G. Knott. Cambridge, University Press, 1923.
2. Boylan, R. K., "Atmospheric Dust and Condensation Nuclei." *Proc. R. Irish Acad.*, (A) 37: 58–70 (1926).
3. Bradbury, N. E., "The Absolute Values of the Mobility of Gaseous Ions in Pure Gases." *Phys. Rev.*, 40: 508–523 (1932).
4. Cullen, T. L., "On the Exhalation of Radon from the Earth." *Terr. Magn. atmos. Elect.*, 51: 37–44 (1946).
5. Erikson, H. A., "Mobility of Ions." *Phys. Rev.*, 17: 400 (1921); 18: 100–101 (1921); 20: 117–126 (1922); 24: 502–509 (1924); 26: 465–468 (1925); 28: 372–377 (1926); 29: 215–216 (1927); 33: 403–411 (1929); 34: 635–643 (1929).
6. Fleming, J. A., ed., *Physics of the Earth—VIII, Terrestrial Magnetism and Electricity*. New York, McGraw, 1939.
7. Gagge, A. P., and Moriyama, I. M., "The Annual and Diurnal Variation of Ions in an Urban Community." *Terr. Magn. atmos. Elect.*, 40: 295–306 (1935).
8. Gish, O. H., and Sherman, K. L., "Electrical Conductivity of the Air to an Altitude of 22 Kilometers." *Nat. Geogr. Soc. Contrib. Tech. Papers, Stratosphere Series*, No. 2, pp. 94–116 (1936).
9. —— *Ionic Equilibrium in the Troposphere and Lower Stratosphere*. Trans. Wash. Meeting 1939, Un. géod. géophys. int., Ass. Terr. Magn. Atmos. Elect., Bull. 11, pp. 474–491, 1940.
10. Hess, V. F., *The Electrical Conductivity of the Atmosphere and Its Causes*. Trans. by L. W. Codd. London, Constable, 1928.
11. —— "On the Concentration of Condensation Nuclei in the Air over the North Atlantic." *Terr. Magn. atmos. Elect.*, 53: 399–403 (1948).
12. Hogg, A. R., "Some Observations on the Average Life of Small Ions and Atmospheric Ionization Equilibria." *Beitr. Geophys.*, 41: 32–55 (1934).
13. —— "Continuous Observations of the Rate of Production of Small Ions in the Atmosphere." *Beitr. Geophys.*, 43: 359–378 (1935).
14. —— "The Mobility of Small Ions of the Atmosphere." *Beitr. Geophys.*, 47: 31–59 (1936).
15. —— "Atmospheric Electric Observations." *Beitr. Geophys.*, 41: 1–31 (1934).
16. —— "The Intermediate Ions of the Atmosphere." *Proc. phys. Soc. London.*, 51: 1014–1027 (1939).
17. Israël, H., and Schulz, L., "The Mobility-Spectrum of Atmospheric Ions—Principles of Measurements and Results." *Terr. Magn. atmos. Elect.*, 38: 285–300 (1933).
18. Jaffé, G., "Zur Theorie der Ionisation in Kolonnen." *Ann. Physik*, 42: 303–344 (1913); "Sur l'ionisation des diélectriques liquides par l'emanation du radium." *Radium, Paris.*, 10: 126–134 (1913); "Zur Theorie der Ionisation in Kolonnen, II." *Ann. Physik.*, 1: 977–1008 (1929).
19. Kennedy, H., "The Large Ions in the Atmosphere." *Proc. R. Irish Acad.*, (A) 32: 1–6 (1913).

20. Landsberg, H., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, Supp., 3: 155–252 (1938).
21. Langevin, P. "Sur les ions de l'atmosphère." *C. R. Acad. Sci., Paris*, 140: 232–234 (1905).
22. Loeb, L. B., *Fundamental Processes of Electrical Discharges in Gases*. New York, Wiley, 1939. (See pp. 86–100)
23. Luhr, O., and Bradbury, N. E., "Corrected Values for the Coefficient of Recombination of Gaseous Ions." *Phys. Rev.*, 37: 998–1000 (1931).
24. Mauchly, S. J., "Studies in Atmospheric Electricity Based on Observations Made on the *Carnegie*, 1915–1921." Res. Dept. Terr. Magn., *Carneg. Instn. Wash. Publ.* 175, 5: 385–424 (1926).
25. Moussiegt, J., "Mesures de la conductibilité et de l'ionisation de l'air dans les Alpes." *C. R. Acad. Sci., Paris*, 208: 216–217 (1939).
26. Nolan, J. J., and de Sachy, G. P., "Atmospheric Ionisation." *Proc. R. Irish Acad.*, (A) 37: 71–94 (1927).
27. Nolan, J. J., and Nolan, P. J., "Further Observations on Atmospheric Ionisation at Glencree." *Proc. R. Irish Acad.*, (A) 41: 111–128 (1933).
28. —— "A New Method for Counting Atmospheric Ions and Determining Their Mobilities." *Proc. R. Irish Acad.*, (A) 42: 15–19 (1935).
29. —— "Atmospheric Electrical Conductivity and the Current from Air to Earth." *Proc. R. Irish Acad.*, (A) 43: 79–93 (1937).
30. Nolan, P. J., "Experiments on Condensation Nuclei." *Proc. R. Irish Acad.*, (A) 47: 25–38 (1941).
31. —— "The Recombination Law for Weak Ionisation." *Proc. R. Irish Acad.*, (A) 49: 67–90 (1943).
32. —— and Galt, R. I., "The Equilibrium of Small Ions and Nuclei." *Proc. R. Irish Acad.*, (A) 50: 51–68 (1944).
33. Nolan, P. J., and Fahy, E. F., "Experiments on the Conductivity of Atmospheric Air." *Proc. R. Irish Acad.*, (A) 50: 233–256 (1945).
34. Overhauser, A. W., "The Clustering of Ions and the Mobilities in Gaseous Mixtures." *Phys. Rev.*, 76: 250–254 (1949).
35. Parkinson, W. D., "Factors Controlling the Atmospheric Conductivity at the Huancayo Magnetic Observatory." *Terr. Magn. atmos. Elect.*, 53: 305–317 (1948).
36. —— *Review of the Literature on the Ion Balance of the Lower Atmosphere Since 1939*. Trans. Oslo Meeting 1949, Un. géod. géophys. int., Ass. Terr. Magn. Atmos. Elect., Bull. 13 (in preparation).
37. Pollock, J. A., "A New Type of Ion in the Air." *Phil. Mag.*, 29: 636–646 (1915).
38. Sayers, J., "Ionic Recombination in Air." *Proc. roy. Soc.*, (A) 169: 83–101 (1938).
39. Sherman, K. L., "Atmospheric Electricity at College-Fairbanks Polar Year Station." *Terr. Magn. atmos. Elect.*, 42: 371–390 (1937).
40. —— "Total and Uncharged Nuclei at Washington, D. C." *Terr. Magn. atmos. Elect.*, 45: 191–204 (1940).
41. Swann, W. F. G., "On Certain New Atmospheric-Electric Instruments and Methods." *Terr. Magn. atmos. Elect.*, 19: 171–185 (1914).
42. Thomson, J. J., *Conduction of Electricity Through Gases*, 3rd ed. Cambridge, University Press, 1933. (See p. 7)
43. Torreson, O. W., "On the Value of the Ratio of the Number of Uncharged Nuclei (N_0) to the Number of Charged Nuclei of one Sign ($N\pm$) at Washington, D. C." *Terr. Magn. atmos. Elect.*, 39: 65–68 (1934).
44. —— and others, "Scientific Results of Cruise VII of the *Carnegie*. Oceanography III, Ocean Atmospheric Electric Results." *Carneg. Instn. Wash. Publ.* 568 (1946).
45. Torreson, O. W., and Wait, G. R., "Measurements of Total Nuclei, of Uncharged Nuclei, and of Large Ions in the Free Atmosphere at Washington, D. C." *Terr. Magn. atmos. Elect.*, 39: 47–64 (1934).
46. Wait, G. R., "Diurnal Variation of Concentration of Condensation-Nuclei and of Certain Atmospheric-Electric Elements at Washington, D. C." *Terr. Magn. atmos. Elect.*, 36: 111–131 (1931).
47. —— "The Intermediate Ion of the Atmosphere." *Phys. Rev.*, 48: 383 (1935).
48. —— "Some Experiments Relating to the Electrical Conductivity of the Lower Atmosphere." *J. Wash. Acad. Sci.*, 36: 321–343 (1946).
49. —— and Torreson, O. W., "The Large-Ion and Small-Ion Content of the Atmosphere at Washington, D. C." *Terr. Magn. atmos. Elect.*, 39: 111–119 (1934).
50. —— "Diurnal Variation of Intermediate and Large Ions of the Atmosphere at Washington, D. C.," *Terr. Magn. atmos. Elect.*, 40: 425–431 (1935).
51. —— "Recombination of Ions from Gas Flames." *Phys. Rev.*, 57: 1071 (1940).
52. Weiss, R., und Steinmaurer, R., "Messungen der Luftionen in Innsbruck." *Beitr. Geophys.*, 50: 238–251 (1937).
53. Whipple, F. J. W., "Relations Between the Combination Coefficients of Atmospheric Ions." *Proc. phys. Soc. Lond.*, 45: 367–380 (1933).
54. Wright, H. L., "The Association Between Relative Humidity and the Ratio of the Number of Large Ions to the Total Number of Nuclei." *Terr. Magn. atmos. Elect.* 39: 277–280 (1934).
55. —— "Sea-Salt Nuclei." *Quart. J. R. meteor. Soc.*, 66: 3–12 (1940).
56. Yunker, E. A., "The Mobility Spectrum of Atmospheric Ions." *Terr. Magn. atmos. Elect.*, 45: 127–132 (1940).
57. Zeilinger, P. R., "Ueber die Nachlieferung von Radiumemanation aus dem Erdboden." *Terr. Magn. atmos. Elect.*, 40: 281–294 (1935).
58. Zupancic, P. R., "Messungen der Exhalation von Radiumemanation aus dem Erdboden." *Terr. Magn. atmos. Elect.*, 39: 33–46 (1934).

PRECIPITATION ELECTRICITY

By ROSS GUNN

Physical Research Division, U. S. Weather Bureau

Introduction

Although atmospheric electricity does not play an important part in the control of world-wide weather phenomena, it does have considerable bearing on special weather problems. Because of the basic electrical nature of all matter, most mechanical and thermodynamical energy transformations are accompanied by some type of electrical phenomenon. It is not surprising, therefore, that the production of precipitation in the atmosphere frequently gives rise to interesting and important electrical effects. These phenomena share many of the peculiarities of weather because of the extreme complexity of charge production and transport processes in the atmosphere. One serious difficulty in the proper evaluation of precipitation electric phenomena is that the data thus far collected have been limited to a relatively few geographical regions and with few exceptions have covered only short periods of time. Many of the data are contradictory.

In view of the complexity of the electrical phenomena accompanying precipitation, it might appear that the most rapid progress on basic problems would be made if laboratory investigations were undertaken. However, it is apparent that this would be extraordinarily difficult for the same reason that investigations of weather in the laboratory have met with considerable difficulties. Electrification of the atmosphere and of precipitation is related to the characteristics and development of cloud structures, and these have not yet been successfully reproduced under controlled conditions. The scientist is forced, therefore, to collect data wherever and whenever available and to attempt to deduce from them the main characteristics and important processes of nature.

The principal practical problems of precipitation electricity that require intensive work are (1) to describe the detailed electrical processes responsible for the production of lightning, (2) to describe the mechanisms responsible for the maintenance of the observed negative free charge on the surface of the earth, and (3) to describe those processes that transfer free electrical charge to aircraft flying through natural precipitation and to devise a method for counteracting such processes. Solution of these practical problems requires a detailed understanding of other still more basic questions: (1) How is a free electrical charge placed on precipitation? (2) Why does charge of a selected sign appear principally upon the larger precipitation elements? (3) What are the mechanisms responsible for the separation of positive and negative charges? and (4) How large are the electric fields so produced? Quantitative understanding of these fundamental problems will provide a suitable foundation for the solution of the more practical problems.

Earth electrification processes that become manifest through easily obtainable measurements are intimately related to a dual basic process consisting first of the deposition of free charge on precipitation particles and then of the subsequent mechanical separation of charges having opposite signs. Mechanisms responsible for placing a free electrical charge of selected sign on the precipitation particles are not well understood, but they are basically related to atomic forces that are both physical and chemical in nature.

The occurrence of electrification and of coexisting available electrical energy implies the expenditure of mechanical work to establish the electrified state. In the earth's atmosphere this systematic mechanical work comes principally from gravitational forces and accelerations due to turbulent atmospheric motion, acting on precipitation particles. Large-scale separation of electrical charges will occur only when the acting forces operate on particles of one sign in a way quite different from the way they operate on particles carrying the opposite electrical charge. Because precipitation particles normally fall in the earth's atmosphere, the observed separation of charges always implies that the aerodynamic characteristics of carriers of the positive charge are notably different from the characteristics of carriers of the negative charge. This aerodynamic contrast between the particles carrying opposite charges is of importance in the description of all large-scale atmospheric electrifications.

Observed Free Electrical Charge on Precipitation

Early investigators were of the opinion that the free electrical charge carried to the earth by rain was adequate to replenish the normal discharge current observed in fair-weather areas throughout the earth. Later work has recognized that other processes are also important, but the original ideas stimulated the first measurements. The literature on the subject is contradictory in many cases, and the actual values of free electrical charge carried down by precipitation particles differ so much from place to place and with different meteorological situations that average values are of questionable significance. Better agreement between various observers is secured if the electrical characteristics of precipitation are classified in accordance with three distinct types of rainfall: (1) continuous or quiet rain (*Landregen*), (2) shower or squall rain (*Böenregen*), and (3) electrical storm rain (*Gewitter*). This useful classification was adopted by Gschwend in his measurements

of charge and mass of individual falling raindrops reported in 1920 [11].

Continuous or Quiet Rain. Quiet rain, usually associated with smaller droplets, carries relatively smaller charges *per drop* than storm rain and is more likely to be predominantly of one sign. The majority of experimenters have found that quiet rain is usually positive, although examples of continuous negative rain are known. The ratio of positive to negative charges reported generally varied from 1.1 to 1.5 when adequate samples were taken. In spite of the excess observed positive charge, it was found that the negative charge *per droplet* in most cases was greater than the positive charge per droplet. This implies, and measurements show, that usually a larger number of positively charged droplets fall. A dependence upon the rate of rainfall has usually been observed. A long series of measurements by Chalmers and Pasquill [4] in England showed that the number of particles carrying positive charges was some 70 per cent more than the number carrying negative charges, and that the total positive free charge delivered at the earth was some 30 per cent greater than the negative. It is interesting to note in this connection that, if data are collected over a sufficiently long period, the net free transported charge may be a small fraction of the total. For example, Scrase [23] measured the convected charge continuously for a two-year period. His results showed that positive charges predominated one year and negative the other, but for the entire interval the positive charge exceeded the negative by 10 per cent.

Shower or Squall Rain. Squall-type showers share the electrical characteristics both of quiet rain and of rain falling in typical electrical storms. Wide excursions in the electrical characteristics are normally observed. The analysis of rain falling from postthunderstorm showers and from those that have not quite proceeded to the point of active charge separation will be of the utmost value in determining the basic precipitation charging processes.

Electrical Storm Rain. The rain falling in a typical electrical storm is usually characterized by large droplets and large free charges that approximate a few hundredths of an electrostatic unit per drop. Gschwend [11] pointed out the remarkable fact that such rain frequently changes its sign. It has been found in active storms that after a very few consecutive droplets of one sign had been measured the chance of capturing a droplet having an opposite sign was very large. The free charge brought down by individual droplets falling from active thunderclouds has been measured on the ground by Gschwend [11], Banerji and Lele [3], and Gunn [12]. Gschwend found that the largest charges were associated with positive droplets, while the other experimenters found the largest charges associated with negative ones. In very active electrical storms, Gunn found a general trend connecting the charge on a droplet and its radius. The electrification of these droplets increased on the average until an electric field that approximated 2.5 esu cm^{-1} was established on the surface of the droplets. Negatively charged droplets were nearly twice as massive as the positively charged ones, and each carried about 25 per cent more charge.

The sign and magnitude of the integrated free charge transported to the earth by precipitation in electrical storms is uncertain. A number of measurements in thunderstorms have shown that negative charge is usually transferred to the earth by the acting mechanisms. For example, Banerji [1] has estimated that the excess negative charge convected to the ground by rain falling from an active thunderstorm in India was 2×10^3 coulombs.

The electrical state in an active storm is so complicated and confused that there is doubt as to the value of precipitation data as a guide to the interpretation of basic electrical processes. This uncertainty in interpretation has become serious as a result of a recent paper by Simpson [25]. He has presented evidence suggesting that the charge on falling rain is deposited by conduction or corona currents discharged near the surface of the earth by the electric fields usually present. Although a number of earlier experimenters looked in vain for such a correlation, Simpson now reports a good correlation between the sign of the free charge and the direction of the electric field. A careful check of the facts in this matter by independent observers is badly needed.[1]

Measurements taken in an aircraft at various altitudes up to 26,000 ft have been reported by Gunn [14]. These measurements were made in regions far above surface corona discharges and may be the only ones that give a clear-cut indication of the charge-producing mechanisms in the earth's atmosphere. In a weak cold front exhibiting no thunderstorm activity, positive charges averaging 0.033 esu per drop were observed from 10,000 to 26,000 ft. Negative charges averaging 0.040 esu per drop were measured between 4000 and 20,000 ft. Positive particles were not observed below 10,000 ft and negative ones were not detected above 20,000 ft. The freezing level during these measurements was at 11,000 ft. Direct measurement in the vicinity of the plane showed that the electric field did not exceed 25 v cm^{-1}, and therefore thunderstorm activity was negligible. A coherent set of similar data [15] taken in an active thunderstorm gave notably greater free charges on the precipitation and suggests that the interpretation of such collected data will be extraordinarily difficult.

Snow. Gschwend made a number of measurements of the charge carried by individual snowflakes. Positive charges, in general, exceeded negative charges, and it is important to note that the charge on newly formed snow is nearly one hundred times larger than the charge on quietly falling, and presumably aged, snowflakes.

Nakaya and Terada [20] found that snow carried a preponderately negative charge unless the flakes had frozen water droplets attached. Recently, Pearce and Currie [22] remarked on the large number of essen-

1. (Note added in proof.) W. C. A. Hutchinson and J. A. Chalmers have just published a paper, "The Electrical Charges and Masses of Single Raindrops," *Quart. J. R. meteor. Soc.*, 77:85–95 (1951), that provides needed data on this subject.

tially neutral snowflakes. However, of the flakes measured, they found free positive charges on twice as many flakes as carried negative charges. They also found that snow drifting over the ground was strongly negative. Chalmers and Pasquill [4] reported that snow is predominantly negative. It seems apparent from the literature and from other measurements that the sign and magnitude of the charge on falling snow depends critically upon its crystalline structure and mode of formation. As an illustration of this fact, this author found that snowflakes falling quietly with an average velocity of 48 cm sec^{-1} carried average positive charges of 0.00067 esu, whereas simultaneously falling negative flakes of average charge 0.0010 esu fell with a velocity of 80 cm sec^{-1}. The difference in sign was definitely correlated with the rate of fall. It is probable that the rate of fall is determined by the structure and density of the flake, which, in turn, is determined by its mode of formation. It is a fair inference from the data, therefore, that the opposite electrical charges result from grossly different developmental histories.

Average values of free charge carried by both positive and negative individual droplets of various kinds of precipitation, as measured by Gschwend [11], Banerji and Lele [3], Chalmers and Pasquill [4], and Gunn [12, 14, 15] are summarized in Table I.

Table I. Average Free Electrical Charge on Individual Droplets (esu $\times 10^3$)

Observer	Altitude (ft)	Charge	Quiet rain	Shower rain	Electrical storm rain	Quiet snowfall	Squall snowfall
Gschwend (1921)	surface	+	0.24	1.75	8.11	0.09	5.64
		−	0.53	5.43	5.88	0.06	4.78
Banerji and Lele (1932)	surface	+		6.4	6.9		
		−		6.7	7.3		
Chalmers and Pasquill (1938)	surface	+	2.2	1.3	3.7*		10.5
		−	3.0	2.3	9.2*		5.7
Gunn (1947)	4,000	+		—			
		−		24			
	12,000	+		41			
		−		100			
	20,000	+		63			
		−		—			
Gunn (1949)	surface	+			15	0.67	
		−			19	1.0	
Gunn (1950)	5,000	+			81		
		−			63		
	10,000	+			148		
		−			112		
	15,000	+			123		
		−			76		
	20,000	+			52		
		−			62		

* Actual lightning activity doubtful.

Cloud Elements. If raindrops are formed by the association of cloud elements, it is obvious that the free electrical charge collected on cloud particles is of fundamental importance. A number of measurements have been made on the charges carried by clouds and fog, notably by Wigand [27]. He found that in a dry fog the cloud elements carried a positive electrical charge, but sometimes negative charges were measured. The magnitude of the charge varied from a few to a few hundred elementary charges. Using a specially instrumented aircraft [18], the author made a number of measurements of the charges on cloud particles in small swelling cumulus clouds and found that the charges were usually negative. The average charge on each element was estimated to approximate 22 elementary charges. Scrase [24] found that cloud elements in heavy wet fogs frequently carried a negative charge approximating 35 elementary units. Accumulation of information on the electrical charges carried by cloud droplets under various meteorological conditions is urgently needed.

Processes Responsible for the Electrification of Precipitation

Because all large-scale atmospheric electrifications derive their energy originally from expenditure of mechanical or gravitational work and because this work can be converted only when a free electrical charge is attached to some physical entity like a raindrop, it is of utmost importance to understand in detail the physical processes whereby free charge, of either sign, can be systematically deposited on droplets. One of the outstanding characteristics of precipitation elements in the atmosphere is the enormous surface area exposed to the atmospheric ions and to the chemical activity of the air. It has been noted frequently that precipitation elements in the air share many of the remarkable properties of a colloidal suspension.

Droplet charging processes may be divided into two major categories: first, charging processes which are of a basic nature and dependent upon the physical and chemical properties of water and air; and second, charging processes which are critically dependent upon special environmental conditions.

Basic Processes. As a common example of electrification by a basic process, one may mention the separation of electricity produced by friction. The rubbing together of materials having contrasting physical properties usually results in the selective transfer of electrons in the outer orbits from one material to the other. Dry ice crystals sliding along the metallic wing of an aircraft communicate large amounts of negative electricity to the aircraft and positive electricity to the ice crystal. It is well known that snow blowing along the ground acquires a strong negative charge which is very likely of similar frictional origin.

One of the important basic processes that produce electrical effects in the atmosphere results from the chemical adsorption of ions at the surfaces of precipitation particles. Systematic polarization and orientation of surface molecules frequently result. This orientation produces electrical double layers that are responsible for electrophoresis and a number of allied surface phenomena [9]. In pure water the polarization of the surface molecules is such that the outer surface is made up of negative charges, while some 10^{-8} cm below this negative surface a positive distribution of

exactly the same amount of electricity exists. Inside this double layer, a more or less random distribution of both positive and negative free charges is re-established. The net result of the double layer on the surface of a spherical drop is that the potential inside the drop may be greater than that outside by a fraction of a volt. This does *not* mean that free electrification is produced, since it must be remembered that exactly equal amounts of positive and negative electricity are encompassed by the double layer. However, if the double layer is subsequently broken and mechanical work is done on it or if ions are selectively captured, measurable amounts of free electrification may result.

One familiar charge-producing mechanism, intimately related to this double-layer process, is the so-called waterfall effect first studied by Lenard [19]. Subsequent experimentation has shown that when a droplet of pure water is broken up by mechanical means, the residual droplets carry positive charges while the adjacent air acquires both positive and negative ions. Electrifications produced by breakup, atomization, splashing, or bubbling are all intimately related to the double-layer characteristics; this subject has received much study [2, 5]. Although one widely quoted theory invokes such mechanisms to describe thunderstorm electricity, quantitative agreement with observation is quite unsatisfactory.

One aspect of double-layer electrification may be of importance in the atmosphere when hail is produced. Dinger and Gunn [6] discovered that when water freezes in the atmosphere a large amount of air is entrapped in the ice in the form of tiny bubbles. Upon melting, these bubbles are released, and upon breaking the surface they transfer to the adjacent air a negative charge of 1.25 esu gram^{-1} while the melted water droplet retains an equal and opposite positive charge. The free charge thus produced is appropriately distributed near the freezing level and is sufficiently large to explain the presence of active cloud electrification. These experimenters also discovered that the freezing of relatively pure water is accompanied by transient changes in the contact electromotive force amounting to 6 or 10 v. Because the charge distribution at the surface, responsible for the contact electromotive force, is in the nature of a double layer, they did not believe it contributed to a net electrification of a freezing droplet. Using special dilute solutions, Workman and Reynolds [29] have reexamined this latter process and report that potential differences exceeding even 100 v are produced under special circumstances.

It is important to note that the magnitudes of electrical effects in all double-layer phenomena depend *critically* upon the purity of the water. Banerji [2] has remarked that impurities commonly existing in precipitation in the atmosphere are usually sufficient to reduce the expected electrical effects to small values.

Environmental Processes. It is an observed fact that the atmosphere is pervaded by relatively large numbers of ions of both signs. The small negative ions move 10–40 per cent faster than the positive ions when acted on by the same force. Therefore, the negative ions usually determine the charge captured by an initially uncharged and insulated body.

It has long been known that an insulated conductor supported in an ion stream becomes charged due to the selective capture of ions. Pauthenier and Moreau-Hanot [21] have formulated this process in a quantitative theory that appears to agree well with experiment.

Wilson [28] has also pointed out that a droplet falling in an electric field is polarized, and as it falls it selectively captures, because of its motion, the more mobile ions in the volume it sweeps out. Experimental results confirm the reality of this process [10]. Since this important charge-separating effect depends upon the existence of an initial electric field, its application to the description of the electrical properties of the atmosphere is obscure. The mechanism is useful in describing changes in the electrical state subsequent to the establishment of an electric field by some more fundamental mechanism.

Electrification occurring when rime is deposited on a conducting surface or on graupel is considered important by Findeisen [8], who has based a theory upon it. The effect is real, but its quantitative relation to thunderstorms has not yet been completely worked out. The application of this mechanism is attractive because it correlates the observed high electrification occurring near the freezing level with theory.

An environmental process first investigated by Gunn [13] depends on the differential migration of atmospheric positive and negative ions under the influence of a systematic transfer of water molecules. He showed that the transfer of water vapor towards a condensing droplet results in a transfer of momentum to both the positive and negative ions in the vicinity and the establishment of a greater concentration of the most mobile ion adjacent to and upon the condensing droplet. The charge capable of being transferred to such a droplet is related to the vapor stream velocity and to the thermal kinetic energy of the molecules, and hence in the atmosphere is something less than 0.1 v. It should be clear that reversing the direction of the water-vapor stream will reverse the sign of the charge on the evaporating or condensing droplet.

Association Processes. In attempting to understand the basic mechanisms responsible for the surprisingly large electrical charge sometimes carried by precipitation, one process should be emphasized. Without discussing the details of association, it seems certain that rain produced below the freezing level results from the association of an extremely large number of cloud particles. In typical cases, the number of cloud particles associated to produce a single raindrop is surprisingly large, and if any process systematically transfers even small charges of a given sign to the cloud particles, then the total charge may be large. Gunn [13] worked out a complete "association theory" of electrical storm activity based on this idea. He remarked that a number of physical and chemical forces could be expected to transfer small charges to the cloud particles. To illustrate the theory, he adopted the notion that each cloud particle was an electrical concentration cell and that

the mean potential between the droplet and the outside air approached 60 mv. By estimating the total charge resulting from the association of typical cloud elements, relatively large free charges per droplet were calculated. It was found that precipitation electrical phenomena could be well described, both qualitatively and quantitatively, by such an hypothesis. The theory is not considered complete but it does serve to emphasize the probable importance of association mechanisms in the production of highly charged precipitation.

Separation of Free Electrical Charges and the Importance of Droplet Size

Although all matter is composed of an enormous number of electrical charges, it is a general rule that every small volume of space contains as many positive charges as negative charges. A short calculation will show that this indeed must be the case, because any systematic separation of charges of opposite sign immediately sets up surprisingly large electrical forces that always act in such a direction as to restore electrical neutrality.

An important exception to the general rule of neutrality occurs in the earth's atmosphere when free charges of one sign become selectively attached to the larger or smaller precipitation elements. As an example, suppose that, for some reason, all of the elementary cloud particles in a given volume selectively capture a positive (or negative) charge. When rain is formed, these cloud particles associate to form a highly charged raindrop. Suppose that, simultaneously, neutralizing negative (or positive) charges for each droplet are immediately outside and are attached to air molecules or other very small molecular aggregates. Gravity acts on both types of charged carrier, and they fall at a velocity determined by the acting aerodynamic and electrical forces. When the droplets are small, so that gravity does not give the particles high velocities, the neutralizing negative (or positive) charges are carried along with the falling positive (or negative) droplets as a result of electric fields. Thus, after a preliminary small separation, gross separations of the type observed in thunderstorms do not result.

Unfortunately, the literature does not contain an adequate discussion of the important problem of charge separation as influenced by the size of the droplet. Since the droplet size and the acting forces are of the utmost importance in understanding charge separation and lightning processes in the atmosphere, it has seemed worth while to discuss this matter here.

Consider a cloud of infinite extent, lying parallel to the earth's surface and composed of but two types of particles: (1) raindrops upon which positive charge, for example, is selectively deposited; (2) particles (assumed to be small) with a sufficient number of them carrying a total charge just large enough to neutralize the positive charge on the rain droplets. Assume, at first, that the whole cloud system contains exactly as much positive electricity as negative, each uniformly distributed. It will therefore be neutral and no electric fields will exist.

It is noted first that an electric field produced by the separation of charges always acts in such a direction as to prevent the separation. Thus, if positively charged droplets fall with respect to negatively charged droplets, the electric field thus produced acts to support raindrops while simultaneously it drags the small negative elements downward.

It is clear that the electric field can never grow to exceed a value greater than that of the field which will support the droplet. Thus, if E_m is the electric field when the droplets are completely supported by it, q is the charge on the droplet and m is its mass, while g is the acceleration due to gravity, one may equate the electrical and gravitational forces and write

$$E_m = \frac{mg}{q}. \tag{1}$$

This electric field is an absolute maximum in the atmosphere for droplets of a given size and is large enough in general to cause spark discharges.

The equilibrium electric field E_m, described above, is never realized practically because of the conductivity of the earth's atmosphere, which always acts to discharge and reduce any electric field so produced. In order that one may formulate quantitatively the actual equilibrium when the droplets are allowed to fall in an electrically conducting atmosphere, a one-dimensional solution will be obtained by considering the transfer of charge within a prism one square centimeter in cross section and extending vertically through the cloud. The electric current per unit area, i, due to the convected charges on the precipitation, is

$$i = \sum (n_+q_+v_+ + n_-q_-v_-), \tag{2}$$

where n is the number of charged particles per unit volume, q the charge on each particle, and v the velocity of fall, and where the subscripts denote the sign of the transported charge. This downwardly transported net free charge per unit time, i, not only charges the conducting earth below but also supplies charge to replace that conducted upward as a result of the normal ionic conductivity of the atmosphere and the generated electric field. If Q is the total free charge per unit area deposited on the surface of the earth, then, equating the rate of supply to the rate of loss of charge, one has

$$\frac{dQ}{dt} + \sigma E = \sum (n_+q_+v_+ + n_-q_-v_-), \tag{3}$$

where σ is the normal ionic conductivity of the earth's atmosphere, and E is the electric field generated by the charge separation. Under the assumed geometrical conditions, the surface charge density Q is related to the produced electric field E by the relation

$$E = 4\pi Q, \tag{4}$$

from which one may write, for regions near the earth's surface, that

$$\frac{1}{4\pi}\frac{dE}{dt} + \sigma E = \sum (n_+q_+v_+ + n_-q_-v_-). \tag{5}$$

Before integration of this expression is possible, the distributions and the velocities of fall must be expressed as a function of the gravitational forces and electric field. Under most conditions, the velocity of fall may be determined from the terminal velocity of fall of a spherical body in the earth's gravitational field together with known values for the electric field and the mobility of the particle. The terminal velocity of fall, V, for droplets of various sizes has been accurately determined and may be read from tables [6]. The mobility u is defined as the velocity of the particle in unit electric field, whence one may write approximately

$$v = V + uE. \tag{6}$$

Thus, droplets carrying charges of one sign move faster than their normal terminal velocity, while those of opposite sign move slower. Substituting this approximation in (5), assuming that the droplets are all the same, and integrating, one finds that the electric field increases with the time t in accordance with the following relation,

$$E = \frac{n_+q_+V_+\left[1 + \dfrac{n_-q_-V_-}{n_+q_+V_+}\right]}{\sigma - n_-q_-u_-\left[1 + \dfrac{n_+q_+u_+}{n_-q_-u_-}\right]} \left[1 - \epsilon^{-4\pi[\sigma - n_-q_-u_-(1+(n_+q_+u_+)/(n_-q_-u_-))]t}\right], \tag{7}$$

whence, approximating, the maximum *equilibrium* field is given nearly enough by

$$E = \frac{n_+q_+(V_+ - V_-)}{\sigma + n_-q_-u_-}, \tag{8}$$

where all quantities are now expressed as positive numbers. Attention is drawn to the fact that the selection of signs given above is arbitrary, and that in nature a negative instead of a positive charge frequently comes down on rain.

In interpreting equation (7), it is noted that when the positive carriers are small and have very low terminal velocities of fall, their actual velocities closely approximate the velocities of the carriers of negative charges. Thus the difference in terminal velocities becomes small, and the electric field approaches zero. In nonprecipitating clouds, therefore, one would expect that the measured electric fields would be very small; this is in accordance with direct observation [16]. When the rain droplets become reasonably large, the electric fields increase to large values. In fact, according to equation (7), the electric field is proportional to the rainfall intensity and to the free electric charge carried by the larger droplets. Since the negative carriers are very small compared to a raindrop, one may ignore their velocity and calculate Table II from equations (1) and (8), using the best available data [13, p. 94] to show how the equilibrium electric field increases with the size of the raindrop. It is interesting to note in Table II that, while cloud droplets produce a negligible field, the equilibrium field for large droplets is great enough to produce a discharge in air and thus initiate lightning. Equation (8) is therefore consistent with observation.

Using balloons, Simpson and Robinson [26] made measurements purporting to show that the electric fields inside active electrical storm clouds are "of the order of 100 volts/cm." This conclusion is seriously in error, for actual measurements in aircraft show that fields of 1000 v cm^{-1} commonly occur in such clouds without producing a lightning stroke [16]. The electric field on the belly of a B-25 aircraft at the beginning of an energetic lightning stroke has been measured as 3400 v cm^{-1} [14].

Table II. Electric Field and Droplet Size

	Droplet radius (cm)	Electric field to support droplet, (v cm^{-1})	Maximum equilibrium electric field (v cm^{-1})
Fog	5×10^{-4}	1,500	0.5
Drizzle	1×10^{-2}	24,000*	10.8
Medium rain	5×10^{-2}	24,000*	1,930
Excessive rain	1×10^{-1}	24,000*	24,300*

* Because the effective dielectric strength for long discharge paths in air approximates only 3000 v cm^{-1}, active lightning strokes would prevent such high values of electric field from maturing.

From equation (7) it can correctly be inferred that an increase in the conductivity of the atmosphere will reduce the generated electric field. It is not impossible that sudden localized increases in the electrical conductivity of the air, due to a lightning discharge or localized radioactivity, would so increase the conductivity that the generated electric fields would be small even with large droplets and big free charges. Thus, lightning would be suppressed. This matter requires further investigation and may be of importance in the artificial suppression of lightning discharges by localized dissipation of radioactive material into the atmosphere.

The analysis presented above properly emphasizes the dual and interrelated character of thunderstorm electricity as compared with charge production and separation. Electric storm fields would not exist if charges were not actively separated. The analysis shows that separation cannot take place unless the forces acting on the positive charges are different from the forces acting on the negative charges. This implies, in turn, the necessity for a selective deposit of a charge of definite sign on rain particles and a deposit of a charge of opposite sign on lighter cloud particles or air molecules.

Electrification of Aircraft Flying through Precipitation

It was found during World War II that aircraft flying in cold areas systematically lost all radio communication and navigational facilities whenever they encountered dry ice-crystal clouds or snow. Pilots flying through such precipitation in mountainous areas without usable radio navigational facilities continually faced dangerous situations that adversely affected the delivery of urgent war goods to Alaska.

Hundreds of flights made by the Army-Navy Precipitation-Static Research Team near Minneapolis,

through all kinds of precipitation and under varying meteorological conditions, established the basic mechanisms responsible for precipitation static. By using a specially instrumented aircraft, it was found that the two most important sources of electrification of the aircraft were (1) closely adjacent, highly electrified cloud centers, and (2) friction of snow and ice crystals as they slid over the wings at low temperatures. Ordinary rain or shower clouds showing little vertical convection were relatively inactive.

The Research Team was able to demonstrate that the interference with communications on the aircraft resulted from St. Elmo's fire or corona discharges from the aircraft antenna or closely adjacent structures. Because of the intermittent pulselike nature of the corona discharge, adjacent radio circuits were strongly shock-excited and rendered insensitive to the ordinary radio signals.

Normal thunderstorm activity produces large electric fields in the atmosphere; when the plane is in the vicinity of these fields, corona currents frequently are produced on the aircraft. Because an airplane traverses a typical electrically active cloud center in a relatively short interval of time, this type of disturbance (while especially severe) does not persist for long and therefore is not a serious handicap to navigational radio communication. The main difficulty arises from the fact that lightning sometimes strikes the aircraft or that very intense electric fields break down the insulating wire used on antennas.

The second type of electrification is self-produced by the aircraft as a result of frictional effects of snow or ice crystals as they slide over metallic parts of the aircraft. In frontal conditions this type of electrification may last for hours and, because radio navigational facilities must be constantly employed on aircraft, it is evident that such continuous electrification constitutes a dangerous operational hazard.

The frictional charge produced on the airplane proper when flying in dry snow or ice crystals is always negative, while flakes leaving the plane after sliding along its surface carry a positive charge. Experimental investigations have established the fact that the rate of charge production depends on the character of the metallic surface intercepting the precipitation. The charging increases with snow or ice-crystal density and with the cube of the air speed. As one might expect, the Research Team found that the charging rate was dependent upon the temperature, being relatively small near the freezing temperature and increasing as the temperature dropped to about −15C.

It is impractical to review the detailed effects that result from flying through precipitation. Interested readers are urged to read the extensive technical reports of the Precipitation-Static Project [7, 18].

As an illustration of the severity of precipitation static, it seems worth while to give some numerical results obtained off Yakutat, Alaska, in a typical upslope storm. Cloud and charging conditions were singularly uniform and serious electrification was observed for more than three hours. A four-engine B-17 aircraft, cruising at 165 mph, generated an average current of 750 μa. This current transferred a negative charge to the airplane and raised its potential to more than 450,000 v. The electrical energy dissipated, therefore, was 330 w. It was not surprising that corona discharge from the antenna was initiated causing such severe radio interference as to override urgently required communications.

Precipitation static is still a serious operational hazard because the present high operating speed of aircraft has greatly increased the charging rates that were already troublesome at low speed. The modern trend toward still higher speeds will ultimately demand housed antennas for communication purposes. Such construction will greatly assist in meeting future requirements.

REFERENCES

1. Banerji, S. K., "Does Thunderstorm Rain Play Any Part in the Replenishment of the Earth's Negative Charge?" *Quart. J. R. meteor. Soc.*, 64:293–299 (1938).
2. —— "On the Interchange of Electricity between Solids, Liquids and Gases in Mechanical Actions." *Indian J. Phys.*, 12:409–436 (1938).
3. —— and Lele, S. R., "Electric Charges on Raindrops." *Nature*, 130:998–999 (1932).
4. Chalmers, J. A., and Pasquill, F., "The Electric Charges on Single Raindrops and Snowflakes." *Proc. phys. Soc. Lond.*, 50:1–15 (1938).
5. Chapman, S., "Mechanisms of Charge Production in Thunderclouds." Abstract. *Phys. Rev.*, 68:103 (1945).
6. Dinger, J. E., and Gunn, R., "Electrical Effects Associated with a Change of State of Water." *Terr. Magn. atmos. Elect.*, 51:477–494 (1946).
7. Edwards, R. C., and Brock, G. W., "Meteorological Aspects of Precipitation Static." *J. Meteor.*, 2:205–213 (1945).
8. Findeisen, W., "Über die Entstehung der Gewitterelektrizität." *Meteor. Z.*, 57:201–215 (1940).
9. Gilbert, H. W., and Shaw, P. E., "Electrical Charges Arising at a Liquid-Gas Interface." *Proc. phys. Soc. Lond.*, 37:195–213 (1925).
10. Gott, J. P., "On the Electric Charge Collected by Water-Drops Falling through a Cloud of Electrically Charged Particles in a Vertical Electric Field." *Proc. roy. Soc.*, (A) 151:665–684 (1935).
11. Gschwend, P., "Beobachtungen über die elektrischen Ladungen einzelner Regentropfen und Schneeflocken." *Jb. Radioakt.*, 17:62–79 (1920).
12. Gunn, R., "The Free Electrical Charge on Thunderstorm Rain and Its Relation to Droplet Size." *J. geophys. Res.*, 54:57–63 (1949).
13. —— "The Electricity of Rain and Thunderstorms." *Terr. Magn. atmos. Elect.*, 40:79–106 (1935).
14. —— "The Electrical Charge on Precipitation at Various Altitudes and Its Relation to Thunderstorms." *Phys. Rev.*, 71:181–186 (1947).
15. —— "Free Electrical Charge on Precipitation Inside an Active Thunderstorm." *J. geophys. Res.*, 55:171–178 (1950).
16. —— "Electric Field Intensity Inside of Natural Clouds." *J. appl. Phys.*, 19:481–484 (1948).
17. —— and Kinzer, G. D., "The Terminal Velocity of Fall for Water Droplets in Stagnant Air." *J. Meteor.*, 6:243–248 (1949).

18. —— and others, "Technical Reports on Precipitation Static." *Proc. Inst. Radio Engrs., N. Y.*, 34:156P-177P; 234–254 (1946).
19. Lenard, P., "Über Wasserfallelektrizität und über die Oberflächenbeschaffenheit der Flüssigkeiten." *Ann. Phys., Lpz.*, 47:463–524 (1915).
20. Nakaya, U., and Terada, T., "On the Electrical Nature of Snow Particles." *J. Fac. Sci. Hokkaido Univ.*, 1:181 (1934).
21. Pauthenier, M., et Moreau-Hanot, M., "Contrôle expérimental du mouvement de petites sphères métalliques dans un champ électrique ionisé." *C. R. Acad. Sci., Paris*, 194:544–546 (1932).
22. Pearce, D. C., and Currie, B. W., "Some Qualitative Results on the Electrification of Snow." *Canad. J. Res.*, 27(A):1–8 (1949).
23. Scrase, F. J., "Electricity on Rain." *Geophys. Mem.*, Vol. 9, No. 75 (1938).
24. —— "The Air-Earth Current at Kew Observatory." *Geophys. Mem.*, Vol. 7, No. 58 (1933).
25. Simpson, G. C., "Atmospheric Electricity during Disturbed Weather." *Terr. Magn. atmos. Elect.*, 53:27–33 (1948).
26. —— and Robinson, G. D., "The Distribution of Electricity in Thunderclouds, II." *Proc. roy. Soc.*, (A) 177: 281–329 (1941).
27. Wigand, A., "Ladungsmessungen an natürlichen Nebel." *Phys. Z.*, 27:803–808 (1926).
28. Wilson, C. T. R., "Some Thundercloud Problems." *J. Franklin Inst.*, 208:1–12 (1929).
29. Workman, E. J., and Reynolds, S. E., "Electrical Phenomena Resulting from the Freezing of Dilute Aqueous Solutions." Abstract. *Phys. Rev.*, 75:347 (1949).

THE LIGHTNING DISCHARGE

By J. H. HAGENGUTH

General Electric Company, Pittsfield, Massachusetts

The thunderstorm process and the theories relating to the formation of charges in the clouds are explained in another article.[1] This paper will contain information on the lightning discharge only.

The mechanism of the lightning discharge can be divided into three regions of interest: (1) cloud-to-cloud discharges, (2) cloud-to-ground discharges, and (3) phenomena on the ground end of cloud-to-ground discharges. From the practical point of view, the greatest effort has been devoted to understanding and interpreting the phenomena associated with the lightning stroke when it contacts man-made installations [1, 2]. Qualitative data have been obtained to give confidence in the principles of protection used to guard buildings and electrical transmission systems against the effects of lightning. It becomes progressively more difficult to determine the mechanism and physics of the lightning stroke as it occurs away from the earth.

Leaders

The study of the mechanism of strokes to ground has been accomplished primarily by means of photography [15, 16]. The lightning stroke starts at the cloud in the form of a stepped leader, as illustrated in the upper part of Fig. 1. The steps have an average length of 50 m. The time interval between successive steps is of the order of 50 μsec. The average velocity of the individual steps is of the order of 5×10^9 cm sec^{-1}, while the velocity of the total step mechanism is of the order of 1.5×10^7 cm sec^{-1}. Thus, the total time required for the stepped leader to reach the earth may be greater than 0.01 sec.

Table I. Characteristic Data for Leaders and Return Strokes

	Minimum	Average	Maximum
Stepped leaders			
Length of steps, *m*	10	50	200
Time interval between steps, μ*sec*	15	50	100
Velocity of propagation of step, *cm sec*$^{-1}$	—	5×10^9	—
Velocity of propagation of pilot streamer, *cm sec*$^{-1}$	1×10^7	1.5×10^7	5×10^7
Continuous leaders			
Velocity of propagation, *cm sec*$^{-1}$	—	2×10^8	—
Return stroke			
Velocity of propagation, *cm sec*$^{-1}$	2×10^9	5×10^9	1.5×10^{10}

1. Consult "Precipitation Electricity" by R. Gunn, pp. 128–135 in this Compendium.

After the leader reaches the earth, the photographic film shows a much brighter illumination traveling upwards from the earth toward the cloud through the channel established by the stepped leader. This is called the *return stroke*. The average velocity of propagation is 5×10^9 cm sec^{-1}.

Subsequent to this first discharge there may be other discharges. These are also initiated by a leader, but of a different type, called *dart leader* or *continuous leader*, with a velocity of 2×10^8 cm sec^{-1}. As in the case of the first discharge, return strokes result on contact of the continuous leader with the earth. Occasionally leaders on discharges subsequent to the first have been observed to have a few steps. Data on leader and return-stroke characteristics are given in Table I.

Branching

Frequently the first discharge shows branching produced by a stepped-leader process. Branching is always in the direction of propagation of the leader. In general, subsequent leaders do not show branching, choosing the path where ground previously has been reached, resulting in a return stroke. In many cases more than one branch may reach ground simultaneously. In other cases ground is reached at a different point on subsequent discharges, following and completing a branch established on the first discharge. Upward branching has also been observed at the cloud end of the stroke. Photographs show that branching is the result either of discharges between portions of the cloud or of discharges from a different charge center in the cloud through part of the previously established channel. There are indications that branching is more profuse in hilly, wooded country than in flat, bare country.

Theories have been developed to explain the leader mechanism [8, 14]. These theories consider the presence of a pilot streamer which progresses more or less continuously, connecting the individual bright tips of the original leader. As the pilot streamer progresses, the ionization in the upper part of the streamer slows up and recombination of ions occurs. The impedance of the path therefore increases, and hence the potential difference across the established channel is also increased. As the potential is raised, reionization occurs down the channel. The speed and intensity of ionization increase as ionization reaches the tip of the pilot streamer. This increases the energy at the tip and results in greater illumination. The pilot streamer then proceeds for another 50 m and the process is repeated.

The breakdown or ionization process begins in a part of the cloud characterized by a high field gradient. At atmospheric pressure, 30,000 v cm^{-1} are required to begin ionization. At the reduced pressure in the clouds

and due to the presence of water droplets, it is estimated that a gradient of 10,000 v cm^{-1} may be sufficient. The process is started by the acceleration of one or more free electrons in the air. By collision with air molecules, these electrons liberate more and more electrons as they advance in the field, the number increasing very rapidly in the form of an avalanche. Large numbers of positive ions are left behind in the field. The resulting positive space charge increases the field gradient sufficiently for the attraction of photoelectrons. These in turn produce greater ionization and rapidly complete the breakdown process.

ten to several hundred feet. However, the evidence is insufficient to determine whether ground streamers occur on every lightning discharge.

If light intensity is considered a measure of the current amplitude in the return stroke, then in the majority of the cases photographed the first in a series of multiple strokes appears to carry current of the highest amplitude. This is not always the case. Oscillographic evidence also indicates that the first current peak does not always have the highest amplitude.

In many photographs taken in South Africa [9], the light intensity of the channel decreases as the

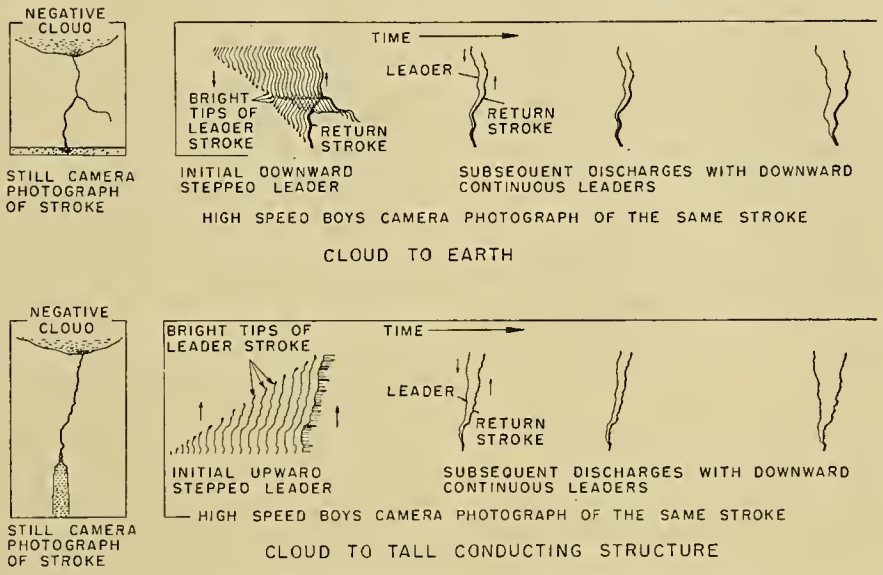


FIG 1.—Schematic diagram showing the mechanism of the lightning discharge from cloud to ground.

Laboratory experiments on long sparks have confirmed the stepped-leader process from negative point to plane, but not from positive point to plane. More sensitive measuring methods are needed to record the complete development of the streamers or leaders. There is considerable doubt as to the existence of the pilot streamer. The theory of the long discharge is not yet sufficiently developed, however, to provide a better answer to the formation of the lightning stroke by means of the stepped leader.

The Return Stroke

As the leader approaches the ground, the charges in the ground begin to move in the direction of the approaching leader. As the leader touches the ground, the charges in the channel and the ground charges can neutralize each other, resulting in the return stroke. Since the channel is partly ionized, further ionization can result at a more rapid rate. The current in the channel increases rapidly and the velocity of propagation upward becomes of the order of ten times that of the continuous downward leader.

There is some evidence that streamers from the ground start up toward the approaching leader, thus producing contact between leader and charges in the air [11]. The length of such streamers may vary from

return stroke progresses toward the cloud. Points of discontinuity are observed, particularly at the junctions of the main channel and the branches. There is a question of whether current surges proceed from the branch to ground or from the ground to the branch. The velocity in most cases exceeds 10^{10} cm sec^{-1}, and therefore cannot be accurately measured from the photographs. Measurements in other parts of the world do not seem to indicate such drastic changes in intensity in the channel as the return stroke progresses toward the cloud. This difference may be due to ground conditions. In those cases where charges are readily available, the current in the return stroke can be maintained up to higher altitudes at higher amplitudes than in ground with high resistivity.

Data are not available on current amplitudes in the leader strokes, the relation between currents and charges in the leader stroke, the charges available in the ground, and the wave shape and amplitude of the return stroke. The mechanism involved in the wave shape of the current in the return stroke is not too clear. It is thought by some that the front is formed shortly before the leader touches the ground and completed on contact [12]. Other theories indicate that maximum peak current is reached at the ground end after the current wave has traveled a short distance upwards in the channel.

Stroke Mechanism for High Buildings

Observations at the Empire State Building in New York [11] have shown that the starting mechanism of the stepped leader is quite different (bottom part of Fig. 1). In most cases the stroke starts as a stepped leader at the building rather than at the cloud. The length of the steps, the time interval between steps, and the velocity of propagation of the steps fall within the range of leaders from cloud to ground. Another significant difference is the absence of a return stroke after the leader has reached the cloud. Instead, a continuous flow of current of the order of magnitude of a few hundred amperes is observed. Frequently the stroke current stops without any further manifestation. In many cases, however, the initial discharge is followed by subsequent continuous downward leaders from the cloud to the building, followed by a return stroke upward from the building, as in the case of strokes reaching the ground in open country.

This sequence leads to the conclusion that charges in the cloud are not sufficiently concentrated to provide a return stroke. In spite of the large drainage of charges (as much as 80 coulombs) in the preliminary continuous-current period and a rather well-ionized channel, the cloud can precipitate a continuous leader to the building. This may be due to a more rapid increase of potential gradients within the cloud or to rapid interchange of charges from other charge centers within the cloud. The reason for establishing a continuous leader in a channel, quite well ionized by the preceding continuing current, is not clear. Photographs and oscillograms show that the return stroke, coupled with a heavy current discharge or peak current, is governed principally by conditions in the ground.

Continuous Current

Oscillographic and photographic evidence from the Empire State Building investigation seems to indicate that successive discharges are always connected by continuing current flow. Other investigators have made measurements which indicate that the current between successive discharges may drop to zero. The fact that successive lightning discharges in a stroke have the same shape is an indication that sufficient ionization remains in the channel for subsequent leaders to choose the same path. In some cases the time interval between such successive discharges is 0.5 sec. In the laboratory it was found that on establishing a long, sixty-cycle arc, a series of multiple discharges take place within a few hundreds of microseconds. The shape of these discharges differs greatly, indicating that ionization has ceased. The currents available in this case are of the order of a few amperes.

To solve this question it is necessary to greatly enlarge our knowledge on deionization time of the air and to determine more accurately whether current flow exists during deionization.

The Channel of the Lightning Stroke

The channel of the lightning stroke almost invariably follows an irregular path. This path and its branches are probably determined by the conditions in the atmosphere surrounding the tips of the stepped leaders as they progress downward. Space charges produced by the leader mechanism, and perhaps charge distributions involved in the thunderstorm process itself, may be largely responsible for the field distortion resulting in the irregular pattern of the lightning channel. The diameter of the channel is apparently a function of the rate of rise of current flowing in the channel and the amplitude of current flow. Experiments have shown that the channel diameter experiences equilibrium when the density of the current flowing reaches 1000 amp cm^{-2}. Much greater current densities, as high as 30,000 amp cm^{-2}, are reached when current flowing in the channel has a rate of rise of a few thousand amperes per μsec. As the result of such measurements, the probable maximum diameter of the channel was deduced as of the order of 5 cm. Other computations and experiments indicate diameters as high as 23 cm.

The change in current density and consequent enlargement of the channel as a result of ionization, heating, and disassociation along the path of the lightning discharge, results in pressure effects. For continuous discharges involving only low currents, the pressure effects are so negligible that thunder cannot be heard. Such observations permit the deduction that pressure effects will be greater the higher the rate of rise of current flowing in the channel and the greater the amplitude of the current. This is confirmed by laboratory experiments where the pressure effects are associated with high surge currents, while low-current discharges produce negligible pressure.

At times the path of the lightning channel is affected by wind. In such cases a stationary camera film and lens will record the multiplicity of the discharge. In some such cases the stroke path changes its shape considerably, indicating perhaps the disruption of current flow. However, it is quite possible that differences in wind velocity at various heights are responsible for distortion of the pre-established channel. In other cases the channel retains its precise form to the end of the stroke.

The Multiple Stroke

The formation of multiple strokes [6, 11] has been explained as the extension of the stroke channel to new charge centers in other portions of the cloud. Alternatively, it has been suggested that new charge centers develop toward the original channel by means of the leader process. In some strokes the regularity of the time interval between successive discharges has been suggested to be due to repeated charging of the original stroke center. Statistical data on the number of multiple discharges in a stroke are given in Table II.

Potential Involved in the Stroke Formation

Based on potential gradients measured at the earth and within clouds, it now appears that a lightning discharge can be initiated and completed with average gradients of the order of 100 v cm^{-1}. On this basis the total voltage required to initiate a stroke of 10,000-ft

length has been estimated at twenty to thirty million volts. Considering the formation of the channel by means of the stepped-leader process, it is reasonable to assume that the average gradient for the lightning stroke may be considerably less than that required to break down a gap in the laboratory (30,000 v cm^{-1}, uniform field; 5000 v cm^{-1}, nonuniform field—large gaps).

It is not known what gradients exist at the point where the leader is initiated in the cloud. The lower density of the air, and the presence of waterdrops with their associated charges, may greatly alter the gradient requirements initiating a discharge.

Cloud-to-Cloud Discharges

While the occurrence of cloud-to-cloud discharges is relatively much greater than that of cloud-to-ground discharges—estimates range between 50:1 and 0.7:1—the mechanism involved is not as well known because these strokes are frequently obscured by the cloud masses. However, the available photographs seem to indicate that cloud-to-cloud discharges are initiated by stepped leaders in the same manner as that of the first discharge in a cloud-to-ground stroke. Return strokes are not known to occur.

Measurements of field changes associated with cloud-to-cloud strokes also indicate the absence of return strokes. The measurement of the wave form of atmospherics, however, has disclosed that some types of cloud-to-cloud strokes result in wave forms similar to cloud-to-ground strokes, but of smaller amplitude and separated by short, quiet intervals. These probably arise from multiple discharges within the clouds. Photographic evidence indicates that discharges within clouds can take place from the tip to the bottom of the cloud, as well as in a horizontal direction within clouds.

The importance of cloud strokes lies in their effects on radio reception and on the safety of airplanes. It has been suggested that airplanes may trigger off cloud-to-cloud discharges, and evidence is accumulating that the susceptibility of planes to lightning strokes increases as the size and speed of the planes become greater.

Phenomena on Ground End of Cloud-to-Ground Discharges

To safeguard electrical installations against damage from lightning strokes, a large number of measurements have been made to determine the characteristics of lightning strokes at the ground end. Such measurements have involved the determination of voltages on transmission lines (maximum measured—5,000,000 v), but have dealt principally with the statistical evaluation of the current in the stroke, the charges involved in the stroke, the wave shapes of currents, and other data needed to provide reliable protective systems.

Figure 2 shows a composite oscillogram of a lightning stroke current to the Empire State Building. In this case an upward leader, followed by a long period of continuing current flow, initiated the stroke and terminated at 0.25 sec. At this moment a continuous downward leader to the building resulted in a return stroke or current peak of approximately 15,000 amp in which the time to half value of the crest current was roughly 40 μsec. A short period of continuing current was followed by a second current peak of 4000 amp. A total of four current peaks were measured in this stroke, the third one having the highest amplitude, 23,000 amp. The total duration of the stroke was approximately 0.46 sec.

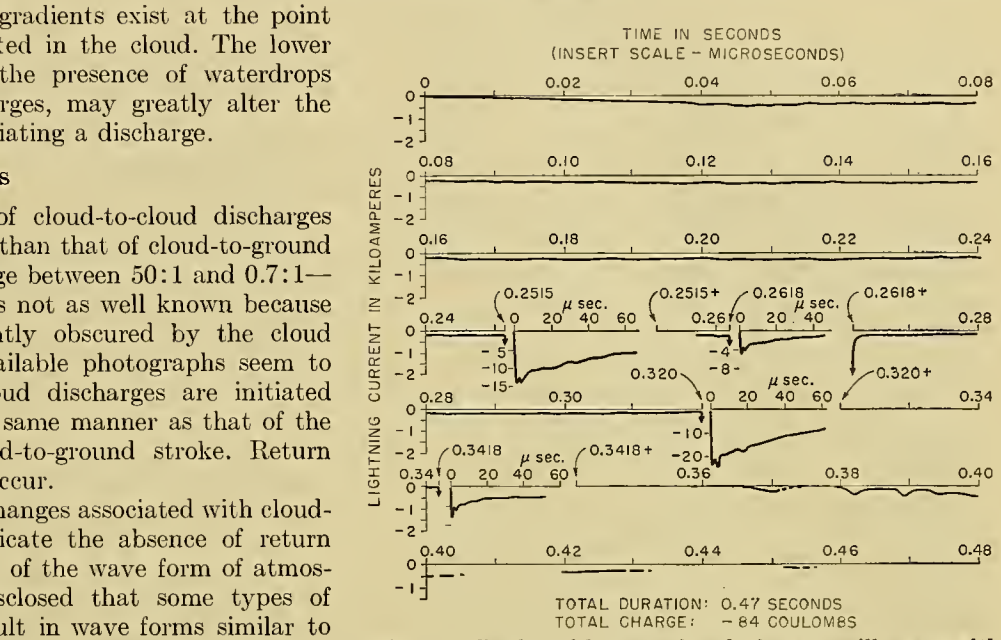


Fig 2.—Replot of low-speed cathode-ray oscillogram with inserts of high-speed cathode-ray oscillograms showing current peaks, Empire State Building, New York City, 1940.

In strokes to open country, the stroke starts at the cloud rather than at the ground, and consequently a current peak would be the first measurement made. The value of continuing current between current peaks is expected to vary greatly and may even be zero. Nevertheless, this oscillogram shows all the principal components of a cloud-to-ground stroke and the relative relations between amplitudes and wave shapes of the two principal components—current peaks and continuing currents—as well as the multiple character of the stroke.

From the available measurements, it has been possible to derive statistical data which are shown in Table II. These statistics in all cases give the minimum and maximum curves obtained from published data for 90, 50, and 10 per cent of the strokes, as well as the maximum available values. The great spread in some of the data is due to the various methods of observations used.

Stroke Current

Item 1 of Table II shows the amplitudes of current peaks measured in the path of the lightning stroke, with an average peak current of 7000 amp. A much greater number of records, which show a surprising agreement, have been obtained on transmission line towers (Item 2). Fifty per cent of the lower currents are in excess of 10,000 amp, while the maximum meas-

ured was 130,000 amp. By making various assumptions, the lightning stroke currents responsible for tower currents measured were deduced as shown in Item 3. The correctness of some of these assumptions is questionable, and the values shown in the table are probably

TABLE II. CHARACTERISTICS OF LIGHTNING STROKES

Item		Per cent of strokes with values in excess of those shown in Table I			Maximum
		90%	50%	10%	
1. Current peaks measured in stroke path, *kiloamperes*	Min.	2.2	6.0	20.0	—
	Max.	5.0	8.6	27.5	160
2. Current amplitudes in steel towers, *kiloamperes*	Min.	1.0	8.8	28.4	—
	Max.	5.3	12.2	35.8	130
3. Lightning stroke currents computed from Item 2, *kiloamperes*	Min.	2.4	13.3	50.0	—
	Max.	10.3	40.0	101.0	220
4. Rate of rise of stroke currents, *kiloamperes per μsec*	Min.	1.8	8.5	27.5	—
	Max.	2.4	14.5	35.0	45
5. Wave front of stroke currents, *μsec*	Min.	0.2	1.0	5.2	—
	Max.	0.9	2.51	5.5	10
6. Wave tail of current peaks—time to half value, *μsec*	Min.	11.6	30.3	78.0	—
	Max.	26.0	45.5	80.0	120
7. Charges in current peaks, *coulombs*	Min.	—	—	—	—
	Max.	0.04	0.23	1.03	5.6
8. Total stroke charges, *coulombs*	Min.	2.3	10.4	86.0	—
	Max.	4.2	22.2	100.0	165
9. Total duration of strokes, *sec*	Min.	—	0.0006	0.2	—
	Max.	0.1	0.37	0.68	1.6
10. Number of current peaks in lightning strokes	Min.	1.0	1.8	4.0	—
	Max.	1.3	3.0	11.0	42
11. Time interval between current peaks, *sec*	Min.	—	0.02	0.1	—
	Max.	0.03	0.088	0.18	0.5

too high. The statistical evidence indicates, however, that only very few lightning strokes will have current peaks in excess of 60,000 amp.

Wave Shape of Current Peaks

Of great importance for evaluating the effect of lead length in lightning protective systems is the knowledge of the length of the front and the rate of rise of current of the current peaks. The number of records available is relatively small. However, the data taken by different investigators with different means of recording are in sufficient agreement to allow confidence in the results. The limits of wave fronts and rates of rise measured indicate that inductive drops are a serious consideration. In 50 per cent of the cases the rate of rise was 12,000 amp per μsec or greater, which would result in approximately a 6000-v drop per foot of conductor.

The duration of the current peaks is of importance in estimating or determining the strength of insulation subjected to lightning strokes. For practical reasons this is usually expressed in terms of microsecond duration of the current wave while it rises from zero to crest and decays to half value. The statistical evidence shows that 50 per cent of the waves have a time to half value of approximately 38 μsec or more.

The charges in the current peak are related by an unknown factor to the charges laid down by the leader. The charges measured at the ground end of the stroke are expected to be lower than those in the leader because of losses during the formation of the leader. The charges given in the table are based on the integrated product of current in amperes and time in seconds, using only the portion of the current peak between its start and its decrease to half value. In all cases but one, the charges thus determined were less than one coulomb and resulted from lowering a negative charge from the cloud. The maximum value of 5.6 coulombs resulted from lowering a positive charge from the cloud.

The Composite Stroke

The lightning stroke may consist of a number of current peaks and continuing current flow. The total duration of the stroke measured by different investigators varies between 0.0006 sec and 0.35 sec for the 50 per cent level. The longest duration measured is about 1.5 sec. Fifty per cent of the strokes have two or more current peaks, while a maximum of 42 current peaks has been measured. The time interval between successive current peaks at the 50 per cent level varies between 0.02 sec and 0.09 sec, with a maximum of 0.5 sec between successive discharges.

Of considerable interest is the total charge in the lightning stroke. This is principally the charge conducted by the continuous currents. The average charge measured is approximately 18 coulombs, while a maximum of 165 coulombs has been recorded. Many of these measurements were obtained from strokes to tall buildings, and in all cases they represent the charges at the ground end of the stroke. They represent the total charges conducted in the channel through the point of measurement. There should be some difference in the total charges conducted when the stroke is initiated at the cloud and when it is initiated at a tall structure. In the latter case, the total charge in the channel can be measured, while for strokes initiated at the cloud, charges laid down from the cloud are not necessarily included in the measurement.

Measurement at the ground end of the stroke will not necessarily record all charges involved in a lightning stroke, particularly in the not-too-rare cases where the stroke changes its path at the ground end or at both ends. It is obvious that in such cases the total charges involved in the stroke mechanism can be considerably greater.

Charge determination by means of electric field measurements of thunderstorms has indicated a maximum charge of 200 coulombs, a value about 25 per cent higher than at the Empire State Building. A further indication of total charges involved in lightning strokes has been obtained from damage produced by lightning strokes on metal parts, particularly of airplanes. Comparison of such damage with similar effects produced in the laboratory has indicated charges at least as high as 300 coulombs, and possibly in excess of 500 coulombs. In some of these high total charges, strokes to ground were also involved, but since the total stroke mecha-

nism is not known, it cannot be determined whether all of the charges were conducted to ground or to what extent they occurred within the cloud only.

Polarity

The polarity of lightning strokes, in the great majority of cases measured, is negative; that is, negative charges are lowered from the cloud. The relation for grounded structures is approximately 15:1 in favor of negative charges. Since most measurements have been made on transmission line towers and other rather well-grounded structures, it has been pointed out that this would be conducive to a higher number of negative strokes on account of the possibility of guiding a negative leader to the structure by means of streamers produced at the positive grounded object. In the case of a positive stroke, a negative streamer would not be as likely to occur. From this reasoning it is expected that strokes to open ground may include a lower percentage with negative polarity. Measurements of field changes seem to favor this view because in many of these cases the ratio of negative to positive polarity cloud-to-ground strokes is of the order of 6:1 or less.

In cases where direct strokes to ground have been measured by means of oscillographs or other devices which permit detailed examination of the stroke current throughout its duration, it was found that the majority of the strokes were entirely of negative polarity. In some cases, however, the polarity of the continuing current flow changed, usually toward the end of the stroke. In other cases one of the current peaks was of positive polarity, while the remaining current peaks resulted from a negative cloud charge. The polarity relations throughout the length of the stroke are probably governed by the mechanism of multiple discharges which was discussed previously. The process must be governed by the means of charge exchange within the cloud during and following the first leader stroke to ground, as well as by the rate of production of charges within the original stroke center in the cloud.

The fact that charges of negative polarity are lowered to the ground in the majority of the cases seems to indicate that negative charges predominate in the bottom of the cloud. The fact that some of the highest current peaks measured were of positive polarity might be accepted as proof that centers of positive polarity can exist in the lower portions of some storm clouds.

Stroke Density

It is extremely difficult to obtain information on stroke density. Analysis has shown that the density per square mile is approximately one-half of the number of storm days from the isokeraunic map. This was partially confirmed by a two-year count in a region with 27 storm days per year where the average stroke density was approximately 15 strokes per square mile per year. Observed figures depend on the size of the area under observation and rapidly decrease with increased area. Accurate data would be of value to determine lightning risks.

Forms of Lightning

Names like *streak*, *bead*, *ribbon*, *fork*, *heat*, *sheet*, and *ball lightning* have been used to describe various observed forms of lightning discharges. Streak lightning is the normally observed type as described in the foregoing discussion. Bead, ribbon, fork, and heat lightning probably have the same characteristics as streak lightning. Bead lightning is assumed to be a form of streak lightning. The appearance of the beads may be caused by variations in the luminosity along the channel, perhaps caused by brush discharges. Ribbon lightning is probably a streak lightning stroke with multiple discharges where the channel is blown along by the wind. Fork lightning is the term used for strokes with several apparently simultaneous paths to ground. As explained, these ground termini may be formed simultaneously or during successive discharges of multiple strokes. Heat lightning is a form of streak lightning at a distance sufficiently far away so that thunder is not heard. Sheet lightning usually occurs in clouds between the lower and upper atmosphere over a considerable area. This areal distribution of sheet lightning, and its long persistence, constitute the principal difference between it and streak lightning.

Ball lightning is described as a luminous ball of reddish color, with an average diameter of 20 cm. When seen emerging from the cloud, its velocity has been estimated as high as 100 m sec^{-1}; while on the ground it travels at 1 to 2 m sec^{-1}. Usually the ball explodes after an average life of 3 to 5 sec. There is much controversy regarding this form of lightning. It has been described as an optical illusion caused by retention of vision of a heavy lightning discharge in the retina of the eye. However, the testimony of a few apparently well-qualified observers seems to indicate that such phenomena may exist. No reliable explanation for the existence of ball lightning has been offered. However, several theories exist explaining the phenomenon on the basis of chemical and physical reactions. Some photographic evidence submitted as possibly caused by ball lightning has been proved to have had an erroneous interpretation.

Measurements of Lightning Characteristics

The most exact measurements can be made at the ground end of a stroke. For this purpose a variety of instruments have been developed.

The Lichtenberg figure on photographic film [7, 13] placed between two electrodes is a corona discharge. The size of the figure is a measure of the voltage applied and the polarity of the discharge. By suitable shunts, currents can be measured. The accuracy of such devices has been estimated at 25 to 50 per cent.

A very simple method of measuring the peak value and polarity of surge or lightning currents is the magnetic link [4]. It consists of a number of steel strips of high retentivity. The magnetic flux associated with a lightning stroke magnetizes the link. The magnetization measured after exposure gives a very accurate indication of current crest. Two links at different spacing from

the conductor carrying the lightning surge are used to increase the accuracy, particularly where currents of opposite polarity may flow. These links will measure only the maximum peak in a multiple stroke.

Fulchronographs [17], devices where a number of links are mounted on a rotating wheel, permit determination of the variation of current with time. The resolution for continuing currents is good. The high-speed wheels used permit separation of amplitudes of successive high current peaks, but the speed is insufficient to determine the wave shapes of current peaks.

Magnetic links applied in circuits in which inductances or capacitances are used in combination with resistance permit determination of the maximum rate of current change of a current peak.

The photographic surge-current recorder [10] measures the intensity of light produced by a surge current across a short spark gap. A range of 0.1 amp to 150,000 amp is claimed for this device.

The cathode-ray oscillograph [3] is the most versatile, but also the most elaborate device for measuring surges. Several types of oscillograph must be used to record the high-speed (current peaks) and slow-speed (continuing currents) components of a complete lightning stroke. The development of the sealed-in cathode-ray tube has made this instrument simpler and subject to automatic operation.

Of considerable value for determining the mechanism of the discharge are the so-called *Boys cameras* [5, 15]. With high-speed rotating film, or high-speed rotating lenses, the speed of propagation of the leaders and return strokes can be analyzed. The low-speed cameras give valuable information on the time sequence of current peaks and the continuing-current flow between multiple discharges. It has been possible to obtain fairly good correlation between density of the film exposed to a stroke and the current flow causing the illumination, principally for the continuing components. A microphotometer has been used to correlate these quantities. For accuracy it is necessary to avoid overexposure of the film, as well as to have highest sensitivity for the weaker illumination. For this purpose multiple-lens cameras are used with apertures varied to cover the complete range of exposure expected in a straight-line relation.

Electric field measurements permit investigation of lightning-stroke phenomena from the point of view of the charges involved. Some of the measurements made differentiate the leader, the return stroke, and the continuing-discharge portions of strokes. The wave shapes of atmospherics and correlation with various forms of lightning discharges have been investigated. Such measurements have been extended to cover distances of hundreds of miles. By using two or more instruments, the distance of the source of atmospherics can be determined with good accuracy.

A rather useful means of determining lightning characteristics is the examination of damage produced by lightning. By reproducing similar effects in the laboratory it is possible to determine approximately the type of stroke responsible for the damage, as well as the amplitude of current peaks and the charge conducted in the stroke. Damage produced on metal parts of airplanes is in many respects the only means by which lightning currents occurring in the clouds can be determined.

Lightning Protection

The statistical knowledge gained over the past fifteen to twenty years has largely confirmed the effectiveness of a properly installed lightning-rod system (advocated by Benjamin Franklin more than 150 years ago) to protect ordinary buildings and houses against direct strokes. Such systems consist of interconnected air terminals mounted on the highest points of the structure, and connected to the ground at several points. The "Code for Protection Against Lightning" issued by the National Bureau of Standards and designated as *Handbook H40*, and the *British Code of Practice C. P. 1:1943* published by the British Standard Institution, contain the essential rules to follow for installation, materials, size, and other factors essential for an effective durable installation.

For continuity of electrical service to homes, farms, and factories, many practices have evolved, based directly on the many lightning studies and their results. For protection of electrical apparatus, the lightning arrester is a commonly accepted device. Such arresters have the property of conducting surge currents to ground at a reduced voltage. Any system power current (follow current) which may flow through the arrester immediately following the lightning discharge is interrupted, and the line is restored to its original condition. Many different types of arresters are in use. In some cases plain gaps are used for protection. These cannot limit the voltage applied to the apparatus to voltages as low as arresters. After sparkover of the gap occurs, circuit power current usually follows and must be interrupted by opening a breaker.

Transmission lines of higher voltage ratings can be effectively protected by the use of ground wires properly suspended above the transmission wires to intercept the direct strokes. The lightning currents contacting the ground wire and the towers to which they are connected will raise the tower potential by virtue of the resistance of the tower to ground and the current in the stroke. To this is added the inductive voltage drop caused by the inductance of wire and tower and the rate of rise of current on the current front. By reducing the ground resistance to less than one ohm per 12 kv of circuit voltage, it has been possible to reduce outages on circuits of 66 kv and above to an extremely low value. Ground resistance can be reduced by the use of buried wires—counterpoise—or deeply driven ground rods.

Lower voltage circuits have been made lightning-resistant in many cases by the use of wood as insulation in addition to the normal porcelain or glass insulators. Reclosing circuit breakers are another tool for preventing circuit outages. These breakers are able to open a circuit and reclose it in as little as one-fifth of a second by means of suitable relaying. Proper balance between

circuit breaker duty and protective means for reducing the number of flashovers must be considered.

Conclusions

The physical phenomena of the lightning discharge are not entirely understood. Photographic evidence indicates a stepped-leader initiation of the stroke followed by a return stroke from its ground terminal. Theoretical stipulation of a pilot streamer has not been proven.

To complete the understanding of the physical phenomena involved in lightning discharges, more information is required on these principal questions:

1. The gradient at the point of origin of the stroke.
2. Gradient distribution within clouds, as well as beneath clouds.
3. Gradients at the ground end of a stroke.
4. Existence and character of ground streamers prior to stroke contact with the ground.
5. Ionization processes within the stepped leader, continuous leader, and the return stroke.
6. Ionization and deionization of the stroke channel with special reference to continuing current discharges.
7. Influence of ground conditions on the return-stroke process.

For further progress on the protection problem, statistical evidence is desirable on:

1. The wave shape of lightning stroke currents of both current peaks and continuing currents.
2. The distribution of such currents in the ground network of protective installations, as well as in the earth.
3. The lightning stroke density in various regions of the earth.
4. The incidence of lightning to various structures as determined by height and physical location with regard to other structures and natural terrain.

From the numerous investigations of lightning undertaken during the last twenty-five years, it has been possible to devise protective systems and practices which reduce damage due to lightning to a negligible factor on electrical installations as well as on buildings.

REFERENCES

The sources of information are extensive. References to 750 papers are found in [1] and [2] below. To keep the number of references at a reasonable figure, only a few additional references are listed.

1. *A.I.E.E. Lightning Reference Book, 1918–1935*. New York, American Institute of Electrical Engineers, 1937.
2. *A.I.E.E. Lightning Reference Bibliography, 1936–1949*. New York, American Institute of Electrical Engineers, 1950.
3. Flowers, J. W., "The Direct Measurement of Lightning Current." *J. Franklin Inst.*, 232: 425–450 (1941).
4. Foust, C. M., and Kuehni, H. P., "The Surge-Crest Ammeter." *Gen. elect. Rev.*, 35: 644–648 (1932).
5. Hagenguth, J. H., "Lightning Recording Instruments." *Gen. elect. Rev.*, 43: 195–201, 248–255 (1940).
6. Larsen, A., "Photographing Lightning with a Moving Camera." *Rep. Smithson. Instn.*, pp. 119–127 (1905).
7. Lee, E. S., and Foust, C. M., "The Measurement of Surge Voltages on Transmission Lines Due to Lightning." *Trans. Amer. Inst. elect. Engrs.*, 46: 339–356 (1927).
8. Loeb, L. B., and Meek, J. M., "The Mechanism of Spark Discharge in Air at Atmospheric Pressure." *J. appl. Phys.*, 11: 438–447, 459–474 (1940).
9. Malan, D. J., and Collens, H., "Progressive Lightning, III." *Proc. roy. Soc.*, (A) 162: 175–203 (1937).
10. McCann, G. D., "The Measurement of Lightning Currents in Direct Strokes." *Trans. Amer. Inst. elect. Engrs.*, 63: 1157–1164 (1944).
11. McEachron, K. B., "Lightning to the Empire State Building." *J. Franklin Inst.*, 227: 149–217 (1939).
12. —— and McMorris, W. A., "The Lightning Stroke: Mechanism of Discharge." *Gen. elect. Rev.*, 39: 487–496 (1936).
13. Peters, J. F., "The Klydonograph." *Elect. World, N. Y.*, 83: 769–773 (1924).
14. Schonland, B. F. J., "Progressive Lightning, IV." *Proc. roy. Soc.*, (A) 164: 132–150 (1938).
15. —— and Collens, H., "Progressive Lightning." *Proc. roy. Soc.*, (A) 143: 654–674 (1934).
16. Schonland, B. F. J., Malan, D. J., and Collens, H., "Progressive Lightning, II." *Proc. roy. Soc.*, (A) 152: 595–625 (1935).
17. Wagner, C. F., and McCann, G. D., "New Instruments for Recording Lightning Currents." *Trans. Amer. Inst. elect. Engrs.*, 59: 1061–1068 (1940).

Further articles containing comprehensive references are:

18. Bruce, C. E. R., and Golde, R. H., "The Lightning Discharge." *J. Instn. elect. Engrs.*, 88: 487–505 (1941).
19. McEachron, K. B., "Lightning and Lightning Protection" in *Encyclopaedia Britannica*, Vol. 14. Chicago, 1948. (See pp. 114–116)
20. —— and Hagenguth, J. H., "Lightning and the Protection of Lines and Structures from Lightning" in *Standard Handbook for Electrical Engineers*, A. E. Knowlton, ed. New York, McGraw, 8th ed., 1949. (See pp. 2230–2254)
21. Meek, J. M., and Perry, F. R., "The Lightning Discharge" in *Reports on Progress in Physics*, 10: 314–357. Phys. Soc., London, 1944–45.

INSTRUMENTS AND METHODS FOR THE MEASUREMENT OF ATMOSPHERIC ELECTRICITY*

By H. ISRAËL

Institute for Atmospheric Electric Research of Buchau a. F.

System of Units

Various systems of units are used in electricity. Each of these is an entity in itself and is built up on a definite fundamental physical relationship:

1. The "electrostatic cgs system" is based on Coulomb's law of the force effects of interacting electrical charges and arbitrarily takes the proportionality factor to be nondimensional and equal to unity.

2. The "electromagnetic cgs system" is based on the Biot-Savart law of the forces between an electric current and a magnetic pole, and again the proportionality factor is set equal to unity.

3. Giorgi's "natural m-sec-v-amp system" assumes voltage and amperage as fundamental quantities, with the second (sec) as the time unit and the meter (m) as the length unit.

In the first two systems the electrical quantities are reduced to the mechanical units, cm, g, and sec (cgs), which appear in complicated, nonvisualizable combinations. The third system of units is more suitable to the scope of physics in general; therefore, its adoption for consistent use in the field of atmospheric electricity is recommended. In this system, the international standards for volt, ampere, etc., are employed.

Auxiliary Equipment and Practical Suggestions

The atmospheric electrical problems of measurement consist predominantly of the measurement of potential differences of medium magnitude and of minute electrostatic charges. They belong, therefore, to the domain of electrometry proper. Although increasing use is being made of "vacuum tube electrometers" (d-c current amplifiers or d-c voltage amplifiers), thorough familiarity with the electrometer is a prerequisite for work in atmospheric electricity.

All electrometers are based on the principle of electrostatic attraction or repulsion. An exception is the capillary electrometer which utilizes the change in the surface tension of mercury when traversed by an electric current. Therefore, strictly speaking, the capillary electrometer represents a type of galvanometer [103]. Quadrant electrometers have long oscillation periods and are therefore suited chiefly for the measurement of constant potential differences, whereas filament electrometers adjust themselves aperiodically and almost instantaneously at sensitivities that are not excessively high. The highest sensitivities are attained with the modern modifications of the quadrant electrometer.

Figure 1 shows schematically the well-known principle of the quadrant electrometer. Special electrometer types are: the model developed by Elster and Geitel [37], particularly suited to photographic recording; the Dolezalek electrometer [30, 31] for high sensitivities; a special design by Schultze [119]; the "Binanten" electrometer [49] and the "Duanten" electrometer of Hoffmann [48, 50, 51]; the Benndorf electrometer [11, 12], the familiar, low sensitivity quadrant electrometer for mechanical recording of potential gradients.

Figures 2 and 3 are sectional drawings of the bifilar electrometer without auxiliary potential and of the unifilar electrometer with auxiliary potential according to Wulf's design [152]. More modern designs for attaining the highest sensitivites include those of Lindemann and Keeley [85], Compton [26, 27], Shimizu [127], Swann [139] and Perucca [104, 105].

The development of the so-called electrometer tubes with their characteristically high insulation of the control grid and very low grid current (10^{-13} amp and less) has resulted in the substitution of vacuum-tube electrometers for the ordinary electrometers. Many types of electrometer tubes are now commercially available. During continuous operation, maintenance of a constant zero point is sometimes difficult; some amelioration can be provided by using oversized filament batteries and by the use of selected pairs of tubes in a bridge circuit (for further details see, for example, [22, 28, 36, 46, 47, 69, 77, 82, 106, 109, 114]).

There are several types of circuits available for use with quadrant and filament electrometers provided with auxiliary potentials: In the *idiostatic circuit* the needle (lemniscate or electrometer filament) and one pair of quadrants (knife edge) are grounded; the unknown voltage is applied to the other pair of quadrants (knife edge). The deflection is proportional to the square of the unknown voltage, thus making possible a-c measurement. In the *quadrant circuit* the needle is at a fixed high voltage; one pair of quadrants is grounded, and the unknown voltage applied to the other. Deflection is proportional to the unknown voltage. The *heterostatic* or *needle circuit* has both pairs of quadrants connected to a fixed auxiliary voltage of opposite sign and the unknown voltage applied to the needle. This is the most frequently used and most sensitive circuit, and has the lowest capacitance. In the *current circuit* the voltage measurement is made at the terminals of a resistor which is usually of the high-ohmic type.

Measurements of atmospheric electricity are almost exclusively electrostatic measurements and therefore call for high quality of the dielectric materials. The best insulating materials are the natural products, amber and sulfur. Good substitutes include high quality plastic dielectric materials such as hard rubber (ebonite), plexiglass, and Trolitul. Meticulous surface treat-

* Translated from the original German.

ment is particularly important. Such treatment includes high polish, drying, dust removal, paraffin coating in some cases, and protection against direct sunlight. When working in the open air, careful maintenance of the dryness of the dielectric surfaces by drying with hot air or by electrical heating is essential. The relative humidity on the insulator must not exceed approxi-

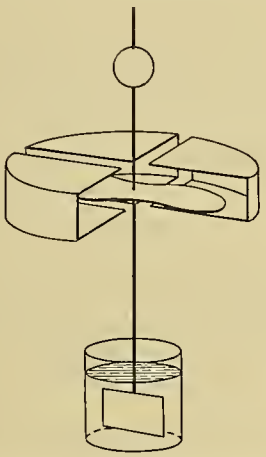


FIG. 1.—Schematic diagram of a quadrant electrometer.

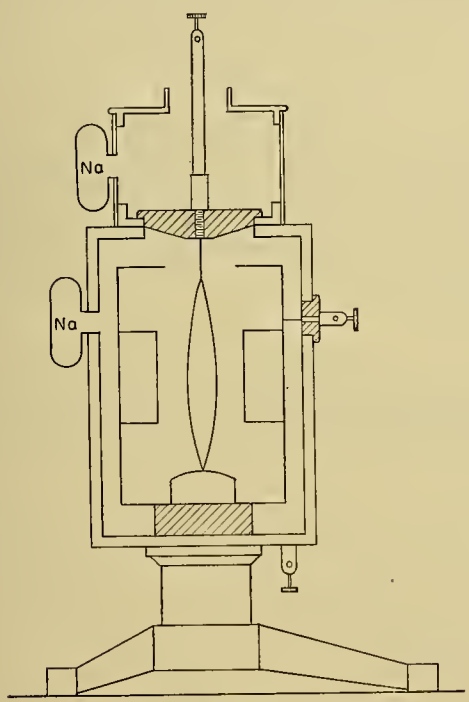


FIG. 2.—Wulf's bifilar electrometer.

mately 80 per cent. Figure 4 gives one example of a suitable form of open-air insulator which has held up well in practice.

To combat insects and spiders it is helpful to paint the metal parts about the air gap with an adhesive substance (fly-paper glue) and also with aromatic insect repellants. To eliminate air-borne spider threads in summer and fall, mechanical methods must be applied.

In all electrostatic measurements it is essential to maintain good grounding of all parts to be maintained at zero potential. In general, grounding to a water pipe is satisfactory. It is particularly important that all parts to be grounded are connected to the same ground

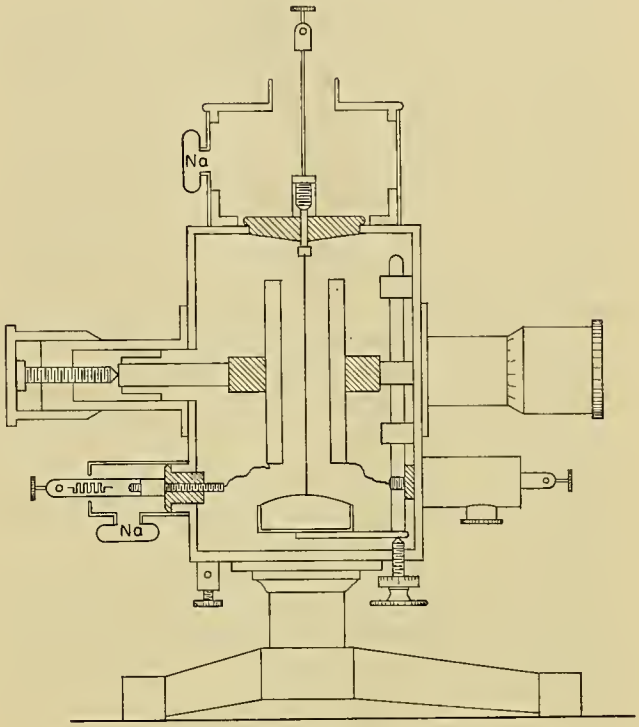


FIG. 3.—Wulf's unifilar electrometer.

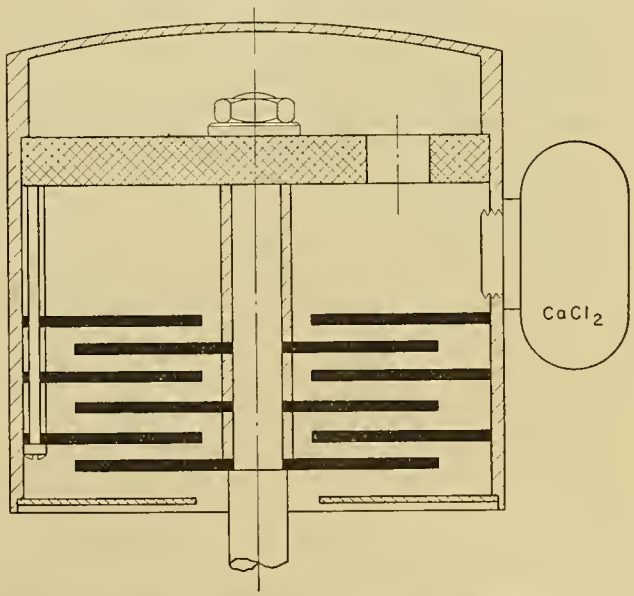


FIG. 4.—Open-air insulator for large mechanical load (antennas and the like).

connection. All other leads must be protected from induction by shielded cables.

With high electrometer sensitivities undesirable disturbances may appear as a result of certain unforeseeable phenomena in the insulating dielectric material, such as the formation of deposits and polarization phenomena. Direct creep of charges across an insulator

can be prevented by double insulation separated by a grounded metal plate. In all cases, when working with maximal sensitivities, it is judicious to protect the insulators against either direct or indirect influence of electric fields by shielding them.

On many occasions, especially when working with movable parts such as rotary collectors, unwelcome concomitant phenomena are possible as a result of the Volta effect. The latter must be determined and corrected for by check tests.

In electrostatic work, the almost universal contamination of rooms by a-c fields produced by the electric light network is a source of serious disturbance. Its effect can, in general, be easily recognized and eliminated when working with electrometers. When amplifiers are used such a-c fields may easily lead to errors of measurement and misinterpretation (apparent atmospheric electrical potential gradient in rooms) because of unintended and undetectable rectifier action.

Batteries are preferable to rectified a-c lines as sources of constant potential, especially in electrometric work. Portable high-voltage sources for counting-tube measurements in open country are described elsewhere [133, 136].

High ohmic resistors up to about 10^{13} ohms are fabricated commercially by evaporation of thin platinum films on quartz or amber. For homemade high ohmic fluid resistors, radioactive (Bronson) resistors, and other possibilities, see von Angerer [3].

Ions and Other Atmospheric Suspensions

Ionizers of the Atmosphere. The ionization of the lower atmospheric layers, aside from occasional local ionizers of subordinate significance (waterfall effect, combustion gases), is caused by the α, β, and γ radiation of radioactive substances and by cosmic radiation. The special methods for measuring radioactive substances and cosmic radiation are treated elsewhere in this Compendium.[1]

The ionization in a sealed chamber is due to radiation from the chamber walls (mathematical determination according to von Schweidler [120], experimental determination in mines [17]), radiation from the earth, radiation from the atmosphere, and cosmic radiation. The mathematical estimation of the radiation from the earth and the atmosphere is made as follows:

If ρ_{earth} and ρ_{air} are the concentrations of radium, or of its *RaC*-equivalent, in a cubic centimeter of earth or of air, μ_{earth} and μ_{air} the absorption coefficients of the corresponding γ-radiation in earth or in air, and if K is the "Eve number" (4.0×10^9 in the absence of secondary radiation, approximately 5 to 6×10^9 in thick-walled ionization chambers), the ion production q (*i.e.*, number of ion pairs formed per cubic centimeter per second) is given by

$$q_{\text{earth}}(0) = 2\pi\rho_{\text{earth}}K/\mu_{\text{earth}}, \quad \text{Radiation instrument directly on the earth's surface} \quad (1)$$

$$q_{\text{earth}} = 4\pi\rho_{\text{earth}}K/\mu_{\text{earth}}, \quad \text{Radiation instrument in the ground (caves, tunnels, etc.)} \quad (2)$$

$$q_{\text{earth}}(h) = q_{\text{earth}}(0)\phi(h\mu_{\text{air}}), \quad \text{Radiation instrument at an altitude } h \text{ above the earth's surface} \quad (3)$$

wherein

$$\phi(x) = e^{-x} - x\int_x^\infty \frac{e^{-u}}{u}\,du,$$[2]

$$q_{\text{air}}(0) = 2\pi\rho_{\text{air}}K/\mu_{\text{air}}, \quad \text{Radiation instrument on the surface of the earth} \quad (4)$$

$$q_{\text{air}} = 4\pi\rho_{\text{air}}K/\mu_{\text{air}}. \quad \text{For great altitudes} \quad (5)$$

Portable radiation devices, which are convenient to manipulate, have been described elsewhere [78, 79, 151], as have counting-tube instruments for field work [133, 136].

Conductivity, Concentration of Ions, and Ion Mobility. If an electric potential is applied to two electrodes in an ionized gas, a weak current begins to flow. Corresponding to the two oppositely flowing ionic currents, the current density is the combination of two terms:

$$i = \epsilon(k_1 n_1 + k_2 n_2)E = \Lambda E, \quad (6)$$

where ϵ is the charge of the ion and equal to 1.6×10^{-19} amp sec, E is the field intensity, n_1 and n_2 denote the number of positive and negative ions, and k_1 and k_2 represent the mobility of the positive and negative ions, respectively. The expression $\epsilon(k_1 n_1 + k_2 n_2) = \Lambda$ is designated as the (total) conductivity of the gas. The terms

$$\lambda_1 = k_1 n_1 \epsilon \text{ and } \lambda_2 = k_2 n_2 \epsilon$$

are the positive and negative polar conductivities of the ionic conductor.

FIG. 5.—Schematic curve of the current-potential relationship (characteristic) in an ionized gas at rest.

If the voltage is increased starting from zero, a gradual decrease in ion content results; the current does not rise in proportion to the voltage and remains constant after a given value of voltage has been attained. Accordingly a distinction is made between *ohmic*

1. Consult "Radioactivity in the Atmosphere" by H. Israël, pp. 155–161.

2. For a tabulation of this function, see [86].

current (I), *semi-saturation current* (II), and *saturation current* (III); see Fig. 5.

To attain clearer experimental conditions, it is now a general practice to employ the aspiration condenser in connection with the so-called *method of perpendicular velocities* as follows: When air containing ions flows through a condenser to which an electric field has been applied, the trajectory of an ion is found to be the resultant of the two mutually perpendicular forces, namely, that of the air current and that of the field. If M is the aspirated quantity of air in cubic centimeters per second, C is the condenser capacitance ($C = L/[2 \ln (R/r)]$ for cylindrical condensers having radii R and r and length L), and V is the potential in volts, the expression

$$k_g = M/4\pi CV \qquad (7)$$

represents the *limiting mobility* of the condenser and states that all ions whose mobility is greater than, or

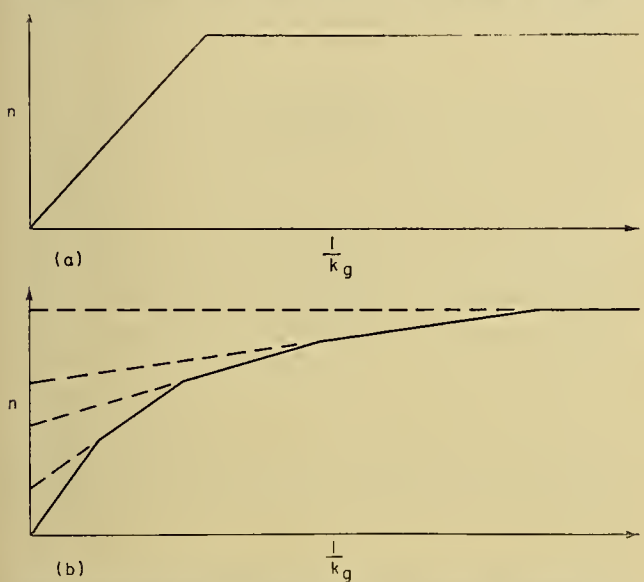


FIG. 6.—Schematic diagram of the current-potential characteristic in the aspiration condenser (*a*) in the presence of one type of ion, (*b*) in the presence of several types of ions.

at least equal to, this mobility are deposited. Of those ions, whose mobility k is smaller, only the percentage k/k_g is deposited.

It is evident that, in the foregoing, the characteristic relationship between current and potential in the condenser will be a broken linear curve. This curve will resemble Fig. 6*a* when but one type of ion is present, and assume the form of Fig. 6*b* when several types of ions occur. From the preceding statement, the following conditions are found for the measurement of ions:

Conductivity: M must be chosen so great or V so small that operations take place in the first segment of the curve that rises from the origin of the coordinate system.

Ion Counts: Determined from the saturation current.

Ion Mobility: Each break in the curve yields, according to equation (7), the mobility of a corresponding type of ion.

Ion Spectrum (numerical distribution on the basis of the individual mobilities): The number pertaining to a given mobility results from the magnitude of the change in slope of the characteristic; expressed in terms of differentials, the ion spectrum is determined by the second differential quotient of the characteristic.

Additional methodological details are given elsewhere [54, 56, 57]. With respect to "edge disturbances" (effect of the inhomogeneity of the field at the edge of the condenser), see Itiwara [73], Israël [56], and others. For special designs, that homogenize the field of cylindrical condensers, refer to Becker [8], Swann [137], and Scholz [115].

Well-known instruments that are easy to manipulate include the Gerdien aspirator [39, 40] for measurements of conductivity; the Ebert ion counter [32, 33, 35] and the Weger aspirator [58, 142] for measurements of the concentration and mobility of small ions (see references [6, 41, 42, 142] for errors of the Ebert instrument), and the Israël ion counter [55] for counts of medium and larger ions.

Mobility measurements by means of divided condensers are described in the literature [18; 19; 29; 54, pp. 179 ff.]; a differential method of very great resolving power had been proposed by Benndorf (*e.g.*, see [54, 56, 57]). Recording devices for conductivity measurements are also described in the literature [82, 108, 115, 116, 138]; recording instruments for counting ions have been devised by Nordmann [94–96], Leckie [82], Langevin [81], Hogg [52, 53], and others.

Schering's method [110, 111] for recording conductivity, which is still used occasionally, is somewhat different. A wire from 10 to 20 m long is freely suspended and surrounded by a cylindrical wire net having a radius of approximately 1 m. The wire is charged at certain time intervals and its voltage drop is recorded, for instance, by a Benndorf electrometer.

Rates of Ion Formation and Recombination; Mean Life of Ions. Under conditions of equilibrium, the rate of ion production q and the numbers n, N, and N_0 of the small ions, large ions, and uncharged suspensions, respectively, have the following relationship:

$$q = \alpha n^2 + \eta_1 nN + \eta_2 nN_0 + \eta_3 NN_0 + \eta_4 N^2 \qquad (8)$$

where

α = recombination coefficient between small ions,

η_1 = recombination coefficient between small and large ions,

η_2 = recombination coefficient between small ions and uncharged particles,

η_3 = recombination coefficient between large ions and uncharged particles,

η_4 = recombination coefficient between large ions.

The last two terms of equation (8) are insignificant as compared to the others, because η_3 and η_4 are smaller than the other coefficients by several orders of magnitude.

In order to determine the individual recombination coefficients, synchronous measurements of all participating constituents are necessary [92, 93]. Owing to the prerequisite of ionization equilibrium, such meas-

urements are very difficult [60, 63]. However, according to von Schweidler [121–123], the following visualizable approximation can be made: Under natural conditions near the ground where moderate to high values of N and N_0 prevail, we can write instead of equation (8)

$$q = \alpha n^2 + \beta n = \beta' n, \tag{9}$$

since the quadratic term becomes less important in comparison to the linear one, as the concentration of large ions in the air becomes greater. The coefficient β', which has the unit of sec^{-1}, is designated as the *vanishing constant*. Its reciprocal then is the *mean life of small ions*, in analogy to radioactive phenomena.

In practice, the air is introduced into an ionization chamber (condenser, sealed on all sides). By applying a high voltage, the value of n present is determined, and then the characteristic current-potential relationship is recorded. The values for q and n are determined by means of Method I (ohmic current), or q and β are found by Method II (semisaturation current); for more details, see the references previously cited.

Condensation Nuclei and Dust Particles. Aside from the hydrometeors (fog, clouds, and precipitation), the atmospheric content of suspended particles is somewhat arbitrarily divided into condensation nuclei and dust particles. Condensation nuclei, hygroscopic particles whose radius is approximately 10^{-5} cm or less, are counted by producing condensation upon them in supersaturated air and determining the number of resulting droplets. Two types of such nuclei counters are well known. One was developed by Aitken [1] and the other by Scholz [117, 118]. For details, see these papers as well as others [23, 71, 80].

The measurement of dust particles is made by means of the Owens counter [9] and the Konimeter.[3]

The Electric Field of the Atmosphere

If an uncharged conductor of any given shape is introduced into the earth's field, a separation of electric charges is found on this conductor, due to electrostatic induction. The conductor assumes the potential V_L, which is the potential of a given equipotential layer in the earth's field. This layer intersects orthogonally the electrically neutral line nn of the conductor's surface. It is the potential of the reference point B (see Fig. 7). The following possibilities exist for measurements of the electric field.

1. The body is temporarily grounded in the position shown in Fig. 7. Under such conditions it assumes the zero potential of the earth and takes on a charge $Q = -CV_L$ (where C = capacitance) from which, when it is introduced into a field-free space ("shielding"), the difference in potential of V_L with respect to the ground can be determined (electrostatic induction method). Variations of the method illustrated in Fig. 7 include:

a. The body is grounded, then insulated, and transferred to another point in the field. In this way its electrostatic induction is changed. Measurement of this "free induction charge" furnishes a measure of the difference in potential of the reference points of both positions. (For the principle of the movable conductor see [2; 7; 108; 137, p. 182]).

b. If after temporary grounding the body remains in position, its changes in electrostatic induction give a measure of the variation of the field; by means of high-ohmic leaks, the arrangement is converted into a field variometer [59].

c. If a metal plate is mounted flush with the earth's surface, grounded and exposed to the field, and thereupon shielded against the latter in an insulated state, it furnishes the surface density of the electrostatic inductance charge and thus the field intensity at the ground level (Wilson test-plate [146–149]). Rhythmical exposure to and shielding from the field produces an a-c current [87, 89–91, 124].

2. If provisions are made for the removal of the induction charge at a point P not situated on the neutral line of the conductor, a new neutral line is produced to which a different equipotential surface V_R

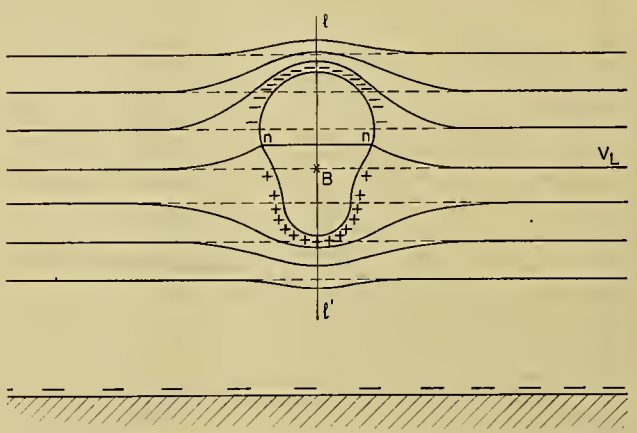


FIG. 7.—Electric field conditions produced by an uncharged conductor.

orthogonally connects (see Fig. 8). A connected measuring instrument indicates the potential of the reference point R with respect to the ground. Discharge of the electrostatic induction is attained by so-called *collectors* as enumerated herewith:

a. Point collector, based on the point discharge flow (no longer in use).

b. Flame collector, utilizing the ionization of the gases of combustion to conduct the charge away; a variant is the glow collector.

c. Water-dropper collector, operating by capacitive charge separation and discharge.

d. Radioactive collector, which provides for discharge by ionization of its environment.

The discharge proceeds according to an exponential law:

$$U_t = U_0 e^{-\kappa t/C}, \tag{10}$$

where U_t is the potential difference at the time t, U_0 the potential difference at the time zero, C the capacitance, and κ the discharge constant. The discharge constant κ or its reciprocal, the "apparent" or "transition"

3. Described in the Zeiss catalogue, Jena.

resistance, serves as an index of the efficiency of a collector. The value of κ fluctuates from approximately

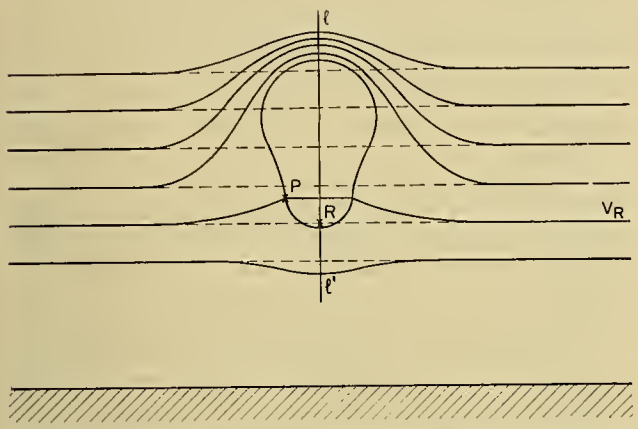


FIG. 8.—Electric field conditions produced by a collector.

unity for glow collectors to 50–100 $\times$ 10^{-12} for highly radioactive collectors.

The "external resistance," in other words, the resistance of the insulation of the collector from the ground, must be great compared to the transition resistance of the collector, as otherwise its readings become inaccurate. Attempts to operate the collector "short-circuited" were found to be subject to disturbances, for example, by the wind, and should therefore be avoided [77].

Exact field measurements can only be performed by means of the Wilson test-plate on completely level terrain. Any other type of arrangement disturbs the field and yields only more or less acceptable approximations. The so-called technique of *reduction to the free plane* can only be used as an approximation [16]. For this reason a comparison of "absolute values" of the field will always remain somewhat problematical. Therefore, consideration of the relative periodic and aperiodic variabilities of the electric field is fundamentally more important.

With respect to the selection of "undisturbed days" in the treatment of recorded data, see, among other sources, Israël and Lahmeyer [72].

The Vertical Current

The difficulties involved in measuring and recording the vertical electric current stem from the generally minute current density of about 10^{-12} amp m^{-2} and the disturbing effects of the electrostatic inductance of the field or of its changes on the receiver system. Galvanometric methods are applicable only to measurements of vertical currents intensified by thunderstorms. Therefore, accumulation methods (accumulation of inflow of charges over a given time) are used in most cases (Ebert [34], Simpson [128], and Wilson [146–149]). The disadvantage of these methods is that in the accumulation period the potential of the collecting body changes somewhat. Simpson circumvents this disadvantage by continuous draining of the charges by means of a water-dropper collector, whereas Wilson obviates the difficulty by a compensation method (see Fig. 9). Methods for continuous recording are described by Scrase [125] (accumulation method, quadrant electrometer) as shown in Fig. 10, and by Kasemir [77] (direct recording by use of a d-c amplifier). Scrase compensates for the effects of field fluctuation by using a supplementary field collector connected to alternate quadrants in the quadrant electrometer, whereas Kasemir largely suppresses these effects by increasing the capacitance. The vertical current can also be computed from the potential gradient and the conductivity, according to equation (6).

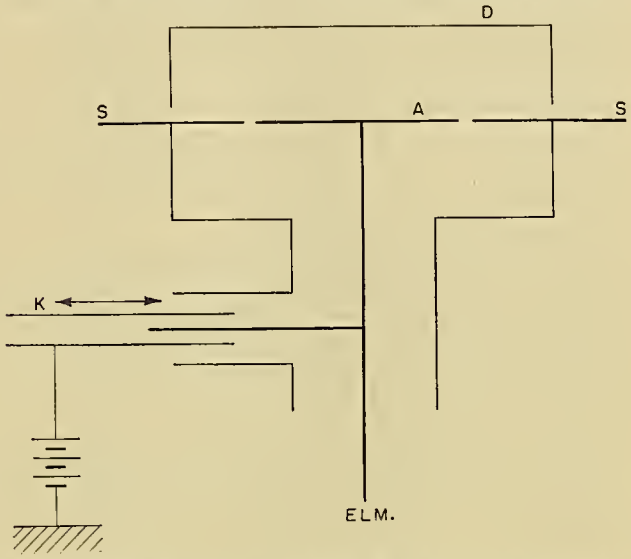


FIG. 9.—Diagram of the Wilson test-plate method (*K* is a variable condenser and *ELM* is the electrometer).

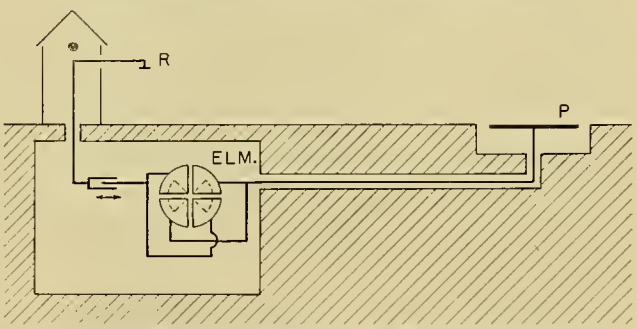


FIG. 10. Schematic diagram of the vertical-current recorder after Scrase (*R* is the radioactive collector, *P* is the test plate, and *ELM* is the quadrant electrometer).

The Space Charge

Measurement or recording of the space charge is made by means of the following methods: The *cage method* consisting of the measurement of the potential difference between the wall surface and the center of a wire cage whose volume ranges from approximately one to several cubic meters [75, 76]; the method of the *change of the potential gradient with altitude* according to Poisson's equation [13, 97, 145]; Obolenski's *filter method* [102]; or the method of *synchronous counts of positive and negative ions* [53, 82]. The cage method is disturbed in most cases by Volta effects [15].

Investigations of Thunderstorm Electricity

Field. Measurements of atmospheric electricity in conjunction with electrical storms require a very rapid response of the equipment. This requirement eliminates collector measurements. The measurements are based on the application of electrostatic induction techniques of field measurement (Wilson test-plate, Wilson elevated sphere, mechanical collectors, antennas) in combination with high-speed recording instruments including cathode-ray oscillographs.

Current. The vertical current can be determined with the aid of the Wilson test-plate in connection with a galvanometer or a capillary electrometer [148] or by recording the current through a point collector mounted in an exposed position [143, 150].

Precipitation Charge. Precipitation is collected in an insulated vessel. Splashing of the drops is prevented by lining the bottom with velvet, brushes, etc. Records are obtained either by the accumulation technique or by the "current-circuit method." For charge measurements on individual drops see Chalmers and Pasquill [25], Gschwend [43], and Gunn [44].

Investigations of Lightning Discharges.[4] The following methods can be employed: The optical method involves photography, using, for example, the Boys camera [20, 21].

In the electrical method the Klydonograph is used for the direct measurement of peak voltages in the lightning discharge by means of Lichtenberg figures; the measuring range is approximately 2–18 kv [83].

The magnetic method is based on the magnetization of small steel rods by currents flowing through lightning rods of various types [38] as, for example, the Fulchronograph [140, 141], or on the measurement of the magnetic field of the lightning discharge channel, and on numerous other special methods some of which are used in combination with high-tension lines [84, 99, 100, 144].

Radio Interferences. The electromagnetic pulses (atmospheric) originating from electric discharges in the atmosphere are investigated with respect to their number, direction of incidence, place of origin, and pattern. Excitation of an oscillating circuit coupled to a nondirectional antenna gives the number; use of rotating loop antennas gives the directional distribution; ranging with cathode-ray direction finders furnishes a bearing on the point of origin; finally, the pattern of the disturbance is established by use of aperiodic d-c amplifiers. For literature of the foregoing see: Appleton and collaborators [4, 5], Bureau [24], Lugeon [88], Norinder [98, 101], Schindelhauer [112, 113], and numerous others.

Measurements in the Free Atmosphere

The methods used for determination of the individual factors of electricity in the free atmosphere are fundamentally the same as those used at ground level. They must, however, be properly modified to allow for the special operational conditions obtaining in either the free-flying or captive carriers of the measuring instruments, such as manned free balloons, gliders, motor-driven aircraft, dirigibles and blimps, automatic recording balloons, captive balloons, and kites. For a summary of older and more recent aerological methods of atmospheric electricity see Israël [65].

Problems of Present-Day Research in Atmospheric Electricity

In concluding this discussion of equipment and methods, a few ideas may be presented on the most urgent problems of modern research in atmospheric electricity.

Although measurements have been made for almost two hundred years, the phenomena of atmospheric electricity have been considered a discrete part of the physics of the atmosphere; it is only recently that certain fundamental relationships to other atmospheric phenomena have been uncovered. Through the gradual detection of the inner relationships between atmospheric-electric and meteorological processes, new light is being shed on many of the electric phenomena that hitherto have been unexplained. Moreover, the possibility arises of immediate application of the knowledge gained in atmospheric electricity to meteorology and aerology.

As has recently been pointed out [68], the fundamental concepts of atmospheric-electric phenomena have undergone a mutation in that the tendency for segregation of electric from meteorological processes is slowly disappearing and is giving way to a trend toward correlating them. The stimulus for this development has been primarily the knowledge gained through the ocean expeditions sponsored by the Carnegie Institution. These cruises have revealed the world-wide synchronous component of the electric field, the coupling of this field with the world-wide weather, and the increasingly clear relationships between the local variations and the vertical mass exchange.

This last correlation, in particular, may furnish the key to the major portion of the relationship between atmospheric-electric and meteorological phenomena and, thus, simultaneously indicate the direction in which further research should proceed.

A primary requirement is the extension of the measurements to more than a single atmospheric-electric element, because the prevalent practice of recording the potential gradient alone permits only very limited conclusions. The three fundamental quantities of field intensity (potential gradient) E, conductivity Λ, and vertical current density i, are interrelated in Ohm's law:

$$E\Lambda = i.$$

Any statements regarding processes taking place in the atmospheric-electric circuit under equilibrium conditions require that two of these fundamental quantities be known. If we are to include the nonstationary ("switching-on") processes, all three quantities must be measured. It would be most desirable if this practice, now followed by the large atmospheric-electric

4. Consult "The Lightning Discharge" by J. H. Hagenguth, pp. 136–143 in this Compendium.

observatories, were introduced everywhere. (For details see [62, 72].)

For investigations of the effect of austausch on atmospheric-electric conditions, a simplified case must first be attacked by means of recordings made at altitudes as high as possible. Such investigations, conducted at high mountain observatories, promise freedom from changes in the aerosol that, in the lower atmospheric layers, are the result of the diurnal variation of the vertical mass exchange. We may expect that the atmospheric-electric processes in their entirety are composed of the interaction of low-level phenomena, which proceed according to local time and are caused by austausch variations, and of the world-wide synchronous processes at higher levels. We may also expect the transition to occur at an altitude of a few kilometers [61, 68]. For preliminary results of pertinent investigations see [70].

The following experiment appears to be an additional promising step in this direction. A program could be set up to obtain simultaneous records of atmospheric-electric elements at neighboring stations located at different altitudes. This would offer the possibility of observing the gradual upward penetration of the austausch effect [67].

In this connection, further investigation of the diurnal variations of atmospheric-electric elements in various air masses [64, 66] can be expected to furnish criteria of the degree of stability or instability of atmospheric stratification.

Research in the vicinity of "generators," that is, in the region of thunderstorms, precipitation, and clouds, offers special problems; see, for example, the work by Simpson [129, 130].

Without doubt, the greatest problem of atmospheric electricity is its systematic extension into the third dimension, that is, the development of an atmospheric-electric aerology with *regular* determinations of the conditions existing in the free atmosphere [65]. Recent developments of special methods of measurement suitable for this purpose, such as those by Simpson [131, 132], Rossmann [107], Gunn [44, 45], and Belin [10], furnish the practical means for this extension.

The exploration of the origin and propagation of high-frequency disturbances in the atmosphere (sferics) can materially aid weather reconnaissance and analysis [4, 5, 88, 113] and can be developed into an integrating component of meteorological practice.

REFERENCES

I. Comprehensive discussions of methods and instruments for measuring atmospheric electricity.

Israël, H., "Luftelektrizität, Grundlagen und Messmethoden," *Meteorologisches Taschenbuch*, F. Linke, Hsgbr., Bd. II, SS. 110–175. Leipzig, 1933. Zweite erweiterte Aufl., Bd. V, SS. 274–351, 1939.

Kähler, K., "Luftelektrizität," *Handbuch der meteorologischen Instrumente*, E. Kleinschmidt, Hsgbr., SS. 605–666. Berlin, J. Springer, 1935. "Staubgehalt der Luft und Kondensationskerne." *Ibid.*, SS. 248–262.

Torreson, O. W., "Instruments Used in Observations of Atmospheric Electricity" in *Terrestrial Magnetism and Electricity* (*Physics of the Earth, VIII*), J. A. Fleming, ed., pp. 231–269. New York, McGraw, 1939; Dover, 1949.

II. References cited in text.

1. Aitken, J., *Collected Scientific Papers*, edited by C. G. Knott. Cambridge, University Press, 1923.
2. Angenheister, G., "Die luftelektrischen Beobachtungen am Samoa-Observatorium, 1912/13." *Nachr. Ges. Wiss. Göttingen*, SS. 191–206 (1914).
3. Angerer, E. v., *Technische Kunstgriffe bei physikalischen Untersuchungen*. Braunschweig, F. Vieweg & Sohn, 1944.
4. Appleton, E. V., Watson-Watt, R. A., and Herd, J. F., "On the Nature of Atmospherics—III." *Proc. roy. Soc.*, (A) 111:654–677 (1926).
5. Appleton, E. V., and Chapman, F. W., "On the Nature of Atmospherics—IV." *Proc. roy. Soc.*, (A) 158:1–22 (1937).
6. Baranow, W. J., und Stschepotjewa, E. S., "Über die Anwendung des Ebertschen Ionenzählers zur Bestimmung der Zahl und der Beweglichkeit der kleinen Ionen in der Atmosphäre." *Phys. Z.*, 29:741–750 (1928).
7. Bauer, L. A., and Swann, W. F. G., "Results of Atmospheric-Electric Observations Made aboard the *Galilee* (1907–08) and the *Carnegie* (1909–16)." *Carnegie Instn. Wash. Publ.*, 3:361–447 (1917).
8. Becker, A., "Neuer Zylinderkondensator zur Untersuchung leitender Gase." *Z. InstrumKde.*, 29:258–261 (1909).
9. Běhounek, F., "Der Gehalt der Luft an Ionen und Staub bei Klimaanlagen." *Gesundheitsing.*, 62:249–253 (1939).
10. Belin, R. E., "A Radiosonde Method for Atmospheric Potential Gradient Measurements." *Proc. phys. Soc. Lond.*, 60:381–387 (1948).
11. Benndorf, H., "Beiträge zur Kenntnis der atmosphärischen Elektrizität. Nr. 10—Über ein mechanisch registrierendes Elektrometer für luftelektrische Messungen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 111:487–512 (1902).
12. —— "Über ein mekanisch registrierendes Elektrometer für luftelektrische Messungen." *Phys. Z.*, 7:98–101 (1906).
13. —— "Beiträge zur Kenntnis der atmosphärischen Elektrizität. Nr. 33—Zur Theorie luftelektrischer Registrierungen I." *S. B. Akad. Wiss. Wien*, Abt. IIa, 118: 1163–1195 (1909).
14. —— "Zuschrift an die Schriftleitung." *Phys. Z.*, 25:60 (1924).
15. —— "Zur Raumladungsmessung in der freien Atmosphäre." *Phys. Z.*, 27:576–578 (1926).
16. —— "Beiträge zur Kenntnis der atmosphärischen Elektrizität. Nr. 68—Grundzüge einer Theorie des elektrischen Feldes der Erde, II." *S. B. Akad. Wiss. Wien*, Abt. IIa, 136:190–194 (1927).
17. Bergwitz, K., "Beiträge zur Gammastrahlung des Erdkörpers." *Elster-Geitel-Festschrift*. Braunschweig, 1915. (See pp. 585–600)
18. Blackwood, O., "The Existence of Homogeneous Groups of Large Ions." *Phys. Rev.*, (2) 16:85–101 (1920).
19. Boylan, R. K., "The Mobilities of Atmospheric Large Ions." *Proc. R. Irish Acad.*, (A) 40:76–84 (1931).
20. Boys, C. V., "Progressive Lightning." *Nature*, 118:749–750 (1926).
21. —— "Progressive Lightning." *Nature*, 124:54–55 (1929).
22. Brentano, J., "Der Gebrauch von Verstärkerröhren zur Messung kleiner Energiebeträge." *Z. Phys.*, 54:571–581 (1929).
23. Burckhardt, H., und Flohn, H., *Die atmosphärischen Kondensationskerne*. Berlin, J. Springer, 1939.

24. Bureau, R., *Répertoires des notes préliminaires du L. N. R.* (Lab. Nat. de Radioélectr.). Paris.
25. Chalmers, J. A., and Pasquill, F., "The Electric Charges on Single Raindrops and Snowflakes." *Proc. phys. Soc. Lond.*, 50:1–16 (1938).
26. Compton, A. H., and Compton, K. T., "An Addition to the Theory of the Quadrant Electrometer." *Phys. Rev.*, 13:288 (1919).
27. —— "A Sensitive Modification of the Quadrant Electrometer: Its Theory and Use." *Phys. Rev.*, 14:85–98 (1919).
28. Custers, J. F. H., "Eine evakuierte Verstärkeranordnung zur Messung kleiner Photoströme." *Z. tech. Phys.*, 14:154–157 (1933).
29. Dessauer, F., Hsgbr., *Zehn Jahre Forschung auf dem physikalisch-medizinischen Grenzgebiet.* Leipzig. G. Thieme, 1931.
30. Dolezalek, F., "Ueber ein einfaches und empfindliches Quadrantenelektrometer." *Z. InstrumKde.*, 21:345–350 (1901).
31. —— "Über ein hochempfindliches Zeigerelektrometer." *Z. InstrumKde.*, 26:292–295 (1906).
32. Ebert, H., "Aspirationsapparat zur Bestimmung des Ionengehaltes der Atmosphäre." *Phys. Z.*, 2:662–664 (1901).
33. —— *Arch. Sci. phys. nat. de Genève*, 12:97 (1901).
34. —— "Galvanometrische Messung des elektrischen Ausgleichs zwischen den Ionenladungen der Atmosphäre und der Ladung der Erdoberfläche." *Phys. Z.*, 3:338–339 (1902).
35. —— "Eine neue Form Ionen-Aspirations-Apparates." *Verh. dtsch. phys. Ges.*, 3:34–37 (1905).
36. Ehmert, A., (in preparation).
37. Elster, J., und Geitel, H., "Ein transportables Quadrantelektrometer mit photographischer Registrierung." *Z. InstrumKde.*, 26:322–323 (1906).
38. Foust, C. M., and Kuehni, H. P., "The Surge-Crest Ammeter." *Gen. elect. Rev.*, 35:644–648 (1932).
39. Gerdien, H., "Die absolute Messung der elektrischen Leitfähigkeit und der spezifischen Ionengeschwindigkeit in der Atmosphäre." *Phys. Z.*, 4:632–635 (1903).
40. —— "Messungen der elektrischen Leitfähigkeit der freien Atmosphäre bei 4 Ballonfahrten." *Nachr. Ges. Wiss. Göttingen*, SS. 383–399 (1903).
41. Gish, O. H., "Systematic Errors in Measurements of Ionic Content and the Conductivity of the Air." *Beitr. Geophys.*, 35:1–5 (1932).
42. Graziadei, H. T., "Studie über die Methodik der Ionenzählung." *Phys. Z.*, 34:82–88 (1933).
43. Gschwend, P., "Beobachtungen über die elektrischen Ladungen einzelner Regentropfen und Schneeflocken." *Jb. Radioakt.*, 17:62–79 (1920).
44. Gunn, R., "The Electrical Charge on Precipitation at Various Altitudes and Its Relation to Thunderstorms." *Phys. Rev.*, 71:181–186 (1947).
45. —— "Electronic Apparatus for the Determination of the Physical Properties of Freely Falling Raindrops." *Rev. sci. Instrum.*, 20:291–296 (1949).
46. Hippel, A. v., *Die Elektronenröhre in der Messtechnik.* Leipzig, 1924.
47. Hoar, T. P., "The Use of Triode and Tetrode Valves for the Measurement of Small D. C. Potential Differences." *Wireless Engr.*, 10:19–25 (1933).
48. Hoffmann, G., "Über ein hochempfindliches Elektrometer und den hiermit möglichen direkten Nachweis der Ionisation des einzelnen α-Teilchens." *Phys. Z.*, 13:480–485 (1912).
49. —— "Über ein Elektrometer hoher Empfindlichkeit." *Ann. Phys., Lpz.*, (4) 42:1196–1220 (1913).
50. —— "Über die Bedeutung der Labilisierung bei der elektrometrischen Messung kleiner Elektrizitätsmengen." *Phys. Z.*, 25:6–8 (1924).
51. —— "Die Bestimmung von sehr kleinen Leitfähigkeiten mit dem Vakuumelektrometer." *Phys. Z.*, 26:913–914 (1925).
52. Hogg, A. R., "Atmospheric Electric Observations." *Beitr. Geophys.*, 41: 1–31 (1934).
53. —— "Some Observations of the Average Life of Small Ions and Atmospheric Ionisation Equilibria." *Beitr. Geophys.*, 41:32–55 (1934).
54. Israël, H., "Zur Theorie und Methodik der Grössenbestimmung von Luftionen." *Beitr. Geophys.*, 31:171–216 (1931).
55. —— "Schwere Ionen der Atmosphäre." *Z. Geophys.*, 7:127–133 (1931).
56. —— "Zum Problem der Randstörungen bei Ionenmessungen." *Beitr. Geophys.* 35: 341–348 (1932).
57. —— "Zur Theorie und Methodik der Grössenbestimmung von Luftionen." *Beitr. Geophys.*, 36:24–37 (1932).
58. —— "Der Wegersche Kleinionen-Aspirator als selbständiges Messgerät." *Meteor. Z.*, 54:487–488 (1937).
59. —— "Zur Methodik der luftelektrischen Messungen. I—Die Feinstruktur des luftelektrischen Feldes und der Vertikalstrom." *Beitr. Geophys.*, 55:314–333 (1939). (Correction in *Beitr. Geophys.*, 56:228 (1940).)
60. —— "Ionen und Kerne; eine kritische Studie." *Beitr. Geophys.*, 57:261–282 (1941).
61. —— "Der Elektrizitätshaushalt der Erdatmosphäre." *Naturwissenschaften*, 29:700–706 (1941).
62. —— "Die halbtägige Welle des luftelektrischen Potentialgefälles." *Arch. Meteor. Geophys. Biokl.*, (A) 1:247–251 (1948).
63. —— "Zum Ionisationsgleichgewicht in atmosphärischer Luft." *Meteor. Rdsch.*, 1:344 (1948).
64. —— "Zum Tagesgang des luftelektrischen Potentialgefälles." *Meteor. Rdsch.*, 1:200–204 (1948).
65. —— "Zur Methodik luftelektrischer Messungen. III—Aerologische Messmethoden." *Wiss. Arb. dtsch. meteor. Dienst. i. franz. BesGeb.*, 2:20–31 (1949).
66. —— "Luftelektrische Tagesgänge und Luftkörper." *J. atmos. terr. Phys.* 1:26–31 (1950).
67. —— "Probleme und Aufgaben der Luftelektrizität." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 12, Ss. 75–77 (1950).
68. —— "Zur Entwicklung der luftelektrischen Grundanschauungen." *Arch. Meteor. Geophys. Biokl.*, (A) 3:1–16 (1950).
69. —— "Sprunghafte Änderungen des luftelektrischen Feldes und Gewitter." *Meteor. Z.*, 61:1–8 (1944).
70. —— Kasemir, H. W., und Wienert, K., "Die luftelektrischen Verhältnisse am Jungfraujoch." *Arch. Meteor. Geophys. Biokl.*, (A) (1951) (in press).
71. Israël, H., und Krestan, M., "Zur Methodik luftelektrischer Messungen. II—Die Zählung der Kondensationskerne." *Beitr. Geophys.*, 58:73–94 (1941).
72. Israël, H., und Lahmeyer, G., "Studien über das atmosphärische Potentialgefälle. I—Das Auswahlprinzip der luftelektrisch 'ungestörten Tage.'" *Terr. Magn. atmos. Elect.*, 53:373–386 (1948).
73. Itiwara, Y., "Zur Methodik der Ionenzählung in der freien Atmosphäre." *Phys. Z.*, 32:97–106 (1931).
74. Jaeger, R., und Kussmann, A., "Über Gleichstromverstärkung, ihre Anwendung zu Messzwecken und ihre Grenzen." *Phys. Z.*, 28:645–651 (1927).

75. Kähler, K., "Über die Schwankung der elektrischen Raumladung in der Atmosphäre." *Meteor. Z.*, 40:204–211 (1923).
76. —— und Dorno, C., "Messungen der elektrischen Raumladung der Atmosphäre in Davos." *Meteor. Z.*, 42:434–439 (1925).
77. Kasemir, H. W., "An Apparatus for Simultaneous Registrations of Potential Gradient and Air-Earth Current." *J. atmos. terr. Phys.*, (1951) (in press).
78. Kolhörster, W., "Die experimentellen Grundlagen der Messung der durchdringenden Strahlung." *Z. InstrumKde.*, 44:333–349 (1924).
79. —— "Apparat zur Messung der durchdringenden Strahlung." *Phys. Z.*, 27:62–63 (1926).
80. Landsberg, H., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, (Supp.) 3: 155–252 (1938).
81. Langevin, P., et Moulin, M., "Sur un enregistreur des ions de l'atmosphère." *C. R. Acad. Sci., Paris.*, 140:305–307 (1905).
82. Leckie, A. J., "Luftelektrische Messungen am Bosschalaboratorium der Technischen Hochschule in Bandoeng, Java." *Beitr. Geophys.*, 52:280–333 (1938).
83. Lewis, W. W., and Foust, C. M., "Lightning Investigations on Transmission Lines." *Gen. elect. Rev.* 33:185–198 (1930).
84. —— "Lightning Surges on Transmission Lines—Natural Lightning." *Gen. elect. Rev.*, 39:543–555 (1936).
85. Lindemann, F. A., Lindemann, A. F., and Keeley, T. C., "A New Form of Electrometer." *Phil. Mag.*, (6) 47:577–583 (1924).
86. Linke, F., Hsgbr., *Meteorologisches Taschenbuch*, Bd. IV. Leipzig, Akad. Verlagsges., 1939. (See Table 144, p. 262)
87. Lueder, H., "Elektrische Registrierung von heranziehenden Gewittern und die Feinstruktur des luftelektrischen Gewitterfeldes." *Meteor. Z.*, 60:340–351 (1943).
88. Lugeon, J., "Sur la nécessité d'une station polaire permanente d'observations radiométéorologiques pour les services de prévision du temps." Inst. Nat. Météorol. de Pologne, Varsovie 1935, 96 pp. (and elsewhere).
89. Matthias, A., "Fortschritte in der Aufklärung der Gewittereinflüsse auf Leitungsanlagen." *Elektrizitätswirtschoft*, 25:297–308 (1926).
90. —— "Bisherige Ergebnisse der Gewitterforschung der Studiengesellschaft für Höchstspannungsanlagen." *Elektrizitätswirtschaft*, 26:2-18 (1927).
91. Meinhold. H.. "Allgemeine Untersuchungen über das luftelektrische Feld und Anwendung auf spezielle Gebiete der elektrischen Nachrichtentechnik." *Zentr. f. wiss. Ber.*, (1944).
92. Nolan, J. J., and Enright, J., "Experiments on Large Ions in Air." *Proc. R. Irish Acad.*, (A) 36:93–114 (1923).
93. Nolan, J. J., Boylan, R. K., and de Sachy, G. P., "The Equilibrium of Ionisation in the Atmosphere." *Proc. R. Irish Acad.*, (A) 37:1–12 (1925).
94. Nordmann, C., "Méthode pour l'enregistrement continu de l'état d'ionisation des gaz. Ionographe." *C. R. Acad. Sci., Paris*, 138:1418–1420 (1904); "Enregistrement continu de l'ionisation gazeuse et de la radioactivité par les méthodes de déperdition." *Ibid.*, 138:1596–1599 (1904).
95. —— "Enregisteur à écoulement liquide de l'ionisation atmosphérique." *C. R. Acad. Sci., Paris*, 140:430–433 (1905).
96. —— "Sur certaines expériences relatives à l'ionisation de l'atmosphère, exécutées en Algérie à l'occasion de l'éclipse totale du 30 août 1905." *C. R. Acad. Sci., Paris.*, 141:945–948 (1905).
97. Norinder, H., "Researches on the Height Variation of the Atmospheric Electric Potential Gradient in the Lowest Layers of the Air." *Geogr. Ann., Stockh.*, 3:1–96 (1921).
98. —— "Cathode-Ray Oscillographic Investigations on Atmospherics." *Proc. Inst. Radio Engrs., N. Y.*, 24:287–304 (1936).
99. —— "Rapid Variations in the Magnetic Field Produced by Lightning Discharges." *Proc. phys. Soc. Lond.*, 49: 364–375 (1937).
100. —— "Some Aspects and Recent Results of Electromagnetic Effects of Thunderstorms, Parts I and II." *J. Franklin Inst.*, 244:109–130; 167–207 (1947).
101. —— and Stoffregen, W., "The Nature and Variation of Atmospherics Caused by Lightning Discharges." *Ark. Mat. Astr. Fys.*, Vol. 33A, No. 16, 44 pp. (1946).
102. Obolensky, W. N., "Über elektrische Ladungen in der Atmosphäre." *Ann. Phys., Lpz.*, (4) 77:644–666 (1925).
103. Palmaer, W., "Über das absolute Potential der Kalomelelektrode." *Z. phys. Chem.*, 59:129–191 (1907). (See p. 187)
104. Perucca, E., "Ein neues Elektrometer." *Z. InstrumKde.*, 47:524–527 (1927).
105. —— und Leiss, C., "Neue Konstruktion des Peruccaschen Elektrometers." *Z. Phys.*, 49:604–607 (1928).
106. Rasmussen, E., "Über Gleichstromverstärkung." *Ann. Phys., Lpz.*, (5) 2:357–380 (1929).
107. Rossmann, F., "Luftelektrische Messungen mittels Segelflugzeugen." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 15, 54 SS. (1950).
108. Russeltvelt, N., "Instrumente und Apparate für die luftelektrischen Untersuchungen an dem meteorologischen Observatorium in Ås." *Beihefte d. norweg. Meteor. Inst. f. 1925.* Oslo (1926).
109. Rutgers van der Loeff, M., *De Ionen en de Ionisatiebalans in de Atmospheer.* Dissert., 120 pp., Amsterdam, 1938.
110. Schering, H., "Der Elster-Geitelsche Zerstreuungsapparat und ein Versuch quantitativer absoluter Zerstreuungsmessung." *Ann. Phys., Lpz.*, (4) 20:174–195 (1906).
111. —— "Registrierungen des spezifischen Leitvermögens der atmosphärischen Luft." *Nachr. Ges. Wiss. Göttingen*, SS. 201–218 (1908).
112. Schindelhauer, F., "Die Luftstörungen der drahtlosen Telegraphie." *Wiss. Abh. Reich. Wetterd.*, Bd. 3, Nr. 5 (1937).
113. —— und Israël, H., "Die Peilung von Luftstörungen der drahtlosen Telegraphie zum Zwecke der Wettererkundung." *Forsch. ErfahrBer. Reichs. Wetterd.*, Reihe B. Nr. 12, 40 SS. (1944).
114. Schintelmeister, F., *Die Elektronenröhre als physikalisches Messgerät.* Wien, J. Springer, 1943.
115. Scholz, J., "Über die Messmethoden der elektrischen Leitfähigkeit der Atmosphäre." *Phys. Z.*, 32:130–139 (1931).
116. —— "Über die Verwendung von Selenzellen zur Registrierung des Potentialabfalles von Quadrantelektrometern." *Phys. Z.*, 29:702–705 (1928).
117. —— "Ein neuer Apparat zur Bestimmung der Zahl der geladenen und ungeladenen Kerne." *Z. InstrumKde.*, 51:505–522 (1931).
118. —— "Vereinfachter Bau eines Kernzählers." *Meteor. Z.*, 49:381–388 (1932).
119. Schultze, H., "Ein neues Quadrantenelektrometer für dynamische Messungen." *Z. InstrumKde.*, 27:65–75 (1907).

120. Schweidler, E. v., "Über die Ionisierung in einem geschlossenen Gefässe infolge der Eigenstrahlung der Wand." *Phys. Z.*, 15:685–688 (1914).
121. —— "Beiträge zur Kenntnis der atmosphärischen Elektrizität, Nr. 59. Über das Gleichgewicht zwischen ionenerzeugenden und ionenvernichtenden Vorgängen in der Atmosphäre." *S. B. Akad. Wiss. Wien*, Abt. IIa, 127:953–967 (1918).
122. —— "Beiträge zur Kenntnis der atmosphärischen Elektrizität, Nr. 60. Über das Gleichgewicht zwischen ionenerzeugenden und ionenvernichtenden Vorgängen in der Atmosphäre, II. Mitt." *S. B. Akad. Wiss. Wien*, Abt. IIa, 128: 947–955 (1919).
123. —— "Über die Charakteristik des Stromes in schwach ionisierten Gasen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 133:23–27 (1924).
124. Schwenkhagen, H., "Elektrostatischer Induktionsspannungsmesser." *Elektrizitätswirtschaft* (1942 oder 1943).
125. Scrase, F. J., "The Air-Earth Current at Kew Observatory." *Geophys. Mem.*, Vol. 7, No. 1 (1933).
126. —— "Some Measurements of the Variation of Potential Gradient with Height near the Ground at Kew Observatory." *Geophys. Mem.*, Vol. 7, No. 10 (1935).
127. Shimizu, T., "A Sensitive Electroscope." *Jap. J. Phys.*, 1:107–111 (1923).
128. Simpson, G. C., "Earth-Air Electric Currents." *Phil. Mag.*, (6) 19:715–725 (1910).
129. —— "Atmospheric Electricity during Disturbed Weather." *Terr. Magn. atmos. Elect.*, 53:27–33 (1948).
130. —— "Atmospheric Electricity during Disturbed Weather." *Geophys. Mem.*, No. 84, 51 pp. (1949).
131. —— and Scrase, F. J., "The Distribution of Electricity in Thunderclouds." *Proc. roy. Soc.*, (A) 161:309–352 (1937).
132. Simpson, G. C., and Robinson, G. D., "The Distribution of Electricity in Thunderclouds, II." *Proc. roy. Soc.*, (A) 177:281–329 (1941).
133. Stechhöfer, S., "Erdstrahlungsmessungen mit dem Geiger-Müller-Zählrohr und elektrische Feldstärkemessungen im Gelände." *Z. Geophys.*, 12:68–86 (1936).
134. Stille, U., "Die Umrechnungsfaktoren von internationalen auf absolute elektrische Einheiten." *Z. Phys.*, 121: 34–53 (1943); "Die atomaren Konstanten e, e/m_0 und h." *Ibid.*, 121:133–200 (1943).
135. —— *Umrechnungstafeln.* Braunschweig, F. Vieweg & Sohn, 1944.
136. Suckstorff, G. A., "Eine transportable Zählrohrapparatur und ihre Anwendung im Gelände." *Z. Geophys.*, 11:95–101 (1935).
137. Swann, W. F. G., "On Certain New Atmospheric-Electric Instruments and Methods." *Terr. Magn. atmos. Elect.*, 19:171–185 (1914).
138. —— "An Apparatus for Automatically Recording the Electrical Conductivity of Air." *Yearb. Carneg. Instn.*, No. 16 (1917). (See pp. 279–281)
139. —— "Note on the Theory of the Single Fiber Electroscope." (Abstract) *Phys. Rev.*, 23:779 (1924).
140. Wagner, C. F., and McCann, G. D., "Lightning Phenomena." *Elect. Engng., N. Y.*, 60:374–384; 438–443; 483–500 (1941).
141. —— and Beck, E., "Field Investigations on Lightning." *Trans. Amer. Inst. elect. Engrs.*, 60:1222–1229 (1941).
142. Weger, N., "Eine Verbesserung der Methodik der Kleinionen-Zählung." *Phys. Z.*, 36:15–20 (1935).
143. Whipple, F. J. W., and Scrase, F. J., "Point Discharge in the Electric Field of the Earth." *Geophys. Mem.*, Vol. 7, No. 11 (1936).
144. Wichmann, H., "Methoden zur Messung des gewitterelektrischen Feldes bei Blitzentladungen." *Beitr. Geophys.*, 58:95–111 (1941).
145. Wigand, A., Schubert, J., und Frankenberger, E., "Ein neues Verfahren der luftelektrischen Raumladungsmessung." *Z. Geophys.*, 6:458–463 (1930).
146. Wilson, C. T. R., "On a Portable Gold-Leaf Electrometer for Low or High Potentials, and Its Application to Measurements in Atmospheric Electricity." *Proc. Camb. phil. Soc.*, 13:184–189 (1905); "On the Measurement of the Earth-Air Current and on the Origin of Atmospheric Electricity." *Ibid.*, 13:363–382 (1906).
147. —— "On the Measurement of the Atmospheric Electric Potential Gradient and the Earth-Air Current." *Proc. roy. Soc.*, (A) 80:537–547 (1908).
148. —— "On Some Determinations of the Sign and Magnitude of Electric Discharges in Lightning Flashes." *Proc. roy. Soc.*, (A) 92:555–574 (1916).
149. —— "Investigations on Lightning Discharges and on the Electric Field of Thunderstorms." *Phil. Trans. roy. Soc. London*, (A) 221:73–115 (1920).
150. Wormell, T. W., "Vertical Electric Currents below Thunderstorms and Showers." *Proc. roy. Soc.*, (A) 127:567–590 (1930).
151. Wulf, T., "Über die in der Atmosphäre vorhandene Strahlung von hoher Durchdringungsfähigkeit." *Phys. Z.*, 10:152–157 (1909).
152. —— "Ein Einfadenelektrometer." *Phys. Z.*, 15:250–254 (1914).

RADIOACTIVITY OF THE ATMOSPHERE*

By H. ISRAËL

Buchau Observatory and University of Tübingen

Introduction

The conductivity of the air is a characteristic of the entire atmosphere at all times. It is the result of the most varied ionizing radiation of a corpuscular and electromagnetic nature. In the lower layers of the atmosphere the radiation of the radioactive substances that are contained in the air and in the uppermost strata of the earth's crust is the principal source of this ion formation.

Table I gives a summary of the three decay series of radioactive substances and the physical qualities of the respective types of atoms. As far as the atmosphere is concerned, our interest begins only with the gaseous intermediate transformation products (*i.e.*, the emanations) and therefore a consideration of all preceding elements has been omitted.

The discovery of radioactive substances in the atmosphere was made by Elster and Geitel [19], who thus gave a physical explanation for the weak conductivity of the air, already known to Coulomb in 1785.

Methods of Measurement

There are two possible methods of measuring the radioactive elements of the atmosphere. One method consists of removing the emanations from the air by adsorption or condensation and measuring them in an ionization chamber (emanometry). In the second method, the property of so-called "inductions" is utilized, that is, the property of some radioactive decay products of carrying a positive electrical charge; these products are deposited on negatively charged collectors where their ionizing effect can be examined. This technique is termed the "induction method." The experimental technique of the induction method is simpler than that of emanometry, but quantitatively less reliable. Therefore, emanometry is to be preferred for analysis of the relatively long-lived *Rn*. However, direct emanometry does not work for *Th*- and *Ac*-products because of the short life of *Tn* and *An*, and only the indirect induction method is possible.

Table I. Radioactive Substances in the Atmosphere

Element	Symbol	Rays	Half-life	Disintegration constant (sec^{-1})
Radon	*Rn*	α	3.825 day	2.097×10^{-6}
Radium A	*RaA*	α	3.05 min	3.78×10^{-3}
Radium B	*RaB*	$\beta + \gamma$	26.8 min	4.31×10^{-4}
Radium C	*RaC*	$\alpha + \beta + \gamma$	19.7 min	5.86×10^{-4}
Radium D	*RaD*	$\beta + \gamma$	22 yr	1.00×10^{-9}
Radium E	*RaE*	$\beta + \gamma$	5.0 day	1.61×10^{-6}
Radium F*	*RaF*	$\alpha(+\beta)$†	140 day	5.73×10^{-8}
Thoron	*Tn*	α	54.5 sec	1.71×10^{-2}
Thorium A	*ThA*	α	0.14 sec	4.95
Thorium B	*ThB*	$\beta + \gamma$	10.6 hr	1.82×10^{-5}
Thorium C	*ThC*	$\alpha + \beta + \gamma$	60.5 min	1.91×10^{-4}
Actinon	*An*	α	3.92 sec	0.177
Actinium A	*AcA*	α	2×10^{-3} sec	347
Actinium B	*AcB*	$\beta + \gamma$	36.0 min	3.21×10^{-4}
Actinium C	*AcC*	$\alpha + \beta + \gamma$	2.16 min	5.35×10^{-3}

* Polonium (*Po*) † uncertain

The gaseous emanations radon (*Rn*), thoron (*Tn*), and actinon (*An*) are emitted from the rocks and the soil into the air entrapped in the ground capillaries or directly into the air at the earth's surface and are transported to higher levels by the apparent diffusion of the vertical atmospheric mass exchange (austausch). Because of the interaction between this exchange and the disintegration rate of the emanations and their by-products, a certain characteristic vertical distribution for each element can be expected which, in the mean, can be proven experimentally (see below). The contribution made by these radioactive elements to the formation of ions in the atmosphere decreases rapidly with altitude and ceases approximately at the height of the tropopause.

Emanometry. The international unit of *Rn* is the *curie* (C). The curie is the quantity of radon which is in radioactive equilibrium with one gram of radium and thus emits as many alpha particles per unit time as does one gram of radium. The current which 1 C can maintain at saturation and with complete utilization of its radiation is 9.22×10^{-4} amp. In the course of approximately 3 hr, equilibrium is reached between

* Translated from the original German.

Rn and its three short-lived decay products, *RaA*, *RaB*, and *RaC*. The current is thereby increased to 20.0×10^{-4} amp.[1] These two "current equivalents" are the basis for emanometry.

A closed cylindrical vessel containing an insulated electrode is generally used as an ionization chamber. Between the walls of the vessel and the electrode, a potential of a few hundred volts is applied which causes the ions produced to deposit before their number changes appreciably by recombination (saturation current). As the ion concentration in the air is small, the measurement is usually made electrometrically by the charge or discharge method shown in Fig. 1.

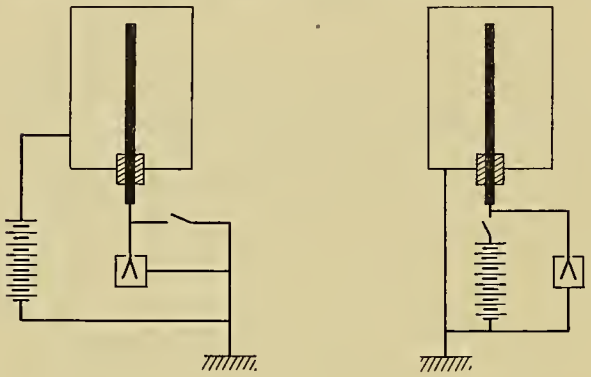


Fig. 1.—Circuit diagram for emanometrical measurements by the charge method (*a*) and by the discharge method (*b*).

Conversion of the measured data into radon units requires various corrections:

1. The "time correction" considers the change of the current equivalent due to formation of *RaA*, *RaB*, and *RaC*. Formerly, measurements were not begun until radioactive equilibrium had been attained between *Rn* and *RaA*, *RaB*, and *RaC*, at the earliest about 3 hr after pure *Rn* had been put into the measuring vessel. This method can now be shortened by using the intermediate values of the current equivalent (Table II).

2. Because of the limitation of the ionization space, not all alpha rays can become fully effective. Great difficulties are encountered in the numerical computation of the correction factor [14, 21, 54]. Therefore, the following empirical corrections developed by Duane and Laborde [16, 17] are used in most cases:

$$J = K\,(1 - 0.517\,O/V)$$
$$J' = K'\,(1 - 0.572\,O/V),$$

where J and J' represent the measured current values for pure *Rn* or for radon in equilibrium with *RaA*, *RaB*, and *RaC*; K and K' are the corrected current values; O is the surface, and V the volume of the ionization vessel.

3. It is difficult to attain saturation in ionization by alpha radiation, because a very strong initial recombination becomes effective in the closely packed ion columns concomitant to their formation. In this case, also, theoretical determination of a satisfactory correction factor is impossible. A generally valid empirical correction cannot be found either, since the saturation deficit is a function of the shape and size of the vessel and of the total quantity of ionizing substances. Therefore, this correction must be determined individually for each type of emanometer for varying quantities of radon. Suggestions for its practical evaluation have been made by Israël [31].

4. The effect of the atmospheric pressure and temperature on emanometrical measurements can be considered according to Lester [36]. In general, however, it remains a minor error compared with the other errors present [6].

Table II. Current Equivalents S_T for Emanometrical Measurements*

t (min)	S_T (10^{-4} amp)	t (min)	S_T (10^{-4} amp)
0	9.22	60	17.25
2	11.03	80	18.24
4	12.21	100	18.91
6	12.97	150	19.78
8	13.48	200	19.98
10	13.80	250	20.00
20	14.69	300	19.95
30	15.34	400	19.80
40	16.00	500	19.61

* Values of S_T are for t minutes after the introduction of pure *Rn* into an ionization chamber, taking into account the disintegration of *Rn*.

These corrections may be avoided if the measurements are compared with those of known quantities of radon. In order to obtain such small, but well-defined *Rn* quantities, the so-called radium standards are used, aqueous solutions of $RaCl_2$ which can be produced easily [8] or which can be obtained commercially. The precautions to be taken when working with such test solutions are discussed elsewhere [8, 29].

Various types of emanometers, especially useful for atmospheric *Rn* measurements, have been developed, for example, by Becker [7], Messerschmidt [42, 43], Janitzky [33], and Israël [28, 31]. Moreover, such devices for measurements in an air stream were developed by Israël [32] and Deij (see [16, 17]). Since the *Rn* concentration in the air is extremely small, a method of enriching it is often used to increase the accuracy. This method utilizes the very great solubility of *Rn* in some organic substances, its ability to condense at the temperature of liquid air, or, most conveniently, its property of being adsorbed by coconut-shell charcoal or activated carbon. By heating the adsorbent to incandescence, the *Rn* will again be given off completely. When large ionization chambers are used at high sensitivity, the "enrichment method" need not be used

1. In this case only 50 per cent of the ionization effect of the alpha rays from *RaA*, *RaB*, and *RaC* is considered, since these inductions deposit on the walls. The beta and gamma radiations from *RaB* and *RaC* constitute less than 1 per cent of the current and are usually not considered. Furthermore, the contribution of *RaD*, *RaE*, and *RaF* is far below the range of accuracy of measurement and may be disregarded. The only manifestation of these substances is the slowly increasing "pollution" of the measuring chamber which can usually be eliminated mechanically.

(lower limit of measurement approximately 2×10^{-17} C cm^{-3}). For further details see Israël [29].

Induction Method. If a charged body is exposed to air containing *Rn*, its surface acquires a certain activity, which becomes greatest with a negative charge. Because of this, it is apparent that the resultant products, principally *RaA*, must be positively charged. According to Eckmann [18], the *RaA* atoms are in fact positively charged immediately after their formation. Hence, they would be useful for an indirect quantitative *Rn* measurement, if they did not soon lose part of their charge by interaction with the ions in the atmosphere. Furthermore, the other by-products also carry, at least in part, electric charges. Finally, it must be considered that these charged inductions move under the influence of the electric field of the atmosphere.

All these possibilities of error greatly reduce the reliability of the indirect method of measurement. This method supplies only relative values which, when obtained under similar working conditions, are more or less comparable with one another. However, the relative values can be correlated with the direct-emission measurements only with very great uncertainty. Even the various attempts at applying corrections, made by Salpeter [49] and Curie [13] did not succeed in eliminating this uncertainty. Only Eve's and Aliverti's modification of the procedure (see below) leads to reliable quantitative results.

The oldest type of a practical technique for carrying out such measurements is that in which a wire is kept at a high negative potential, exposed for several hours, and then wound on a spool and examined in an ionization chamber. The resultant rate of discharge, expressed in volts per hour divided by the length of the wire (in meters), is known as the "activation number." An extensive improvement of this technique has been introduced by Swann [56], who segregated the contributions of various types of inductions by observing the decrease in wire activity with time. Gerdien [22] and Bauer and Swann [4, 5] use an aspiration process for collection. Eve [20] exposes the collecting wire in a large closed vessel, 16 m^3 in volume, and compares the resultant activity with that obtained under otherwise identical conditions, but with a known quantity of *Rn* in the vessel. In Aliverti's method [1, 2], all inductions contained in the atmosphere are deposited in a manner similar to that of electrostatic precipitators. The quantitative values for *Rn* and *Tn* can be obtained from the discharge curves.

Results

From the foregoing discussion, it is apparent that the activation numbers can give only an approximate picture of actual conditions. The values given in Table III represent averages derived from a large number of individual tests.

If we disregard the uncertainties involved, it is apparent from the values given in Table III that the lowest values are found over the oceans, increasing toward the shore, and the highest values in high mountains with strongly emanating igneous rocks. They give clear evidence that the radioactive admixtures get into the atmosphere exclusively from the continents.

The right-hand column of Table III gives mean values of the *Rn* content after improved induction measurements (aspiration method); they probably are much too small, particularly over the continents.

Table III. Average Values Resulting from Measurements of Inductions

Location	Activation number	*Rn* content after induction measurement (C cm$^{-3} \times 10^{-18}$)
Oceans	approx. 10	1.9
Continents	approx. 40	35
High Mountains		
Alps	approx. 100	—
High Cordilleras	450–500	—

Direct *Rn* measurements from various parts of the world are available in large number; a summary of such measurements is presented in Table IV. It will be seen from Table IV that the mean *Rn* content of the atmosphere near the surface over continents amounts to about $100–120 \times 10^{-18}$ C cm^{-3} (omitting the larger values for the high mountains (Innsbruck) and the smaller ones for high altitudes), and over oceans to about $1–2 \times 10^{-18}$ C cm^{-3}. Accordingly, one liter of air over the continent contains about 2000 atoms of *Rn*, over the oceans about 30 atoms. Thus, there can no longer be any doubt that the atmospheric *Rn* content is of purely continental origin; Bongards' assumption [12] that the atmospheric *Rn* is of cosmic origin can therefore be discarded. The same conclusion follows from the decrease of the *Rn* content with height (see below).

As far as the *Th*- and *Ac*-products of the atmosphere are concerned, data are more scarce. On the whole, it can be stated that (1) even over the continents, *Tn*, together with its disintegration products, contributes to the entire ionization probably less than, or at best just as much as, *Rn* with its inductions [44], and (2) *An* with its derivatives contributes hardly more than 3 per cent towards the formation of atmospheric ions.

The exhalation of *Rn* (*Rn* emission of the ground), according to the results obtained so far, is of the order of approximately 40×10^{-18} C cm^{-2} sec^{-1} (Table V).

The exhalation seems to have a single diurnal period with a maximum in the early forenoon [39], but its variation is strongly modified by meteorological factors [35, 61, 63]. The seasonal variation has a maximum in late summer [61]. Precipitation reduces the exhalation considerably, and solar radiation and an increase in temperature raise it; falling atmospheric pressure causes an increase of exhalation, rising pressure a decrease. Frost decreases it very sharply [63] and can stop it entirely [11].

Variations of the Atmospheric Radon Content. Over flat country and valleys the diurnal curves show unequivocally a single period with a maximum during the night (toward morning) and a minimum during the afternoon [9, 39]. In mountainous country the

diurnal curves are much less pronounced and apparently decisively influenced by austausch phenomena [47], inasmuch as for mountain stations well-developed regular diurnal variations appear only when there is a strong convective exchange of air with lower layers.

The seasonal variation seems to run parallel to the temperature curve, as is shown by the increase of the monthly mean from spring to summer [39, 42, 43, 47]; however, in the valley at Innsbruck, the cold season (January) shows the highest *Rn* values. This apparent contradiction can probably be explained by the fact that the temperature influence upon the exhalation is opposed by the effect of the temperature lapse rate on the vertical transport of *Rn*. In other words, during the cold season the *Rn* content of the valley air may increase in spite of reduced exhalation because of very little vertical transport, whereas in summer the effect of the austausch exceeds that of strong exhalation. Annual variations on mountaintops have not been measured as yet.

Table IV. Measurements of the Radon Content of the Atmosphere

Place and time of measurement	Author	Number of tests	Mean ($C\ cm^{-3} \times 10^{-18}$)	Maximum ($C\ cm^{-3} \times 10^{-18}$)	Minimum ($C\ cm^{-3} \times 10^{-18}$)
Montreal 1907–8	Eve [20]	41	60	127	18
Cambridge 1908–10	Satterly [50]	87	108	350	35
Chicago 1908	Ashman [3]	6	95	200	45
Manila 1912–14	Wright and Smith [60]	50	71	154	14
Mount Pauai 1913 (2460 m msl)	Wright and Smith [60]	10	19	34	8
Freiburg, Switzerland 1917	Olujic [45]	36	131	305	54
Innsbruck 1912–20	Schweidler [53]	339	340	1220	0
Seeham-Salzburg 1914–18	Schweidler [53]	207	114	405	0
Innsbruck 1919	Zlatarovic [62]	49	433	1140	40
Halle 1923–24	Wigand and Wenk [59]	—	—	500	300
Halle 1923–24 (0–1000 m msl)	Wigand and Wenk [59]	5	170	—	—
(1000–2000 m msl)	Wigand and Wenk [59]	3	85	—	—
(>2000 m msl)	Wigand and Wenk [59]	3	8	—	—
Novaya Zemlya 1927	Běhounek [10]	—	1*	—	—
Graz 1928	Kosmath [34]	63	142	270	43
Halle 1931	Messerschmidt [43]	704	300	900	140
Turin 1932	Aliverti [1]	26	414	—	—
Innsbruck 1933	Illing [26]	201	436	2580	25
Patscherkofel near Innsbruck 1933 (1980 m msl)	Israël [39]	9	103	610	30
Leiden, Holland					
Land breeze	Israël [30]	—	85	—	—
Sea breeze	Israël [30]	—	21	—	—
Frankfurt am Main 1933–34	Becker [9]	80	154	526	28
Taunus Observatory 1933	Becker [9]	48	125	477	14
Innsbruck 1934	Maček [38]	29	432	1780	72
Bad Nauheim 1935	Schwalb [52]	90	585	9200	0
Innsbruck 1935	Priebsch, Radinger and Dymek [47]	225	312	1560	10
Hafelekar near Innsbruck 1935–36 (2300 m msl)	Priebsch, Radinger and Dymek [47]	128	100	275	7
New York 1941–42	Hess [24]	27	97	481	10
Oceans:					
Pacific Ocean	Bauer and Swann [4]	—	3.3	—	—
Subarctic	Bauer and Swann [4]	—	0.4	—	—
All oceans, far from continents	Mauchly [41]	—	1.2	—	—

* Approximate.

There is a close relationship with meteorological influences: Falling pressure increases the *Rn* content, rising pressure decreases it [35, 39, 47], as is to be expected from the atmospheric influence upon the exhalation. The influence of the wind is twofold: With increasing wind velocity there appears first an increase in *Rn* content (because of increased exhalation), then a decrease (because of predominance of upward transport over increased exhalation). The direction of the wind is also important inasmuch as air masses of maritime origin show a smaller *Rn* content than those of continental origin [9]. Precipitation, particularly that of long duration, decreases the *Rn* content; this can easily be explained by a decrease of exhalation due to the clogging of the ground capillaries (*e.g.*, [39]). Precipitation particles themselves show a measurable content of radioactive inductions [23] which they apparently acquire while falling; on the high seas they are practically inactive, as is to be expected [48].

Table V. Radon Exhalation of the Ground

Place	Author	Exhalation ($C\ cm^{-2}\ sec^{-1} \times 10^{-18}$)
Dublin	Smyth [55]	74
Manila	Wright and Smith [60]	21
Liebenau/Graz	Kosmath [35]	40
Innsbruck	Zupancic [63]	23
Innsbruck	Zeilinger [61]	50
		Mean: 40

The fact that temperature inversion layers with a high aerosol content appear to be especially rich in *Rn* [9] would seem to indicate an austausch effect. On the other hand, this fact may also point to a causal relationship such that, because of selective adsorption by the aerosol particles, the vertical distribution of *Rn* is essentially caused by the distribution of the aerosol [27].

Radon Balance of the Atmosphere. As has been mentioned in the beginning, the gaseous emanations are the connecting links between the primary radioactivity of the ground and that of the atmosphere. By diffusion and the suction effect of the wind upon the ground capillaries, these emanations are brought into the atmosphere where they are distributed to greater heights under the influence of vertical convection. Since, together with their disintegration products, they have only a limited life-span, a height distribution, characteristic for each of the radioactive substances, must develop.

The measurements undertaken so far (see Table IV) show the expected decrease with height, but are not sufficient for the quantitative examination of this relationship. However, there is another possibility of de-

termining the balance (we restrict our examination to *Rn*).

In the mean, the supply from the ground has to maintain an equilibrium with the rate of disintegration of *Rn* in the entire atmosphere. Let us disregard the variations in a horizontal direction and consider a vertical column of air of one square centimeter cross section reaching to the top of the atmosphere; the following equation must then be fulfilled:

$$S\lambda = E,$$

where S is the entire *Rn* content of the column, λ is the disintegration constant of *Rn*, and E is the exhalation. The total content S of the air column can be computed as

$$S = \int_0^\infty s(h)\, dh,$$

where $s(h)$ represents the vertical distribution of *Rn*. For this function $s(h)$ we may write the following differential equation:

$$A\frac{d^2s}{dh^2} + \frac{dA}{dh}\frac{ds}{dh} - \rho\lambda s = 0,$$

in which ρ expresses the density of the air and A the austausch coefficient.

Integrations of this equation have been carried out by Hess and Schmidt [25] for an austausch that is constant with height ($dA/dh = 0$), and by Schmidt [51] with corrections by Priebsch [46], and by Lettau [37] for various assumed values of an austausch coefficient variable with height. Values for S and E as calculated by these various authors are presented in Table VI. The agreement between these values and the observed exhalation of 40×10^{-18} C cm^{-2} sec^{-1} (Table V) can be considered entirely satisfactory. In summary it can be said that the emanation content of the atmosphere over land and water can be completely understood if one considers the solid earth's crust as the almost exclusive source of emanation.

Table VI. Total Radon Content and Rate of Disintegration of a Vertical Column of Unit Cross Section and of Atmospheric Height, after Various Authors

Author	S (C $\times 10^{-12}$)	E (C sec$^{-1} \times 10^{-18}$)
Hess and Schmidt [25]		
A = 50 g cm^{-1} sec^{-1}	13	27
A = 100 g cm^{-1} sec^{-1}	18	38
Schmidt [51] (Priebsch [46])	8	20
Lettau [37]	27	57

Problems

Our knowledge of radioactive substances in the atmosphere is fairly complete as far as their identity, measurability, and origin are concerned. Less clear is the mechanism of the horizontal and vertical distribution of these substances in the atmosphere. As admixtures to the air, they take part in its movement and thus enable us to reach important conclusions regarding air-mass displacements. Therefore, if one considers the question of a suitable continuation of research, the use of radioactive air admixtures as tracers for special meteorological problems opens new possibilities. One experiment in particular suggests itself: the use of radioactive emanations as an indicator of austausch movements on the one hand, and for the determination of the life history of air masses on the other hand. Several of the above-mentioned investigations point in this direction; only a few will be mentioned.

1. The vertical distribution of the individual radioactive component can be considered as the result of austausch [9, 25, 27, 37, 39, 46, 47, 51]; therefore, with proper measurements, it ought to yield, in turn, information about its efficacy and vertical extent on the average as well as in single cases.

2. Air that stagnates in the same climatic region for some time acquires, by contact with the ground, certain characteristics which allow us to define it as an air mass of a given type. One must expect that the air mass also adopts different radioactive properties, according to the exhalation of the ground beneath. Thus, for instance, an air mass located over the ocean for a long time, will show a considerably smaller *Rn* content than air from the continent, as has been shown by measurements of the author on the Dutch coast [30] (see Table IV). Furthermore, the pronounced activity differences which Becker [9] has found above and below an inversion point to the effects of austausch or air-mass characteristics.

3. The radioactivity of precipitation gives an indication of the radioactive character of the air mass from which it falls. The pronounced increase in ionization near the ground during thundershowers [15, 57] is probably due to an especially strong upward transport of low-level (and therefore more strongly radioactive) air in the formation of a thunderstorm and the return of the radioactive admixtures through precipitation.

It is obvious that a more thorough investigation of these relationships from a meteorological and thus atmospheric-electrical and bioclimatological point of view can become of great importance.

The methodological problems, for example, consist of the following:

1. The over-all substitution of measurement by automatic recording.

2. A simultaneous survey of diurnal variations of the radioactive elements in the air at several stations at various altitudes.

3. Attempts at an air-mass classification according to origin and age on the basis of the *Rn* content and the proportion of *Ra*- to *Th*-derivatives.

4. Vertical cross sections of the atmospheric radioactivity, perhaps by means of airplane measurements or by testing the active deposits on the mooring cable of a captive balloon when bringing it in (using Geiger counters).

5. Determination of the radioactivity of precipitation.

REFERENCES

1. Aliverti, G., "Nuovo metodo per determinare quantitativamente il contenuto in sostanze radioattive dell'aria atmosferica." *Nuovo Cim.*, 8:233–245 (1931).
2. —— "Le misure di radioattività atmosferica con il metodo dell'effluvio." *Nuovo Cim.*, 9:313–327 (1932); "Sul metodo dell'effluvio per misure di radioattività atmosferica." *Atti Accad. Torino*, 68:376–385 (1932); "Quantitative Bestimmungen des Luftgehaltes an Radium-Thoriumemanation mittels einer neuen elektrischen Ausströmungsmethode." *Z. Geophys.*, 9:16–22 (1933).
3. Ashman, G. C., "A Quantitative Determination of the Radium Emanation in the Atmosphere." *Amer. J. Sci.*, 176:119–122 (1908).
4. Bauer, L. A., "Reports on Investigations and Projects; Department of Terrestrial Magnetism." *Yearb. Carneg. Instn.*, 15:287–336 (1916); 16:249–284 (1917); 17:233–268 (1918).
5. —— and others, "Researches of the Department of Terrestrial Magnetism: Ocean Magnetic Observations, 1905–1916." *Carneg. Instn. Wash. Publ.*, No. 175, Vol. 3 (1916).
6. Becker, A., "Zur Methodik der Emanationsmessung." *Strahlentherapie*, 15:365–383 (1923).
7. —— "Über die Präzisionsmessung der Radiumemanation." *Z. Phys.*, 21:304–315 (1924). (See also [6])
8. —— und Holthusen, H., "Über absolute Radiumbestimmungen mit dem Emanometer." *S. B. heidelberg. Akad. Wiss.*, Nr. 6, 34 SS. (1913).
9. Becker, F., "Messungen des Emanationsgehaltes der Luft in Frankfurt a. M. und am Taunus-Observatorium." *Beitr. Geophys.*, 42:365–384 (1934).
10. Běhounek, F., "Recherche sur l'électricité et la radioactivité de l'atmosphère au Spitzberg." *J. Phys. Radium*, (6) 8:161–181 (1927).
11. Bender, H., "Über den Gehalt der Bodenluft an Radiumemanation." *Beitr. Geophys.*, 41:401–415 (1934).
12. Bongards, H., "Ein unbekanntes Wetterelement?" *Meteor. Z.*, 36:339–342 (1919); "Messungen des Gehaltes der Luft an radioaktiven Zerfallsprodukten vom Flugzeug aus." *Phys. Z.*, 25:679–682 (1924).
13. Curie, M., *Traité de radioactivité*, Tomes I & II. Gauthier-Villars, Paris, 1910. (See Vol. II, pp. 474–490)
14. Deij, L. J. L., *Metingen van het radongehalte in de atmosfeer.* Leiden, A. W. Sijthoff, 1939.
15. Doan, R. L., "Fluctuations in Cosmic-Ray Ionization as Given by Several Recording Meters Located at the Same Station." *Phys. Rev.*, 49:107–122 (1936).
16. Duane, W., "Sur l'ionisation de l'air en présence de l'émanation du radium." *J. Phys.*, (4) 4:605–619 (1905).
17. —— et Laborde, A., "Sur les mesures quantitatives de l'émanation du radium." *C. R. Acad. Sci., Paris*, 150: 1421–1423 (1910).
18. Eckmann, G., "Über Wanderung und Diffusion der *RaA*-Atome." *Jb. Radioakt.*, 9:157–187 (1912).
19. Elster, J., und Geitel, H., "Über eine fernere Analogie in dem elektrischen Verhalten der natürlichen und der durch Becquerelstrahlen abnorm leitend gemachten Luft." *Phys. Z.*, 2:590–593 (1901).
20. Eve, A. S., "On the Radioactive Matter Present in the Atmosphere." *Phil. Mag.*, (6) 10:98–112 (1905); "On the Amount of Radium Emanation in the Atmosphere near the Earth's Surface." *Ibid.*, 14:724–733 (1907).
21. Flamm, L., und Mache, H., "Über die quantitative Messung der Radiumemanation im Schutzringplattenkondensator." *S. B. Akad. Wiss. Wien*, Abt. IIa, 122:535–542 (1913).
22. Gerdien, H., "Untersuchungen über die atmosphärischen radioaktiven Induktionen." *Abh. Ges. Wiss. Göttingen, Mat.-Nat. Kl.*, Bd. V, Nr. 5, 74 SS. (1907); "Über die spezifische Geschwindigkeit der positiv geladenen Träger der atmosphärischen radioaktiven Induktionen." *Phys. Z.*, 6:465–472 (1905).
23. Gockel, A., und Wulf, T., "Beobachtungen über die Radioaktivität der Atmosphäre im Hochgebirge." *Phys. Z.*, 9:907–911 (1908).
24. Hess, V. F., "On the Radon-Content of the Atmosphere and the Radium-Content of River-Water." *Terr. Magn. atmos. Elect.*, 48:203–206 (1943).
25. —— und Schmidt, W., "Über die Verteilung radioaktiver Gase in der freien Atmosphäre." *Phys. Z.*, 19:109–114 (1918).
26. Illing, W., Dissertation, Innsbruck, 1933.
27. Israël, H., "Emanation und Aerosol." *Beitr. Geophys.*, 42:385–408 (1934).
28. —— "Zur Methodik der klimatologischen Emanationsmessungen." *Der Balneologe*, 1:318–327 (1934).
29. —— *Radioaktivität.* Leipzig, J. A. Barth, 1940. Reprinted by Edwards Bros., Ann Arbor, Mich., 1945.
30. —— "Luftelektrische Messungen in Leiden (Holland)." *Biokl. Beibl.*, 2:129–133 (1935).
31. —— "Zur Methodik der klimatologischen Emanationsmessungen." *Beitr. Geophys.*, 51:35–49 (1937).
32. —— "Das Emanationsdosimeter: Ein Gerät zur Dauerkontrolle mässig hoher Emanationskonzentrationen in Luft." *Wiss. Abh. D. R. Reich. Wetterd.*, Bd. 2, Nr. 10 (1937).
33. Janitzky, A., "Zur Frage der Durchlässigkeit der menschlichen Haut für Radiumemanation." *Der Balneologe*, 2:117–128 (1935).
34. Kosmath, W., "Der Gehalt der Freiluft an Radiumemanation und deren vertikale Verteilung in der Nähe des Erdbodens (nach Beobachtungen in Graz im Jahre 1928)." *Beitr. Geophys.*, 25:95–117 (1930).
35. —— "Die Exhalation der Radiumemanation aus dem Erdboden und ihre Abhängigkeit von den meteorologischen Faktoren." *Beitr. Geophys.*, 43:258–279 (1935).
36. Lester, O. C., "A Rapid and Accurate Method of Correcting Measurements with Emanation Electroscopes for Ordinary Changes in Temperature and Pressure." *J. opt. Soc. Amer.*, 11:637–640 (1925).
37. Lettau, H., "Anwendung neuerer Ergebnisse der Austauschlehre auf zwei luftelektrische Fragen." *Beitr. Geophys.*, 57:365–383 (1941) (Benndorf-Heft).
38. Maček, O., Dissertation, Innsbruck, 1934.
39. —— und Illing, W., "Messung des Emanationsgehaltes der Luft nach der Spitzenmethode." *Beitr. Geophys.*, 43:388–418 (1935).
40. Marx, E., und Wolf, L., "Vorlesungsversuch der Isolierung radioaktiver Substanzen durch Rückstoss." *Phys. Z.*, 24:285–286 (1923).
41. Mauchly, S. J., "The Radium-Emanation Content of Sea Air from Observations aboard the *Carnegie*, 1915–1921." *Terr. Magn. atmos. Elect.*, 29:187–194 (1924).
42. Messerschmidt, W., "Eine neue Methode zur Bestimmung des Emanationsgehaltes der Atmosphäre." *Phys. Z.*, 32:548–549 (1931).
43. —— "Eine neue Methode zur Bestimmung des Emanationsgehaltes der Atmosphäre und ihre Anwendung zur Untersuchung der Zusammenhänge mit den meteorologischen Faktoren und des Einflusses des Emanationsgehaltes der Atmosphäre auf die Messungen der Ultrastrahlung." *Z. Phys.*, 81:84–100 (1933).

44. Meyer, S. und Schweidler, E. v., *Radioaktivität*, 2. Aufl. Leipzig, Teubner, 1927.
45. Olujic, J., Dissertation, Freiburg/Schweiz, 1918.
46. Priebsch, J., "Zur Verteilung radioaktiver Stoffe in der freien Luft." *Phys. Z.*, 32:622–629 (1931).
47. —— Radinger, G., und Dymek, P. L., "Untersuchungen über den Radiumemanationsgehalt der Freiluft in Innsbruck und auf dem Hafelekar (2300 m)." *Beitr. Geophys.*, 50:55–77 (1937).
48. Rouch, J., "Observations d'électricité atmosphérique faites dans l'Antarctique pendant l'Expédition Charcot (1909)." *Ann. météor. Fr.*, 59:117–124 (1911); "Observations d'électricité atmosphérique dans l'Atlantique." *Ibid.*, 61:149–161 (1913).
49. Salpeter, J., "Über den Einfluss des Erdfeldes auf die Verteilung der Radiuminduktion in der Atmosphäre und auf der Erdoberfläche." *S. B. Akad. Wiss. Wien*, Abt. IIa, 118:1197–1205 (1909); 119:107–118 (1910).
50. Satterly, J., "The Amount of Radium Emanation in the Atmosphere." *Phil. Mag.*, (6) 16:584–615 (1908); "On the Amount of Radium Emanation in the Lower Regions of the Atmosphere and Its Variation with the Weather." *Ibid.*, 20:1–36 (1910); "Some Experiments on the Absorption of Radium Emanation by Coconut Charcoal." *Ibid.*, 20:773–788 (1910).
51. Schmidt, W., "Zur Verteilung radioaktiver Stoffe in der freien Luft." *Phys. Z.*, 27:371–378 (1926).
52. Schwalb, K., "Beiträge zur Kenntnis der Radium-Emanation in der Atmosphäre." *Biokl. Beibl.*, 8:82–90 (1941).
53. Schweidler, E. v., "Zusammenfassender Bericht über die Beobachtungen an der luftelektrischen Station Seeham in den Sommern 1908 bis 1915; II. Teil: Ionisierung in geschlossenen Gefässen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 126:1009–1035 (1917); "Zusammenfassender Bericht über die Beobachtungen an der luftelektrischen Station Seeham in den Sommern 1916 bis 1920." *Ibid.*, 129:919–927 (1920).
54. Siegl, L., "Über die quantitative Messung der Radiumemanation im Schutzringplattenkondensator." *S. B. Akad. Wiss. Wien*, Abt. IIa, 134:11–37 (1925).
55. Smyth, L. B., "On the Supply of Radium Emanation from the Soil to the Atmosphere." *Phil. Mag.*, (6) 24:632–637 (1912).
56. Swann, W. F. G., "The Measurement of Atmospheric Conductivity, together with Certain Remarks on the Theory of Atmospheric Radioactive Measurements." *Terr. Magn. atmos. Elect.*, 19:23–37 (1914); "The Theory of Electrical Dispersion into the Free Atmosphere, with a Discussion of the Theory of the Gerdien Conductivity Apparatus, and of the Theory of the Collection of Radioactive Deposit by a Charged Conductor." *Ibid.*, 19:81–92 (1914); "On Certain New Atmospheric-Electric Instruments and Methods." *Ibid.*, 19:171–185 (1914); "The Atmospheric-Electric Observations on the Third Cruise of the *Carnegie*, 1914." *Ibid.*, 20:13–48 (1915).
57. Wait, G. R., and McNish, A. G., "Atmospheric Ionization near the Ground during Thunderstorms." *Mon. Wea. Rev. Wash.*, 62:1–4 (1934).
58. Wigand, A., "Luftelektrische Untersuchungen bei Flugzeugaufstiegen." *Phys. Z.*, 25:684–685 (1924).
59. —— und Wenk, F., "Der Gehalt der Luft an Radium-Emanation nach Messungen bei Flugzeugaufstiegen." *Ann. Phys., Lpz.*, (4) 86:657–686 (1928).
60. Wright, J. R., and Smith, O. F., "The Variation with Meteorological Conditions of the Amount of Radium Emanation in the Atmosphere, in the Soil Gas, and in the Air Exhaled from the Surface of the Ground, at Manila." *Phys. Rev.*, (2) 5:459–482 (1915).
61. Zeilinger, P. R., "Ueber die Nachlieferung von Radiumemanation aus dem Erdboden." *Terr. Magn. atmos. Elect.*, 40:281–294 (1935).
62. Zlatarovic, R., "Messungen des Radium-Emanationsgehaltes in der Luft von Innsbruck." *S. B. Akad. Wiss. Wien*, Abt. IIa, 129:59–66 (1920).
63. Zupancic, P. R., "Messungen der Exhalation von Radiumemanation aus dem Erdboden." *Terr. Magn. atmos. Elect.*, 39:33–46 (1934).

CLOUD PHYSICS

On the Physics of Clouds and Precipitation *by Henry G. Houghton* 165

Nuclei of Atmospheric Condensation *by Christian Junge* 182

The Physics of Ice Clouds and Mixed Clouds *by F. H. Ludlam* 192

Thermodynamics of Clouds *by Fritz Möller* 199

The Formation of Ice Crystals *by Ukichiro Nakaya* 207

Snow and Its Relationship to Experimental Meteorology *by Vincent J. Schaefer* 221

Relation of Artificial Cloud-Modification to the Production of Precipitation *by Richard D. Coons and Ross Gunn* 235

ON THE PHYSICS OF CLOUDS AND PRECIPITATION

By HENRY G. HOUGHTON

Massachusetts Institute of Technology

INTRODUCTION

The entire science of meteorology is concerned with the physics of the atmosphere, but the term *physical meteorology* has been accepted as the designation for only one portion of the science. This division of the field includes topics such as atmospheric optics, atmospheric electricity, solar and long-wave radiation, and the physical processes of condensation and precipitation. The latter is the subject of the present contribution. The discussion will start with a consideration of condensation nuclei and will then proceed in turn to treat the initiation of condensation, the growth of the condensation products, and the formation of precipitation elements. A distinction must be made between condensation in the liquid phase and in the solid phase. There are also differences between the formation of solid and liquid precipitation elements. A brief discussion of the artificial dissipation of fog will also be included.

Some of these topics are discussed by other contributors to this volume.[1] The purpose of the present contribution is to review the entire subject. For this reason no attempt will be made to avoid the topics covered by the other authors. Such duplication is not only essential to a complete discussion of the subject but may also serve to emphasize more clearly some of the different points of view held by various workers in the field. Continuity does not require a discussion of the artificial modification of clouds, and this topic has been omitted in view of the complete treatments in this volume by Coons and Gunn and by Schaefer.

The subject of cloud physics has received much attention in recent years. This is due in large measure to the recent experimental work on the artificial modification of supercooled clouds. Other stimuli have been the problems of aircraft icing, artificial fog dispersal, the propagation of microwave radio energy, and the detection of precipitation areas by radar. In sharp distinction to many other areas of meteorology, the problems of cloud physics are amenable to the techniques of the experimental physicist both in the laboratory and in the free atmosphere.

NUCLEI OF CONDENSATION

The discovery of nuclei of condensation is generally attributed to Coulier who reported on them in 1875. The pioneer in this field was John Aitken [2] whose work on nuclei of condensation extended from about 1880 to 1916. C. T. R. Wilson's name is usually coupled with Aitken's, but his work has proven to be of more value to particle physicists than to meteorologists.

Once it was established that all natural condensation requires the presence of condensation nuclei, attention was focused on their source, nature, and size and on their distribution in time and space. Many of these questions are still in debate. Aitken studied the subject both in the laboratory and in the open air. His insight into the problem and his experimental techniques stand unrivaled as monuments to his name in this field. He developed the expansion type of nucleus counter which has been used in one form or another by all subsequent workers in the field. The careful reading of Aitken's many papers [2] is an absolute requirement for any serious student of condensation nuclei.

Source and Nature of Nuclei. Aitken [2] always referred to nuclei of condensation as "dust particles" although he stated clearly that they were distinct from the type of dust raised, for example, by high winds. He felt that there were two types of nuclei, those with an affinity for water vapor on which condensation begins below saturation, and nonhygroscopic nuclei which require an appreciable degree of supersaturation for the initiation of condensation. In his opinion the first type is the true fog-former while the second type usually produces only haze. Using an Aitken "dust counter," Wigand [56] found that an artificial increase in the dust content of air, such as was produced by beating a carpet in a room, had no effect on the nucleus count. He concluded that such nonhygroscopic dust was inactive as condensation nuclei and proposed that the Aitken dust counter be renamed the *kern counter*, a suggestion that has been generally adopted. Wigand's conclusions were apparently supported by kern counts in the presence of sand and dust storms and by Boylan's laboratory studies [5]. Boylan's results indicated that the introduction of dust such as coal dust and carbon black slightly decreased the kern count. He suggested that this was due to the sweeping action of the dust particles on the kerns. Later Junge [26] experimented with a wide variety of dusts, including some which are not wetted by water, and found that all dusts with a large number of particles smaller than about 10-μ radius increased the kern count. He concluded that particles of any substance can act as nuclei of condensation. He argued that larger particles would fall out in the chamber of the kern counter before the expansion could be made. Junge felt that earlier investigators did not produce dusts with a sufficient number of small particles to increase the kern count significantly. As he pointed out, a dust cloud of several hundred large particles in a cubic centimeter looks much denser than so-called "dust-free" air, which may contain several hundred thousand ultramicroscopic condensation nuclei per cubic centimeter. Junge's results are very convincing and it seems reasonable to accept his conclusions. On the other hand, the evidence still

1. In particular, reference should be made to the papers by Coons and Gunn, Junge, Ludlam, Möller, Nakaya, and Schaefer.

supports Aitken's conclusion that the hygroscopic nuclei are the important cloud producers and that neutral dust is of lesser importance. Boylan [5] found in Dublin that the number of kerns determined with the Aitken instrument averaged fifteen times the number of dust particles determined with the Owens impact dust counter.

Aitken and many others have established that flames and burning materials form tremendous numbers of nuclei and also that heat alone, as from a heated platinum filament, will form nuclei in natural air. Aitken and also Coste and Wright [6] showed that nuclei could be formed by spraying sea water. The latter authors also found that fuming sulfuric acid was an active kern producer. Although flames produce a variety of products it has been established that substances containing sulfur are the most effective fuels. This is significant because of the sulfur content of coal. When sulfur is burned the nonhygroscopic dioxide is formed. This does not easily oxidize to the hygroscopic trioxide in air at normal temperatures. Aitken found that both ozone and hydrogen peroxide are effective oxidizers of sulfur dioxide. He also claimed that ultraviolet solar radiation oxidized the sulfur dioxide. Coste and Wright [6] suggested that at temperatures above 620C, and in the presence of water vapor, hygroscopic nitrous acid is formed. This would explain the production of nuclei by hot filaments. They also suggested that the nitrous acid is the oxidizing agent for sulfur dioxide.

There is general agreement that sea salts are effective as condensation nuclei. Sea-salt crystals have been observed in the atmosphere by Owens [43], Dessens [8, 9], and Woodcock and Gifford [57]. Köhler [27, 28] and others have demonstrated the presence of chloride ion and of other sea-salt anions in water from clouds, fogs, and rain. If the amount of chloride ion, for example, as measured in a bulk sample is divided by the number of drops, it yields a nucleus of reasonable size. Any object which has not been carefully cleaned exhibits the ubiquitous sodium flame when it is heated. On the other hand there is evidence which suggests that the sea is not the principal source of condensation nuclei. The number of nuclei is much smaller over the oceans than over the land. A minimum kern count of nearly zero has been observed over the ocean and the usual count is from several hundred to several thousand per cubic centimeter. Typical values over land range up to 100,000 in rural, and a million or more in urban areas. Simpson [49] shows that if all active nuclei come from the sea surface, the rate of nucleus production must be about 1250 sec^{-1} cm^{-2}. This appears to be unreasonably large if the nuclei are formed by the evaporation of the spray resulting from wave action. A large proportion of the spray drops will be so large that they will fall back to the surface. Findeisen [12] holds that the sea salt indubitably present in the atmosphere is in the form of say 10–20 large particles of about 10^{-10} grams mass per liter of air. Although these are probably all active nuclei, their number is small compared to the total number of nuclei. Their presence would serve to explain the observed chloride content of cloud and rain water. There are isolated observations that the evaporation of sea water (without visible spray) produces kerns, and Aitken found that nuclei were formed by the action of the sun on the foreshore at ebb tide. Attempts to identify the nucleus in an evaporating cloud drop under the microscope have failed. Dessens [8] has caught nuclei from the air on spider webs and caused them to grow into drops or to form crystals by varying the humidity. These are probably the relatively large sea-salt particles referred to by Findeisen [12]. Wright [58] has shown that visibility near the seacoast is a function of the relative humidity, which can be explained by assuming that the nuclei are hygroscopic. Wright felt that he had shown in this way that the nuclei in question were sea salts but Simpson [49] argued that his procedure did not permit the identification of the hygroscopic agent.

It is generally concluded that most active condensation nuclei are hygroscopic particles and that the nonhygroscopic nuclei are unimportant. More information is required to explain this strong preference for hygroscopic nuclei. The hygroscopic and nonhygroscopic kerns are presumably of about the same size; if anything, the hygroscopic nuclei are somewhat smaller. It can be shown that the lowering of the vapor pressure by the salt is not of major importance, since a slightly larger nonhygroscopic particle will support condensation at the same degree of supersaturation. It appears from Volmer's work [52] that the wettability of the substance plays an important role. On a surface which is not wetted by water, condensation occurs in the form of small lens-shaped drops. The work required is greater than when the surface is wetted and increases with the contact angle between the water and the surface. Greater work for the phase change implies a higher supersaturation. Hygroscopic nuclei are liquid drops before cloudy condensation occurs and are therefore perfectly wetted. All other nonhygroscopic surfaces are less easily wetted and it may be that the nonhygroscopic atmospheric dust is largely hydrophobic. Junge's finding [26] that paraffin spheres will serve as condensation nuclei does not contradict this explanation since Volmer shows only that hydrophobic nuclei require a greater supersaturation than hydrophilic nuclei. Junge did not report on the supersaturations used, but since he employed an Aitken instrument it may be assumed that they were from 200 to 300 per cent. An extension of Junge's work with provisions for varying the supersaturation would probably shed some light on this problem.

The available evidence suggests that the hygroscopic nuclei are formed of sea salts and nitrous and sulfuric acids. Other hygroscopic materials may also be important but it has yet to be shown that other such substances regularly exist in the atmosphere in sufficient quantity and in finely divided form. If it be assumed that sea-salt nuclei are formed only by the evaporation of spray, it must be conceded that the source is insufficient to supply a major fraction of atmospheric kerns. Further study on other possible mechanisms for the formation of sea-salt nuclei is

desirable. Perhaps rapid evaporation of sea water in the absence of spray forms nuclei. Aitken's observation that nuclei are formed by the action of sunlight on salt-water beaches should be followed up. Dessens [9] found that salt droplets become supersaturated when the relative humidity is decreased and finally crystallize explosively, occasionally resulting in rupture. He tentatively offered this as a possible mechanism for the production of small condensation nuclei from the relatively large crystals formed when sea spray evaporates, but later expressed the opinion that the fracture mechanism is too rare to be of major importance.

The sulfuric acid in the atmosphere must come primarily from the sulfur in coal, although volcanic activity may contribute a small amount. Much of the research on nuclei has been done in industrial areas such as the British Isles and Germany, so that the importance of sulfuric acid nuclei may have been overemphasized. Although the high concentrations of nuclei in industrial areas is almost certainly due to the products of man-made combustion, no one believes that the cloud and precipitation regimes of the world have been greatly altered by industrialization. It appears that nitrous acid nuclei are also produced by combustion but here it is the high temperature which acts as a catalyst to form the nuclei from the natural constituents of the air. Forest fires are as effective for this as man's furnaces. It is also known that nitrous acid is formed by the action of lightning discharges. It may be that nitrous acid is the principal kern material and that the high kern-concentrations in urban areas are due to the products of combustion and are of only local importance. Adel [1] and Shaw and collaborators [48] have obtained spectroscopic evidence of the presence of about 1 cm of nitrous oxide at NTP in the atmosphere. There is no information on the vertical distribution of the nitrous oxide but these observations suggest that it is a universal constituent of the atmosphere. There may well be other substances, even nonhygroscopic ones, obscured by the abundance of other nuclei in industrial areas, which are the really important nuclei of the free atmosphere.

Size and Size Distribution of Nuclei. The upper limit of nucleus size is set by the settling rate and also in some cases by the method of formation. As an example, the largest sea-salt nucleus probably has a diameter of the order of 5×10^{-3} cm. Because the work of subdivision of a solid or a liquid increases rapidly with the amount of new surface formed, the diameter of the smallest sea-spray nucleus or dust particle produced by erosion or grinding is of the order of 10^{-5} cm. Diameters of nuclei formed from gases such as sulfuric and nitrous acids probably range from 10^{-5} to 10^{-6} cm. Such nuclei are not apt to be as large as sea-salt nuclei because of the small concentration of the gases from which they are formed. All sizes within these rather wide limits are to be expected. (For completeness it may be mentioned that small ions with dimensions of the order of 10^{-7} cm will serve as nuclei only at four to five fold supersaturations and cannot play any role in natural atmospheric condensation.)

Only the larger nuclei can be measured with the visual microscope. The electron microscope opens the way for the measurement of smaller nuclei if a means can be found for the collection of the nuclei on the stage of this instrument. Most measurements of the size of nuclei have been made by indirect means. The large or Langevin ions of the atmosphere are generally believed to be condensation nuclei which have captured a small ion or an electron. Boylan [5] states that about 60 per cent of the nuclei carry a single electronic charge and are identical with the large ions. The mobility (velocity in unit electric field) of such ions may be measured and their size computed from Stokes' law. Such measurements yield diameters in the neighborhood of 10^{-5} cm.

As already mentioned, the size of the nucleus may be determined by assaying a bulk sample of cloud or fog water for the assumed constituent and dividing by the number of drops represented in the sample. This was done first by Köhler [27], who collected rime at a mountain observatory and determined the chloride content by titration. The mean drop size was measured by the corona method. By assuming that the other anions of sea salt were present in the same proportion as in the sea, he computed the average mass of the nuclei to be 1.847×10^{-14} g, equivalent to a diameter of about 5×10^{-5} cm. This procedure is open to the criticism of Findeisen [12] that the salt might be in the form of a few relatively large particles.

Direct measurements of the size of nuclei have been made by Dessens [9] and by Woodcock and Gifford [57]. Dessens caught the nuclei on spider threads and examined them with a visual microscope. He reported radii of nuclei ranging from 0.3 to 0.5 μ. The nuclei were in the form of drops at a relative humidity of 60 per cent. There were others too small to measure, and also some larger ones. Woodcock and Gifford collected nuclei on glass slides 1 mm by 15 mm in size from an airplane flying over the ocean. The counts were corrected for the collection efficiency of the slides. They identified the nuclei as sea salts by determining the relative humidity at the transition between crystal and solution. They present their data in form of the mass distribution of the nuclei. The largest nuclei had a mass of about 2×10^{-9} g and the smallest a mass of near 5×10^{-14} g. These weights correspond to diameters of 24 and 0.7 μ respectively at a relative humidity of 80 per cent. When plotted on logarithmic paper, Woodcock and Gifford's distribution curves of nucleus mass versus number are nearly linear with negative slopes. Their method did not permit the collection of nuclei of mass less than 5×10^{-14} g. The total number of nuclei in a cubic centimeter of air was found to range from less than one to about thirty and to decrease rapidly with elevation in the first 300 m above sea level. Although no simultaneous measurements with an Aitken counter are reported, it is probable that the Aitken count would be large compared with that of Woodcock and Gifford. There is no evidence that these unmeasured nuclei are composed of sea salts.

The methods outlined above of determining the size

of nuclei yield the actual geometric diameter or mass. With the sole exception of the application of mobility measurements to large ion-nuclei, none of these methods can be used in the size range which includes the majority of the natural condensation nuclei of the atmosphere. An indirect measure of the effective size of condensation nuclei may be obtained by causing condensation to occur on them under controlled conditions. To discuss this procedure it is necessary to review briefly the theory of condensation on nuclei as first presented by Thomson [50] for neutral nuclei and Köhler [29] for hygroscopic nuclei. Thomson showed that the vapor pressure in equilibrium with a curved surface is greater than the vapor pressure in equilibrium with a plane surface at the same temperature. The supersaturation required to initiate condensation on the surface of a small sphere, which is wetted by water, is nearly inversely proportional to the radius. This effect is illustrated by the upper curve in Fig. 1. The vapor

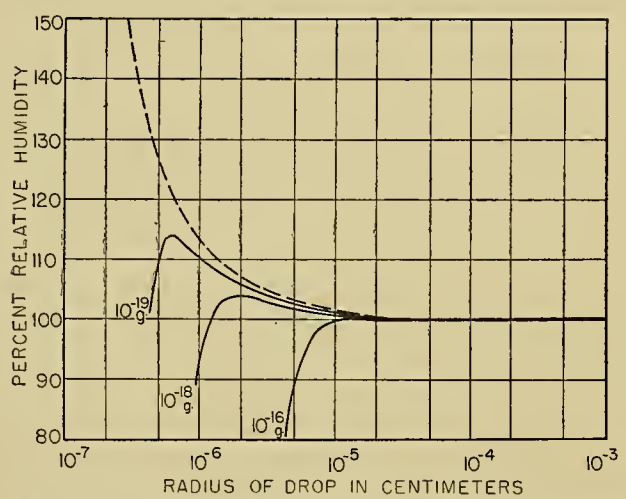


Fig. 1.—Growth curves of sodium chloride nuclei of masses as indicated at 0C. The dashed line is for pure water drops.

pressure over a solution of a hygroscopic salt is lower than that over pure water. In the case of a hygroscopic nucleus, both effects are in evidence and act in opposition. The net result is indicated by the lower curves of Fig. 1. The effect of the dissolved salt is dependent upon its concentration and thus is a function of the mass of the hygroscopic material and of the radius of the droplet. The Thomson effect is a function only of the radius of curvature. The three lower curves of Fig. 1 are for nuclei of sodium chloride of different masses as indicated on the curves. It is readily apparent that hygroscopic nuclei grow more slowly as the relative humidity increases toward saturation. In all cases, the growth curves reach a maximum in excess of 100 per cent relative humidity. This peak relative humidity must be exceeded if the nucleus is to become a cloud drop. The general behavior of these curves has been verified experimentally by Junge [25] who determined the size of hygroscopic nuclei at various relative humidities from the mobility of the nuclei as large ions. Wright [58] has obtained data on the variation of visibility with humidity which seem to confirm the general shape of the curves of Fig. 1 below saturation.

The supersaturation corresponding to the maximum of the curves of Fig. 1 is a measure of the effective size of the nucleus. The geometric size can be determined only if the nature of the hygroscopic salt is known. An adaptation of the Aitken nucleus counter can be used to determine the effective size of the nuclei, as was shown by Junge [25]. He produced expansions of known and increasing amounts in a chamber and determined the number of drops after each expansion. In this way he obtained a nucleus spectrum. Earlier, Aitken [2] performed a similar experiment with his apparatus. The principal difference in the two techniques was that Junge used a large chamber and applied his successive expansions rapidly so that none of the previously activated nuclei would evaporate. Aitken first produced a small expansion and counted the number of drops which fell out in the usual manner. He then proceeded to an increased expansion, waiting each time until the activated nuclei had fallen out of the air. Aitken's procedure is open to the objection that some of the drops might evaporate before falling out and thus leave nuclei to be counted again in subsequent expansions. Junge's method involves the errors of counting the number of drops while they are suspended in the air. Most of Junge's experiments were carried out with artificially produced nuclei, whereas Aitken used natural air. Qualitatively their results were very similar. The evidence is that a large number of nuclei are activated at the lowest expansion used, with smaller numbers requiring greater expansions or supersaturations. Aitken found that all of the ordinary nuclei in the samples of natural air were activated by a relative humidity of 150 per cent. Further increases in the expansion ratio had no effect until the supersaturation required to produce condensation on small ions was reached. Junge found that 110 per cent relative humidity was sufficient to activate all of the nuclei in a sample of outdoor air.

Number and Distribution of Nuclei. Literally thousands of measurements of the number of nuclei in the air have been made with the aid of the Aitken instrument. As ordinarily used, this instrument produces a supersaturation of from 200 to 300 per cent and therefore activates all natural nuclei but not small ions. As will be pointed out below, only a small fraction of the total number of nuclei are activated in natural condensation processes. For this reason the total number of nuclei as determined by the Aitken counter is of limited value since no information regarding the size or the size distribution of the nuclei is obtained. A very complete summary of the measurements which have been made has been given by Landsberg [33], and no attempt will be made here to give any of the detailed results. The maximum concentration was found in cities, with an average of 150,000 per cubic centimeter and a maximum of some four million. In the country, the average is of the order of 50,000 and the maximum near 400,000. Much lower concentrations are found over the oceans with an average of about 1000 and a maximum of about 40,000. At a given

location, the concentration of nuclei has a diurnal and annual variation, the nature of which is largely dependent upon the local conditions. Correlations have been made of nuclei concentration with visibility, air mass, wind direction and force, and so forth. The important question of the variation of the concentration of nuclei with elevation has not been thoroughly studied because of the difficulties in the way of such measurements. Most of these measurements have been made in mountainous regions at various elevations. These show a rapid decrease in concentration with elevation. A few determinations have been made from free balloons. These show a more rapid decrease of concentration of nuclei with increasing elevation than the mountain observations. The average vertical distribution of nuclei from the balloon ascents shows a count of 22,300 per cubic centimeter in the layer from 0 to 500 m decreasing to 80 above 5000 m. Of necessity the balloon flights were made in anticyclonic weather, and the results cannot be considered typical of stormy conditions. In any event, the rapid decrease in concentration with increased elevation indicates that the source of condensation nuclei is at or near the surface. The nuclei are presumably carried aloft by turbulence and convection. This reasoning would suggest that the decrease in concentration with elevation would be smaller in cyclonic than in anticyclonic conditions. No information is available on the change in size of the nuclei with elevation. It would be highly desirable to obtain more information regarding both the number and the size distribution of condensation nuclei in the free atmosphere.

THE INITIAL PHASE OF CONDENSATION

Reference has been made above to the factors which control condensation on both hygroscopic and nonhygroscopic condensation nuclei. Referring again to Fig. 1, it is evident that condensation nuclei will not attain cloud drop size unless the supersaturation corresponding to the maximum of the curves for hygroscopic particles is exceeded. The magnitude of the supersaturation required is dependent on the mass of the hygroscopic material in the nucleus. The supersaturation required in the case of nonhygroscopic nuclei is dependent on the radius of curvature of the nucleus as indicated by the upper curve in Fig. 1. As was pointed out earlier, the critical supersaturation is probably also dependent on how easily the surface of the nucleus is wetted. If the nucleus is composed of a microporous substance, condensation will occur at a lower relative humidity, the exact value depending on the size of the pores. Even in this case the nucleus cannot grow to cloud drop size unless some initial supersaturation occurs. It is therefore generally true that cloudy condensation cannot occur without a small degree of supersaturation. Such data as are available on the size and size distribution of nuclei indicate that the supersaturation required is usually less than one per cent. This is confirmed by the observation that the cloud base corresponds to the saturation level within the precision of the measurements.

As the relative humidity exceeds 100 per cent, condensation occurs first on the largest nuclei, that is, on those requiring the smallest amount of supersaturation to become active. Nuclei are said to be active when they have exceeded their critical supersaturations and are free to grow to cloud drop size. If the condensation is extremely slow, only the very largest nuclei will become active. As the rate of condensation is increased, the rate of condensation on the larger nuclei will not be sufficient to hold the supersaturation down and additional nuclei will be activated. This shows that the concentration of cloud drops is dependent primarily on the initial rate of condensation.

The process which has just been described qualitatively has been investigated theoretically and numerically by Köhler [29] and by Howell [24]. Köhler combined the radius of curvature effect and the effect of the hygroscopic solute with the thermodynamic equations of the condensation process. He did not introduce numerical values and did not consider the effect of a distribution of nuclear sizes. Howell assumed a broad spectrum of nuclear sizes and several different rates of condensation. He was able to show that only a very small fraction of the nuclei were activated and that under reasonable conditions the initial supersaturation was less than one per cent. He places the maximum probable supersaturation at about three per cent. By using different size distributions of the nuclei, Howell found that the concentration of cloud drops was much more dependent on the initial rate of condensation than on the size distribution of the nuclei. As soon as the initial stages of condensation are over, the supersaturation rapidly declines so that it is extremely unlikely that any additional nuclei will be activated. Because of the rapid decrease in supersaturation it is possible that a few of the smaller nuclei which were activated may evaporate. Howell contends that although this presumably occurs, the number of drops which start to form and then evaporate is very small compared to the total. It is reasonably safe to state that the concentration of cloud particles immediately after the initial condensation is very nearly equal to the concentration of activated nuclei and that there is little likelihood of an increase in the concentration of the cloud particles thereafter.

GROWTH OF CLOUD DROPS

When the initial phase of the condensation process is completed, the dissolved hygroscopic material and the radius of curvature have a relatively small effect on the further growth of the drop. The steady-state diffusion equation for the growth of drops as given by Houghton [21] is

$$\Delta(a^2) = 8k(\rho_w - \rho_{0w})\,\Delta t, \tag{1}$$

where $\Delta(a^2)$ is the increment of the square of the drop diameter in the time Δt, ρ_w is the water vapor density at a distance from the drop, and ρ_{0w} is the density of water vapor in equilibrium with the drop. As pointed out by Howell [24], this equation is not valid during the initial condensation phase nor for very small drops

but it may be used to determine the effects of continued condensation on the size and size distribution of the cloud drops. The latent heat of condensation is released at the drop surface and transferred to the air by conduction. As a result the equilibrium temperature of the drop is greater than the air temperature and ρ_{0w} is therefore greater than the saturation water vapor density at the temperature of the air. Since $(\rho_w - \rho_{0w})$ is always positive, supersaturation must exist throughout the condensation process. Computations show that the supersaturation can hardly exceed a few tenths of one per cent for any reasonable rate of lift. As a result of the parabolic relation between the drop diameter and the time in equation (1) the drop diameters will become more uniform as the condensation proceeds. This conclusion has been verified by Howell [24].

Howell [24] attempted to find the conditions most favorable to a broad distribution of cloud drop sizes but was unable to delineate them in a clear-cut fashion. Slow cooling was expected to allow the initial effects of the solute and radius of curvature to exercise a maximum effect in broadening the size of the distribution. The small number of nuclei activated largely compensated for any such broadening. Rapid cooling had the opposite effect and it appears that some intermediate rate of cooling will yield the broadest drop-size distribution. The computed drop-size distributions are of the same general form as those observed in natural clouds, but are quite narrow, corresponding to the more homogeneous half of the clouds measured at Mount Washington, New Hampshire. It does not seem possible to explain the broader size distributions often observed by a uniform lift process of the type treated by Howell. This conclusion appears to be definite, in spite of incomplete knowledge of the condensation nucleus size spectrum, because of the controlling effect of the initial rate of condensation.

Some other mechanism must be sought to explain broader size distributions than those which result from uniform lift. Arenberg [3] suggested that turbulence would bring condensation products of different histories into the same region, thus resulting in a broad size distribution. Arenberg also suggested that the alternate up and down excursions of the drops in a turbulent atmosphere might lead to a broadening of the size distribution. However, excluding the evaporation during the descending branches of the motion, turbulence tends to narrow the size distribution and its net effect is apt to be small.

In nature the uniform lift process adopted by Howell [24] for his computations is subject to important modifications. The rate of lift of different samples of the air at the condensation level will not be the same. Because of the controlling influence of the rate of lift on the concentration and size of the cloud drops it is to be expected that the mean drop sizes of the samples will show a rather wide range. Subsequent mixing of these samples by turbulence will result in a size distribution broader than that produced by a uniform lift. It is to be noted that fine-grain turbulence is required for the final intimate mixing essential to the broadening of the local drop-size distribution. It seems probable that any observed drop-size distribution can be explained on the basis of a suitable variation in the rates of condensation of separate air parcels and their subsequent mixing. No data are available on the distribution of vertical velocities at the condensation level, so that a quantitative verification of this theory is not possible. Advective marine fogs also have broader drop-size distributions than might be expected from uniform cooling at the initiation of condensation. Such fog is formed at a much slower rate of cooling than any type of cloud. One anticipated result is that the drop concentration in advective marine fogs is much smaller than in clouds. The cooling is produced at the underlying surface and even in the usual stable stratification the mechanical turbulence is apparently sufficient to produce a range of cooling rates at the initiation of condensation. Such data as are available suggest that the drop-size distribution of radiation fog is quite narrow, presumably as a consequence of the more stagnant conditions of formation.

The breadth of the drop-size distribution in clouds has an important bearing on the stability of the clouds and on the release of precipitation in clouds which do not reach the freezing level. It is important that further studies of the factors which determine the breadth of the distribution be undertaken. It appears that knowledge of the growth of the drops from condensation nuclei to large cloud drops is now well understood from the physical point of view. The missing information is concerned with those details of the air motion which determine the initial rate of cooling and the mixing of condensation products with diverse histories.

Equation (1) shows that the time required for a drop to grow by condensation increases with the square of the drop diameter if the supersaturation is constant. This suggests that the maximum size of a condensation drop can be estimated by selecting a maximum time and a suitable supersaturation. An analysis of the drop growth process shows that the supersaturation reaches its maximum at the activation of the nucleus and thereafter rapidly declines, adjusting itself so that water will be condensed at the rate called for by the rate of lift. It is thus impossible to select an appropriate supersaturation for the computation of the maximum drop size. The factors that truly determine the maximum drop size are the drop concentration and the amount of water vapor available for condensation. The former is dependent on the initial rate of condensation, the latter on the water-vapor content of the air and the total lift. It is now known that a considerable amount of unsaturated air from the environment is entrained by the rising column in cumulus convection. This desiccates the rising air and thus reduces the maximum drop size. Large drops are favored by slow initial lift, large water-vapor content, and large total lift. These factors are not all mutually compatible, so that the actual maximum drop size is considerably smaller than that computed for optimum values of each of the separate factors. It is generally believed

that the practical maximum diameter reached by condensation is about 100 to 200 μ.

OBSERVATIONS OF DROP SIZE AND LIQUID WATER

Methods of Measurement. Although a complete discussion of methods of measuring cloud drop size and liquid-water content is beyond the scope of this article, some understanding of the measuring techniques is essential to a proper evaluation of the data. The most frequently used method of measuring cloud drop size has been the corona method. The angular diameter of the first- and higher-order diffraction rings observed around a light source is a measure of the average drop size. The method is theoretically unsound for drop diameters less than about 10 μ. Although the possibility of determining the drop-size distribution by the corona method exists, no satisfactory technique has been developed. The method is most applicable to the measurement of the drop size in homogeneous clouds; the coronas become diffuse and difficult to measure when the drop-size distribution is broad.

Mean drop size and some indication of drop-size distribution can be obtained from the rotating multicylinder. This instrument consists of a series (often five) of cylinders of different diameter arranged to be slowly rotated with the axes of the cylinders normal to the wind. The collection efficiency of a cylinder is dependent on the air speed, the cylinder diameter, and the drop diameter. The numerical relations between these quantities and the collection efficiency have been determined by Langmuir and Blodgett [35]. From a comparison of the relative collections of the several cylinders the mean drop size and a measure of the drop-size distribution may be obtained. The liquid-water content (mass of liquid water in a unit volume of air) also may be determined from the measurements. This technique has been used at Mount Washington, New Hampshire (where it was developed by Arenberg), and on aircraft. It was first used only in supercooled clouds where the deposit is in the form of rime but absorbent cylinders are now used at temperatures above freezing. In order to obtain measurable collections the cylinders must be exposed to a few miles of cloud. For this reason the method yields only average values and cannot indicate rapid changes in drop size. In order to obtain a drop-size distribution it is necessary to assume the general form of the distribution curve.

The most direct means for the measurement of drop size and drop-size distribution is the collection and photomicrography of a sample of the drops. Direct photography of the drops in the air has so many inherent difficulties that it cannot be considered to be a practicable method. The usual technique is to expose a suitably coated slide to the windstream and take photomicrographs of the collected drops. Each drop image must be measured individually, and from one hundred to several hundred drops must be so measured to secure a representative distribution curve. The slide surface is covered with a hydrophobic surface, with an oil layer, or it is smoked. If the oil film is used, the drops are immersed thus retarding evaporation and preserving their spherical shape. On the smoked slides the drops leave clear areas which are related to the drop size. The drops on the hydrophobic surface are semiflattened and are subject to evaporation. Because of the finite size of the slides there is discrimination against the smaller drops. In addition, large drops tend to fracture on impact at high air speeds. In spite of these difficulties this general method is the only one which permits the nearly instantaneous determination of the drop-size distribution.

It is clear, even from this brief discussion, that none of the present techniques is entirely satisfactory. It is not believed that improvements in any of the existing techniques will remedy this situation. A new approach to the measurement of drop size and size distribution is badly needed.

Drop-Size Measurements. Köhler [30] has reported on the results of a large number of corona measurements of drop size made at a mountain observatory in northern Norway. Similar measurements have been made by others but Köhler's work may be taken as representative. Köhler made several thousand individual measurements in mountain fog, stratus, stratocumulus, and altocumulus. The absolute range of mean drop diameter was about 5 to 70 μ. The most frequent diameter was found to be about 17.6 μ. He found that the range of the most frequent diameter was smaller for altocumulus and stratocumulus than for fog and stratus. Köhler's measurements yield no information on the size distribution, although he notes that some of the coronas were sharper than others; the sharper coronas corresponded to the narrower size distributions.

Köhler has claimed that his data show what he calls a "mass grouping" such that the sizes in a group represented by $d = d_0(2)^{n/3}$, $n = \pm 1, 2, 3$, etc., are predominant. Here d_0 is the constant modal diameter of the group and d represents the diameters of the drops composing the group. Other investigators have reported similar groupings. If this result were accepted, it would imply that there is a rather fixed drop size resulting from condensation and that other sizes are formed by the combination of drops of this initial size. This does not seem reasonable on physical grounds. No such grouping has been found in the multicylinder or microscopic data. It seems that Köhler's results must be due to some peculiarity of the corona technique and it is no longer believed that such a mass grouping exists.

A large amount of data on drop size, size distribution, and liquid-water content in clouds has been collected at Mount Washington, New Hampshire [51]. All three of the methods described above have been utilized but the bulk of the data has come from the rotating multicylinder. The mean drop diameter was found to be 13 μ, which is somewhat smaller than Köhler's 17.6 μ. The observed range of median diameter was about 5 to 40 μ. It should be noted that the multicylinder mean diameter is a volume median such that one-half of the water is in smaller drops and one-half is in larger drops. The diameter obtained from the corona method probably corresponds to the most frequent size. With the

usual type of size-distribution curve the volume-median diameter is larger than the most-frequent diameter.

As already stated, an indication of the breadth of the drop-size distribution can be obtained from the multicylinder data. The general form of the distribution curve must be assumed. The evidence available suggests that the majority of drop-size-distribution curves are of the general form assumed although there are occasional curves with multiple maxima. For convenience, nine standard volume-distribution curves have been adopted, identified by the letters A through J (I is omitted). The A distribution corresponds to complete uniformity and the succeeding letters to distributions of increasing breadth as shown in Table I.

Table I. Standard Volume-Distribution Curves Used in the Multicylinder Method

Per cent of liquid water in each group	Ratio of diameter of group to volume-median diameter for each distribution								
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
5	1.0	.56	.42	.31	.23	.18	.13	.10	.06
10	1.0	.72	.61	.52	.44	.37	.31	.27	.19
20	1.0	.84	.77	.71	.65	.59	.54	.50	.42
30	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	1.0	1.17	1.26	1.37	1.48	1.60	1.73	1.91	2.22
10	1.0	1.32	1.51	1.74	2.00	2.30	2.64	3.04	4.01
5	1.0	1.49	1.81	2.22	2.71	3.30	4.02	4.93	7.34

As will be seen from Table I, each distribution is made up of seven different drop diameters each representing the percentage of the total water indicated in the first column. The drop sizes are represented as the ratios of the drop diameter to the volume-median diameter so that the distributions may be applied to any volume-median diameter. The frequency of occurrence of the nine drop-size-distribution types at Mount Washington for the months November 1946 through May 1947 is indicated in Table II.

Table II. Occurrence of Drop-Size-Distribution Curves by Type at Mount Washington, N. H.

Distribution curve type	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
Number of occurrences	241	96	47	21	13	5	2	4	7

The predominance of narrow size distributions is striking. This is probably due in part to the high frequency of cloud-cap conditions which do not favor the nonuniform rates of lift apparently required to produce broad size distributions. For this reason Table II cannot be taken as typical of the clouds of the free atmosphere. It is also apparent from Table II that a significant number of broad distributions occur (E through J) which certainly cannot be explained on the basis of uniform lift.

Multicylinder observations were also used to compute the liquid-water content. For the winter season 1945–46 the mean liquid-water content was 0.472 g m^{-3}. The most frequent value was 0.24 g m^{-3} and the range was 0 to 1.44 g m^{-3}. There is a tendency for the higher liquid-water contents to be associated with higher temperatures. There is a similar tendency for the drop size to increase with the temperature.

An extensive series of in-flight measurements of the liquid-water content and drop size of supercooled clouds has been made by the National Advisory Committee for Aeronautics and reported by Lewis and collaborators [37, 38, 39]. The principal instruments used were the rotating multicylinder, a rotating disc icing-rate meter, and a fixed cylinder which gives a measure of the maximum drop size. The measurements were made during three winter seasons and in both the eastern and the western portions of the United States. Although identification of the cloud type was made in each case it was found that, in general, the data did not warrant a more detailed classification than the distinction between cumuliform and stratiform clouds. For three winter seasons the average volume-median drop diameter was found to be 20.5 μ for cumuliform clouds and 14.7 μ for stratiform clouds. The range of volume-median diameters was 3 to 56 μ for cumulus clouds and 3 to 50 μ for stratiform clouds. Nearly 50 per cent of the size distributions as determined from the rotating multicylinder were type A. Evidence is presented that the size-distribution data are unreliable and the authors feel that this technique is not applicable to in-flight measurements. By comparison of the simultaneous observations of volume-median diameter and maximum diameter they concluded that the clouds are more homogeneous than is indicated by the rotating multicylinder.

The liquid-water contents reported by Lewis and collaborators [37, 38, 39] ranged from 0.02 to 2.0 g m^{-3} for cumuliform clouds and from 0.01 to 0.7 g m^{-3} for stratiform clouds. The average values for the three winter seasons were found to be 0.51 g m^{-3} for cumuliform clouds and 0.134 g m^{-3} for stratiform clouds.

In considering these data it must be remembered that they are winter values and that the temperatures were always below freezing and usually markedly below. The data presented show that both the drop size and the liquid-water content tend to increase with the temperature. The mean drop diameter obtained from these in-flight measurements is significantly larger than the Mount Washington mean. Again, this is probably due to the prevalence on the mountain of cloud caps which are presumed to contain smaller drops as a result of the rapid lifting. There is no significant difference in the liquid-water observations in flight and on Mount Washington.

Diem [10] has reported the most extensive set of drop-size-distribution data from the free atmosphere. His measurements were made from aircraft by exposing a small oil-covered slide to the air stream for about $\frac{1}{50}$ sec. The slides were photomicrographed within a minute of collection. The slides undoubtedly discriminated against the smaller drops. Diem states that the collection was satisfactory down to a diameter of 3 μ but there is reason to believe that the discriminatory effect started at a somewhat larger diameter. Diem gives the most frequent drop diameters for six cloud types (Table III). Fair weather cumulus, altostratus

and stratocumulus show the sharpest distributions. The other three cloud types exhibit broad size distributions. It is interesting to note that, in general, the cloud types

Table III. Data from Composite Drop-Size-Distribution Curves
(*After Diem* [10])

Cloud type	Peak diameter (μ)	Range of curve (μ)
Dense cumulus	14.5	3–40
Fair-weather cumulus	8.5	2–20
Stratocumulus	7.9	2–24
Nimbostratus	13.2	2–42
Stratus	12.9	2–42
Altostratus	10.6	2–30

associated with precipitation have broad distributions. Very few of Diem's size distributions would fall in types *A* and *B* of Table I. The difference between Diem's data and the Mount Washington data in this respect is doubtless due in part to the different methods of measurement, but the writer feels that much of the difference is real. As pointed out above, there is good reason to believe that the conditions of formation of the Mount Washington cloud cap favor a narrow drop-size distribution. Until more data from the free atmosphere are available it seems reasonable to accept Diem's data as typical rather than that from Mount Washington.

Drop-size measurements were made in sea fog by Houghton and Radford [22] on the northeast coast of the United States. The fog drops were collected on slides with a hydrophobic surface and then photomicrographed. Sampling errors occurred for drops smaller than about 20 μ but this did not greatly affect the results in view of the relatively large drop size. The volume-median drop diameters ranged from 25 to 75 μ with an average of 45 μ. The drop-size distributions appear rather broad because of the large drop size but most of them correspond to the *C* distribution of Table I, while a few *B* and *D* distributions also occurred. The largest drop measured was 120 μ in diameter. The mean liquid-water content was found to be 0.13 g m^{-3} with a range of from 0.01 to 0.30 g m^{-3}.

The most striking feature of these results is the large size of fog drops as compared to cloud drops and the relatively small variation in the volume-median diameter and in the breadth of the size distribution. Coupled with the large drop size the relatively low liquid-water content shows that the drop concentration is very small. This suggests that the fogs observed formed on a few relatively large nuclei of condensation which quite possibly were sea-salt particles. Chemical analyses of the fog water tended to confirm this conclusion. The chloride content of the water averaged 70 parts per million and ranged from 8 to 480 parts per million. In fog water there appeared to be more sulfate ion in proportion to chloride than there was in sea water. This suggests the presence of some nuclei of industrial origin but does not prove that such products served as nuclei.

Hagemann [18] obtained drop-size distributions in fog in Germany using an adaptation of the oil-covered slide technique. He found that the most frequent size ranged from 9- to 34-μ diameter with an average of 15.6 μ. As pointed out earlier, the volume-median diameter is always greater than the most-frequent diameter, so that Hagemann's data do not differ greatly from those of Houghton and Radford [22]. Although data are not available it is to be expected that the drop size in urban fogs would be smaller than in fogs formed in relatively clean air.

Although more data on the drop-size distribution in clouds of the free atmosphere are badly needed, the information at hand is sufficient to give a good general idea of the end results of natural condensation processes. The most important conclusion is that most clouds of the free atmosphere, especially those associated with precipitation, have drop-size distributions broader than is explicable by a uniform lifting process at the condensation level. As already indicated, the most promising explanation for such a broad distribution is non-uniform rates of lift at the condensation level combined with later turbulent mixing.

THE ICE PHASE

Supercooled Water. The regular existence of supercooled water clouds in the atmosphere is now a matter of common knowledge. Water clouds are much more common than ice clouds at temperatures down to −10C and they have been observed down to −35C and possibly below. Dorsey [11] and others have shown that water in bulk may be supercooled from a few degrees to as much as 20 degrees. The temperature at which water freezes spontaneously is not known with certainty, but theoretical considerations suggest that it is in the neighborhood of −70C. Dorsey found that sealed samples of water had characteristic and reproducible freezing temperatures. The freezing temperature was found to be lower for the cleaner samples such as conductivity water than for natural water from ponds and puddles. Prolonged ageing, or heating to 97C, was found to lower the freezing temperature. Samples maintained a few degrees below their characteristic freezing temperature remained unfrozen indefinitely. Dorsey concluded that his results were best explained by the assumption that freezing is initiated by "motes" of submicroscopic size.

Rau [45] and Heverly [19] have investigated the freezing of supercooled water drops. Heverly reported that the freezing temperature was constant at −16C for drops larger than 400-μ diameter and decreased rapidly with the diameter for smaller drops with a suggestion of a minimum near −40C for very small drops. Rau repeatedly froze a group of 24 drops and found a distribution of freezing temperatures which was apparently independent of drop size. The first two or three freezings lowered the average freezing temperature which thereafter remained constant. Other unpublished investigations have yielded results similar to Rau's and it must be concluded that Heverly's findings were in some way influenced by the experi-

mental technique. Schaefer [46] has found that a supercooled fog may be converted to an ice-crystal fog by cooling a portion of it to −39C. Hollstein, quoted by Weickmann [55], studied the freezing of drops of sodium chloride solution of various concentrations. She found that the freezing point of the solution was equal to the freezing point of the water solvent minus the computed lowering of the freezing point due to the solute. There was some suggestion that the lowest possible freezing point of a sodium chloride solution lies in the neighborhood of −35C at which point the salt may act as a freezing nucleus.

Condensation and Sublimation Below 0C. The ice phase may appear in the atmosphere either by the freezing of the liquid or by the direct sublimation of the vapor to the solid phase. Wegener [54] first suggested that the atmosphere contained sublimation nuclei which would act in a fashion analogous to condensation nuclei to promote sublimation in the vicinity of ice saturation. Findeisen [13] expanded on this concept and based his precipitation theories, in part, on the existence of such nuclei. It was assumed that sublimation nuclei were small solid particles of a shape similar to an ice crystal. Sublimation should occur on a nucleus truly isomorphic with ice at ice-saturation.

The search for sublimation nuclei has been conducted by two experimental techniques. The first of these employs the expansion chamber and the second the dew-method in which the processes are observed on a chilled surface. Cwilong [7] used an expansion chamber of the Wilson-type which could be refrigerated. After the air was cleaned by repeated expansions, he found that a cloud of ice crystals was formed when the minimum temperature during the expansion fell below −41.2C. He felt that small ions were acting as sublimation nuclei below this temperature since the ice cloud formed at smaller expansions than those used to clean the air. In uncleaned outdoor air the transition temperature rose to −32.2C and to −27C when tobacco smoke was added. Cwilong also reported that at −70C there was a distinct change, in that a small shower of quite large grains of ice accompanied the cloud of crystalline dust. Fournier d'Albe [17] repeated and extended Cwilong's experiments with similar apparatus. He was unable to get the dense ice cloud in cleaned air below −41C reported by Cwilong but confirmed the latter's conclusion that only liquid drops are formed above this temperature. Fournier d'Albe also found that no ice crystals were formed until water-saturation was reached or exceeded. He concluded that the ice phase was attained via the liquid phase and suggested that the particles be called *freezing nuclei* rather than *sublimation nuclei*. He also investigated the action of several types of artificial nuclei, including silver iodide, which Vonnegut [53] had reported as causing the crystallization of supercooled clouds. Fournier d'Albe found that silver iodide nuclei caused ice crystals to appear at −7C, but only at water-saturation. Other artificial nuclei such as sodium chloride, sodium nitrate, caesium iodide, and cadmium iodide were found to have no effect on the water-ice transition temperature. It is to be noted that the latter two substances form crystals with lattice constants similar to ice. It was on this basis that Vonnegut selected silver iodide. On the other hand a water-drop cloud formed on cadmium iodide and then evaporated left nuclei on which ice formed at ice-saturation. This suggests that the nuclei of silver, caesium, and cadmium iodides used by Fournier d'Albe may not have been in crystalline form.

Findeisen and Schulz [16] used a much larger expansion chamber with a volume of 2 m^3, which was arranged to permit expansions at rates comparable to those in nature. Their results were similar to those of Cwilong and Fournier d'Albe in that the clouds formed by steady expansions were invariably water or mixed water-ice clouds even at a temperature of −40C. On some occasions when the expansion was interrupted just prior to the attainment of water-saturation, ice crystals were formed in the absence of a water cloud. At an expansion equivalent to a vertical velocity of 5 m sec^{-1} ice crystals were first observed at about −7C and their number increased with decreasing temperature. In the neighborhood of −35C a very large increase occurred. At more rapid expansions the first ice crystals appeared at lower temperatures but the sudden increase occurred at a temperature somewhat higher than −35C. All of these experiments were performed in uncleaned surface air. Very similar results were obtained by Palmer [44], who observed the formation of a few ice crystals at −22C and at a relative humidity of 97 per cent with respect to water. He also observed a rapid increase in ice crystals at −32C in natural surface air. In airplane flights, with a smaller expansion chamber, Palmer found the −32C nuclei only below the haze inversion; at higher altitudes the first crystals appeared at from −41C to −44C.

The most extensive investigations of condensation, freezing, and sublimation on a chilled surface have been made by Weickmann [55]. The advantages of this technique are that the individual particles may be viewed with a high-power microscope, that the temperature and rate of cooling may be controlled precisely, and that the supersaturation with respect to ice is known at all times. The disadvantage is that the condensation occurs on a surface rather than in the air. Weickmann showed rather conclusively that the effect of a properly cleaned surface was small. Weickmann worked mostly at or near −40C. He found that ice crystals formed only when water-saturation was approached. In one series of ten tests, using nuclei from a heated room, ice crystals formed at an average water relative humidity of 97 per cent, the range being from 93 to 104 per cent. Using the residue from evaporated drops as nuclei, a few crystals formed after 33 minutes at relative humidities of from 85 to 90 per cent with respect to water or from 120 to 130 per cent with respect to ice. At or above 100 per cent water relative humidity thousands of ice crystals formed at once. A few substances were found which favored the formation of ice at higher temperatures or lower ice-supersaturations, but in no case were crystals formed near ice-saturation.

No ice formed on soluble nuclei even at low temperatures; it appeared that solid nuclei were required.

Weickmann and others [55, 31, 32] also considered the problem from the theoretical side and showed that the structure of the ice crystal is so unique that it is extremely unlikely that any substances exist in the atmosphere which are truly isomorphic with ice. Weickman concluded that the atmospheric ice phase is formed by the freezing of the liquid on solid freezing nuclei. He felt that the freezing nuclei were often condensation nuclei with a microporous or fissured surface which promoted condensation. He conceded that sodium chloride nuclei might act as freezing nuclei at temperatures below −35C. Although he observed a few nuclei on which ice formed below water-saturation, he preferred to call all of the nuclei *freezing nuclei* rather than *sublimation nuclei.*

There are still many unanswered questions regarding the formation of ice crystals in the atmosphere, but some tentative conclusions can be formed. It seems clear that at all temperatures down to about −40C liquid condensate is more common than ice. Almost all investigators found a critical or transition temperature near −40C although mixed water-ice clouds have been reported both in the free atmosphere and in the laboratory down to at least −50C. The latter measurements may be in error and should be checked. It should not be concluded without further information that −40C is a spontaneous freezing temperature. Since all drops or crystals probably form on some type of nucleus this may be the temperature at which the small soluble condensation nuclei act as freezing nuclei. The computed spontaneous freezing temperature of −70C is based on physical constants, the values of which are not accurately known.

It is probably true that there are no atmospheric nuclei on which sublimation occurs at or below ice-saturation. It has been concluded from this that there are no true sublimation nuclei in the atmosphere except ice itself. On the other hand, ice crystals have been formed below water-saturation, although at considerable supersaturations with respect to ice. In the opinion of the writer it is not proper to require that a sublimation nucleus be active at ice-saturation. Many solid condensation nuclei do not become active until a considerable water-supersaturation is attained but they are still classed as condensation nuclei. Until more information is available it would seem preferable to call all nuclei on which ice forms below water-saturation *sublimation nuclei.* It is conceded that the deposition of the first few molecular layers on such a nucleus may not be in the form of ice, but little is known about this. There is certainly a clear physical difference between ice crystals formed in this way and those which are formed by the freezing of a liquid drop which has already attained cloud drop size. In the latter case it is evident that a freezing nucleus is involved.

On the basis of the definitions given above, it appears that a few sublimation nuclei which are active at temperatures as high as say −10C exist in the atmosphere. The failure to find these nuclei in the small expansion chambers may be attributed to their low concentration. Such low concentrations are adequate and even requisite for the release of precipitation by the ice-crystal mechanism. The evidence is that the large number of ice crystals found in surface air at −32C and in cleaned air at −41C are formed on freezing nuclei rather than on sublimation nuclei. There may also be freezing nuclei in the atmosphere which are active at much higher temperatures than −32C. In some cases these may be solid condensation nuclei, or they may be picked up by collision after the drops are formed.

PRECIPITATION PROCESSES

It has been realized for some time that precipitation elements cannot be formed by a continuation of the processes of cloudy condensation, but that other physical processes are necessary. In general, cloudy condensation leads to the formation of a high concentration of small particles. The precipitation process must convert this multitude of small particles into a smaller number of much larger elements. Since the mass of a raindrop of 1-mm diameter is one million times that of a cloud drop of 10-μ diameter, any proposed precipitation mechanism must be capable of causing the rapid combination of large numbers of cloud elements.

In a classic paper, Bergeron [4] reviewed the possible precipitation mechanisms and concluded that the only one of importance was the colloidal instability of a mixed water-ice cloud at temperatures below 0C. As is well known, the vapor pressure over water is greater than that over ice, at temperatures below freezing. The introduction of a few ice crystals into a supercooled water cloud will result in the relatively rapid growth of the ice crystals at the expense of the supercooled waterdrops. This idea was expanded and extended by others, particularly by Findeisen [13]. Various theories were offered to explain the appearance of the necessary ice crystals in the supercooled cloud. Findeisen's proposal of sublimation nuclei was once accepted as best explaining the observed phenomena but is now questioned for the reasons already discussed. The substitution of freezing nuclei would not alter Findeisen's theory in any important respect.

The ice-crystal theory of precipitation was widely accepted, since it seemed to be in accord with observational evidence. The proponents of the theory categorically stated that all moderate-to-heavy precipitation was initiated in this fashion and that, at most, only drizzle-type precipitation could fall from clouds which did not contain ice crystals. This conclusion was based largely on the observations that the precipitating clouds of middle latitudes extend above the freezing level and that much of the precipitation reaching the ground as rain is melted snow. The most common and apparently conclusive evidence is the glaciation of cumulonimbus prior to the release of precipitation.

It has now been established that many low-latitude clouds which release moderate to heavy precipitation are entirely below the freezing level. This shows that there is another mechanism for the release of precipitation but does not invalidate the ice-crystal theory of

precipitation for those clouds which extend above the freezing level. In the past the extension of a precipitating cloud above the freezing level has been taken as evidence for the operation of the ice-crystal process. In view of the existence of nonsupercooled precipitating clouds a more rigorous criterion must be adopted. It cannot be assumed that the ice-crystal process is operating unless it can be established that ice crystals and supercooled drops are coexistent.

The only other precipitation process worthy of serious consideration is the coalescence of drops in the gravitational field. If the drops are of nonuniform size, collisions will result because of their different terminal velocities of fall. The rate of growth by this process is dependent on the size and size distribution of the drops and on their concentration. Findeisen [14] studied this process in a cloud chamber and found that the resultant growth corresponded closely to what would be expected if each drop coalesced with all drops in its path. Findeisen's measurements were relatively crude and could not reveal the collection efficiency of one drop for slightly smaller drops. More recently, Langmuir [34] has computed the efficiency of collection of small drops by larger drops. The details of these computations are not presented in the reference but it is believed that the results are not completely reliable when the collecting and collected drops are of nearly the same size. In these computations it was assumed that the drops will coalesce if brought into physical contact; the collection efficiency is determined by the aerodynamic forces which tend to cause the smaller drops to follow the air streamlines around the larger drop.

An intelligent appraisal of the two precipitation processes outlined above must be predicated on a quantitative analysis. Unfortunately, few such analyses have been made, and indeed many of the requisite data are lacking. The ice-crystal theory involves a molecular diffusion process. An expression similar to equation (1), derived for the geometric shape of the ice crystal rather than of a sphere, is required to permit a quantitative discussion of this process. The solution of the diffusion equation for geometric forms approximating ice crystals has not been given. In an unpublished study, the writer has obtained a solution for a thin circular disc which might be a useful approximation to some ice-crystal forms. Under the rather severe limitations imposed by the lack of a suitable equation, only approximate results can be obtained. The time required for an ice crystal of mass equivalent to a sphere of 20-μ diameter to grow to an equivalent sphere diameter of 200 μ is of the order of 5 to 10 minutes. It was assumed that the vapor was saturated with respect to water at the optimum temperature of about -15C. To a fair degree of approximation the time required for the growth of the crystal increases with the square of the equivalent sphere diameter. Thus the time required for a crystal to grow to a mass equivalent to that of a raindrop of 1 mm diameter would be of the order of several hours. These numerical values are only approximate and are for the most favorable conditions of supersaturation with respect to ice. The ice-crystal effect is capable of producing crystals of mass comparable to drizzle elements in a few minutes but an excessive time is apparently required to form crystals of mass comparable to raindrops.

With the aid of Langmuir's computed collection efficiencies of drops by larger drops [34] it is a relatively straightforward task to compute the growth of drops by accretion in the gravitational field. Langmuir's paper is so recent that no such calculations have yet appeared in the literature. The writer has made a few preliminary calculations which will have to serve as the basis for the present discussion.[2] The problem was simplified by considering the growth of an initially somewhat larger drop falling through a homogeneous cloud. It will suffice to consider one example in which the diameter of the homogeneous cloud drops was assumed to be 20 μ, the liquid-water content 1 g m^{-3}, and the diameter of the larger drop 30 μ. The growth of the drop under these conditions is presented in Table IV. The relatively long

Table IV. Growth of a Drop of Initial Diameter 30 μ Falling through a Cloud of 20-μ Drops Containing 1 g m^{-3} Liquid Water

Drop diameter (μ)	Time (cumulative) (minutes)	Distance fallen (cumulative) (meters)
30	0	0
40	45	65
60	74	163
100	92	322
200	105	650
500	116	1475
1000	123	2675

time required for the drop to grow to 100 μ, compared with the time required for it to grow from 100 to 200 μ, is striking. The larger the falling drop is in relation to the smaller cloud drops, the more rapid the accretion process. Thus, this process is favored by a broad cloud drop-size distribution. The data in Table IV should be considered as examples and not as definitive numerical values. More refined computations, based on a typical drop-size distribution, should be made.

The two precipitation mechanisms can now be compared on the basis of the approximate numerical results presented above. The ice-crystal effect is more rapid than the collision process in the initial stages and is independent of the drop-size distribution. In the latter stages of growth the collision mechanism is more rapid than the ice-crystal effect, and may also initiate precipitation in clouds which do not contain ice crystals if the drop-size distribution is sufficiently broad. The two precipitation mechanisms taken together appear to be sufficient to explain the formation of precipitation. The writer's concept of the roles of the two processes is as follows: In all cases in which ice crystals are present, the ice-crystal process is dominant in the initiation of precipitation and in causing growth to an equivalent sphere diameter of the order of a few hun-

2. Subsequent to the preparation of this article the writer has extended these calculations. They will be found in "A Preliminary Quantitative Analysis of Precipitation Mechanisms," by H. G. Houghton, *J. Meteor.*, 7: 363–369 (1950).

dred microns. In the absence of ice crystals, the collision mechanism may initiate the precipitation process if the drop-size distribution is broad. Regardless of the process of initiation, the further growth of the precipitation elements is primarily by collision. This includes collisions of the precipitation elements with themselves as well as with cloud elements. In the case of snow, collisions between crystals are common, as is evidenced by even a casual examination of snowflakes. No process depending on the diffusion of water vapor seems to be capable of forming raindrops of 1-mm diameter and larger in the time available. Such precipitation elements must be formed by a collision mechanism. In middle latitudes, it is probable that the collisions are primarily between ice crystals, forming snowflakes which later melt into raindrops. In low latitudes, or in any situation in which a water-drop cloud of large vertical extent exists, once drops of, for example, 100-μ diameter appear, collision with the cloud drops is sufficient to explain the growth of the precipitation elements. More quantitative information on collision processes is badly needed, particularly on collisions between ice crystals and between cloud drops of nearly the same size.

The question naturally arises as to why all clouds do not ultimately release precipitation as a result of collisions between drops of unequal size. It is well known that many clouds produce precipitation which evaporates before reaching the ground, but this is not the complete answer. Langmuir's computations [34] show that for each drop size there is a minimum size of the larger drop below which no collisions will occur. For example, no drop of less than 45-μ diameter will collide with drops of 12-μ diameter. As the smaller drop diameter increases, the minimum diameter of the larger drop approaches that of the smaller drop. These results unfortunately lie in the region where the computations are least reliable. If correct, these results show that clouds composed of small drops are stable even when the drop-size distribution is broad. Diem's data [10] show that clouds such as fair-weather cumulus and stratocumulus contain smaller drops than clouds such as nimbostratus and heavy cumulus. Stratus also contains large drops but is not deep enough to yield more than drizzle. Houghton [20] and others have also suggested that a unipolar electric charge on the cloud drops might serve to inhibit collisions.

Snow. For the most part the discussion above has assumed the initial presence of a water-drop cloud. It may be that snow is often initiated in this way, but it is probable that snow also occurs in the absence of a water cloud. Also, a water cloud will not ordinarily exist for more than a short time in the presence of snow. It has been stated earlier that ice crystals do not form until saturation with respect to water is approached. Apparently few freezing or sublimation nuclei are active at temperatures above -10C, and lower temperatures are generally required. On the other hand, there is evidence that the top of a snow cloud may be warmer than -10C. This suggests that once snow is initiated, nuclei are produced which are active at higher temperatures. An important contribution to this problem was made by Findeisen [15], who found that the more delicate forms of crystals (stellar or dendritic) shed tiny splinters of ice as they fall. These splinters would serve as sublimation nuclei for new crystals at any temperature below freezing.

Approximate computations of the rate of growth of ice crystals suggest that the vapor pressure must be nearer saturation with respect to water than to ice if the observed sizes are to be attained. In this respect the process is quite different from condensation on liquid drops, where the vapor is only very slightly supersaturated with respect to the condensed phase. This difference is due to the much smaller number of ice crystals. The supersaturation increases as required to cause sublimation to proceed at the rate prescribed by the lifting, saturation with respect to water setting the upper limit.

Nakaya and collaborators in Japan [40, 41, 42] have made outstanding contributions to our knowledge of the formation of snow crystals. In one paper Nakaya and Terada [42] have presented useful data on the mass, physical dimensions, and velocity of fall of several types of natural snow crystals. In general, the maximum masses of the crystals are equivalent to a solid sphere of several hundred microns diameter while the velocities of fall are much smaller than those of solid spheres of equivalent mass. Nakaya and collaborators [40, 41] have succeeded in producing in the laboratory all of the types of crystals observed in nature. The two fundamental parameters determining the crystal type are temperature and degree of supersaturation. In general, the more compact crystal forms such as columns, prisms, and plates are formed at low supersaturations and the more open types such as needles and stellar or dendritic crystals are formed at the higher supersaturations. The dependence on temperature was not clearly established in the references, all of which were published before World War II. Dr. Nakaya was able to continue his researches during and after the war, but these results have been published only in Japanese. However, he has prepared all of his material in book form in English, and early publication is anticipated.

Weickmann [55] collected and photographed ice crystals in the free atmosphere. He summarized his observations as follows: In the lower troposphere, the nimbostratus region, there is slight ice supersaturation, the temperature ranges from 0 to -15C, and the crystals are in the form of thin plates and stars; in the middle troposphere or the altocumulus and altostratus region, there is moderate ice supersaturation, the temperature ranges from -15 to -30C, and the crystals are mainly thick plates and prisms; in the cirrus region or the upper troposphere, the temperature ranges from -30 to -60C, and the ice crystals are principally hollow prisms often combined as twins or clusters.

Size of Raindrops. All of the published data on raindrop size were obtained at the surface. A considerable number of such measurements have been published but it will suffice to refer to the rather recent measurements of Laws and Parsons [36]. They used the flour-pellet technique and found that the volume-median diameter

increased with rainfall intensity. For a rainfall rate of 0.05 in. hr^{-1} they give a volume-median diameter of 1.1 mm with a maximum diameter of 4 mm; for 0.5 in. hr^{-1} the volume-median diameter was 1.9 mm and the maximum diameter was 5.5 mm; for 4.0 in. hr^{-1} the volume-median diameter was 2.8 mm and the maximum diameter was 6.7 mm. The size and size distribution of raindrops can be expected to change from the cloud base to the ground because of evaporation, coalescence of drops, the variation of time of fall with drop size, and the fracture of large drops. For studies of the precipitation process it would be necessary to measure the raindrop size in and immediately below the cloud.

Hail. Hail is a special type of frozen precipitation associated with thunderstorms and characterized by extreme sizes much in excess of those of any other precipitation elements. A typical hailstone exhibits an onionlike structure when dissected, which has given rise to the belief that hailstones are formed by repeated excursions above and below the freezing level whereby successive layers of water are frozen onto it. A natural consequence of this theory was the inference that vertical velocities equalling the free-fall velocities of the hailstones exist in the atmosphere (over 100 mph in some cases).

It is now generally accepted that the major growth of the hailstone is by the collection of supercooled water as the stone falls relative to the cloud. In the earlier stages the stone may well travel in a very irregular fashion, up as well as down, but the largest hailstones can hardly be supported by updrafts. The growth of the hailstone is essentially the same as the accretion of ice on aircraft. The layer structure is due to inhomogeneities in the turbulent cloud. A similar layer structure is commonly observed in ice deposits on aircraft. Large hailstones are favored by high vertical velocities, a large vertical extent of the supercooled portion of the cloud, and high liquid-water content. Schumann [47] has shown that the extreme values of these parameters associated with cumulonimbus can lead to hailstones of the observed sizes. Hailstones large enough to damage aircraft have been reported in the clear air surrounding a thunderstorm, suggesting that stones for example, of one or two centimeters diameter may be discharged from the tops of thunderclouds. The terminal velocities of hailstones of 1 and 2 cm diameter are about 12 and 16 m sec^{-1} respectively. Extreme thunderstorm updrafts may readily exceed these velocities. Regardless of their direction of travel with respect to the earth, the hailstones are moving downward through the supercooled water drops at their terminal velocities. The thunderstorm updraft serves primarily to increase the length of the hailstone's path through the cloud and thereby to increase its collection of ice.

ARTIFICAL DISSIPATION OF FOG

Although fog hampers all types of transportation, its effects on air transportation are most serious. Much of the impetus for the development of methods for artificial dissipation of fog has come from aviation interests. Many unsuccessful attempts have been made to dissipate fog, failure being due in most cases to a lack of understanding of the problem. Some successful experiments were not followed up because it was believed that instrument-landing systems would make fog dissipation unnecessary. With the advent of World War II the problem became acute, and the British developed a thermal method now called "Fido." This system has been developed further, since the war, both in England and in the United States where an operational installation has recently been made at a commercial air field. In spite of continued advances in instrument-landing systems, most operators still feel that a cleared region for the final touchdown would greatly increase the safety of the landing. It seems evident that instrument-landing techniques will eventually be developed to the point where fog dissipation is completely unnecessary, but the writer would not care to speculate as to when this will come to pass.

In general, fog can be dispelled by the evaporation of the drops or by the physical removal of the drops from the air. Most of the methods for accomplishing this have been discussed critically by Houghton and Radford [23]. Those methods which were considered reasonably feasible were (1) the direct application of heat, (2) the use of hygroscopic materials to "dry" the air, and (3) the dropping of electrically-charged particles through the fog. The first method is exemplified by "Fido," in which heat is released by the burning of oil in long lines of burners on either side of the runway. The second method was used successfully by Houghton and Radford. The third method was experimented with by Warren, who dropped charged sand on clouds with occasional success.

There is now no doubt that fog can be dispelled by artificial means. Further experimentation with methods already proved practicable is desirable and there is still room for new ideas in both methods and equipment. The basic problem lies in the economics of fog dispersal. The mass of suspended water to be dealt with in even a relatively small volume is large, and it is inevitable that relatively large expenditures of energy will be required to remove it. As an example, the mass of water over an airport runway 6000 ft long and 300 ft wide to a height of 200 ft is about one to two tons, depending on the liquid-water content of the fog. In order to dissipate the fog by evaporation it is necessary both to supply the latent heat of vaporization and to lower the relative humidity of the air to cause the drops to evaporate rapidly. It is generally necessary to reduce the relative humidity to 90–95 per cent to meet the latter requirement. The heat energy required to evaporate the waterdrops and to reduce the relative humidity to 90 per cent in a fog at 10C, containing 0.1 g m^{-3} of liquid water, is about 559 cal m^{-3}. Of this amount, nearly 500 cal m^{-3} are required to reduce the relative humidity of the air.

If we use the figures above, a total of some 5.7×10^9 cal would be required to clear the fog in the volume assumed above. This rather impressive number of calories can be supplied by burning the modest amount of 250 gal of oil at a combustion efficiency of about 70

per cent. The computed energy requirement is for optimum conditions which cannot be realized in practice. Even a light wind steadily brings in more fog, so that the heating must be continuous. It is not practicable to apply the heat uniformly, and in practice much of the heat is wasted in raising the temperature of some of the air much higher than is necessary. The concentrated heat sources produce convection currents which may carry the heat to unwanted heights and suck in additional fog from the sides. For these reasons the practical energy requirements are many times the computed minimum. Working installations are designed to burn on the order of 100,000 gal of oil per hour.

Methods utilizing hygroscopic materials to dissipate fog are also evaporative processes and the basic energy requirements are the same as for the heating methods. Experiments reported by Houghton and Radford [23] indicate that operating systems require from five to ten times the theoretical minimum quantities of hygroscopic material.

Methods of physical removal of the fog drops, such as the charged-sand process, have smaller theoretical energy requirements than the evaporation methods. No basic energy requirement comparable to that given for the evaporative methods can be set and no experimental values are available.

The two evaporative methods which have been subjected to full-scale tests, namely the burning oil and hygroscopic material methods, are demonstrably capable of producing clearings of useful size. In both cases the costs of operation are relatively high and extensive installations are required. The use of hygroscopic materials involves the hazards of corrosion and damage to electrical equipment, although these may be nearly eliminated by proper design. The existence of large oil burners along the sides of the runway is a potential hazard which can also be minimized by proper design. Because of cost and other practical considerations only limited clearings are feasible, so that auxiliary instrumental methods must be available to guide the aircraft into the clearing. Neither method can deal with other conditions of poor visibility, such as dense snow, smoke, and dust. It must be concluded that fog dissipation by these methods is economically marginal and that installations are justifiable only in locations of extreme fog frequency or for urgent military purposes.

The fact that all proved methods of fog dissipation require much more energy than the theoretical minimum offers some hope that more efficient methods can be found. There is certainly room for further work on methods such as the electrified-sand technique where there is reason to believe that the energy requirements are more modest. The ever-recurrent hope that a method will be found which will clear large areas of fog with the expenditure of small amounts of energy is incompatible with physical reality.

CONCLUSIONS AND RECOMMENDATIONS

Our present understanding of the physics of condensation and precipitation is incomplete in many important areas. The writer has attempted to point out deficiencies in each phase of the subject during the detailed discussion. It is hoped that this will be of value to the reader, but it is felt that a more general appraisal of the situation is in order. In particular, an attempt will be made to present the writer's views as to the relative importance of those phases of the subject which need further study. This requires a decision as to the most important contribution to meteorology as a whole which can be expected from the study of cloud physics. The author feels that the complete explanation of precipitation should be the dominant aim of the cloud physicist. Of the elements forecast, precipitation is probably the most important to the majority of people. Further, the complete explanation of the precipitation process involves a knowledge of most of the topics in cloud physics. In making this decision the writer is acutely aware of the possibility that some new discovery will necessitate a refocusing of the entire cloud physics program.

Knowledge of condensation nuclei and of condensation in the liquid phase is incomplete in detail but is relatively satisfactory as compared to other parts of the field. The most important problem in this area is the investigation of the factors determining the breadth of the drop-size distribution. Knowledge of the ice phase in the atmosphere is inadequate. It is imperative that the nature and mode of action of freezing nuclei and sublimation nuclei be determined. This information is essential to an evaluation of and practical utilization of the ice-crystal theory of precipitation. It is equally imperative that a complete study be made of the growth of drops by collision in the gravitational field.

It is the writer's opinion that these problems should be attacked experimentally both in the laboratory and in the free atmosphere. The mechanisms of phase changes can be studied only in the laboratory, and the collision process is also a proper subject for laboratory investigation. However, the most complete laboratory study will not tell us what is happening in the atmosphere. It is therefore essential that flight measurements be made to determine which precipitation process operates under various conditions and to obtain a quantitative verification of the operation of the assumed processes. No adequate instrumentation is available for flight observations, and instrument development is therefore an essential part of the program. Flight measurements are extremely costly and time-consuming and should not be embarked on without careful planning and adequate instrumentation.

REFERENCES

1. Adel, A., "Note on the Atmospheric Oxides of Nitrogen." *Astrophys. J.*, 90: 627 (1939).
2. Aitken, J., *Collected Scientific Papers.* Cambridge, University Press, 1923.
3. Arenberg, D. L., "Turbulence as the Major Factor in the Growth of Cloud Drops." *Bull. Amer. meteor. Soc.*, 20: 444–448 (1939).
4. Bergeron, T., "On the Physics of Cloud and Precipitation." *P. V. Météor. Un. géod. géophys. int.*, Pt. 2, pp. 156–178 (1935).

5. Boylan, R. K., "Atmospheric Dust and Condensation Nuclei." *Proc. R. Irish Acad.*, 37(A): 58-70 (1926).
6. Coste, J. H., and Wright, H. L., "The Nature of the Nucleus in Hygroscopic Droplets." *Phil. Mag.*, Ser. 7, 20: 209-234 (1935).
7. Cwilong, B. M., "Sublimation in a Wilson Chamber." *Proc. roy. Soc.*, (A) 190: 137-143 (1947).
8. Dessens, H., "The Use of Spiders' Threads in the Study of Condensation Nuclei." *Quart. J. R. meteor. Soc.*, 75: 23-26 (1949).
9. —— "Les noyaux de condensation de l'atmosphère." *C. R. Acad. Sci., Paris*, 223: 915-917 (1946).
10. Diem, M., "Messungen der Grösse von Wolkenelementen II." *Meteor. Rdsch.*, 1: 261-273 (1948).
11. Dorsey, N. E., "The Freezing of Supercooled Water." *Trans. Amer. phil. Soc.*, Vol. 38, Pt. 3, pp. 247-328 (1948).
12. Findeisen, W., "Die Kondensationskerne. Entstehung, chemische Natur, Grösse und Anzahl." *Beitr. Phys. frei. Atmos.*, 25: 220-232 (1939).
13. —— "Die kolloidmeteorologischen Vorgänge bei der Niederschlagsbildung." *Meteor Z.*, 55: 121-133 (1938).
14. —— "Messungen der Grösse und Anzahl der Nebeltropfen zum Studium der Koagulation inhomogenen Nebels." *Beitr. Geophys.*, 35: 295-340 (1932).
15. —— und Findeisen, E., "Untersuchungen über die Eissplitterbildung an Reifschichten." *Meteor. Z.*, 60: 145-154 (1943).
16. Findeisen, W., und Schulz, G., "Experimentelle Untersuchungen über die atmosphärische Eisteilchenbildung I." *Forsch. ErfahrBer. Reichs. Wetterd.*, Reihe A, Nr. 27, (1944).
17. Foubnier d'Albe, E. M., "Some Experiments on the Condensation of Water Vapour at Temperatures below 0°C." *Quart. J. R. meteor. Soc.*, 75: 1-14 (1949).
18. Hagemann, V., "Eine Methode zur Bestimmung der Grösse der Nebel- und Wolkenelemente." *Beitr. Geophys.*, 46: 261-282 (1936).
19. Heverly, J. R., "Supercooling and Crystallization." *Trans. Amer. geophys. Un.*, 30: 205-210 (1949).
20. Houghton, H. G., "Problems Connected with the Condensation and Precipitation Processes in the Atmosphere." *Bull. Amer. meteor. Soc.*, 19: 152-159 (1938).
21. —— "A Study of the Evaporation of Small Water Drops." *Physics*, 4: 419-424 (1933).
22. —— and Radford, W. H., "On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 6, No. 4, 31 pp. (1938).
23. —— "On the Local Dissipation of Natural Fog." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 6, No. 3, 63 pp. (1938).
24. Howell, W. E., "The Growth of Cloud Drops in Uniformly Cooled Air." *J. Meteor.*, 6: 134-149 (1949).
25. Junge, C., "Übersättigungsmessungeu an atmosphärischen Kondensationskernen." *Beitr. Geophys.*, 46: 108-129 (1936).
26. —— "Zur Frage der Kernwirksamkeit des Staubes." *Meteor. Z.*, 53: 186-188 (1936).
27. Köhler, H., "The Nucleus in and the Growth of Hygroscopic Droplets." *Trans. Faraday Soc.*, 32: 1152-1161 (1936).
28. —— "Über die Kondensation an verschieden grossen Kondensationskernen und über die Bestimmung ihrer Anzahl." *Beitr. Geophys.*, 29: 168-186 (1931).
29. —— "Zur Thermodynamik der Kondensation an hygroskopischen Kernen und Bemerkungen über das Zusammenfliessen der Tropfen." *Medd. meteor.-hydr. Anst. Stockh.*, Vol. 3, No. 8 (1926).
30. —— "Untersuchungen über die Elemente des Nebels und der Wolken." *Medd. meteor.-hydr. Anst. Stockh.*, Vol. 2, No. 5 (1925).
31. Krastanow, L., "Beitrag zur Theorie der Tropfen- und Kristallbildung in der Atmosphäre." *Meteor. Z.*, 58: 37-45 (1941).
32. —— "Über die Bildung der unterkühlten Wassertropfen und der Eiskristalle in der freien Atmosphäre." *Meteor. Z.*, 57: 357-371 (1940).
33. Landsberg, H. E., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, Vol. 3 (Supp.) pp. 155-252 (1938).
34. Langmuir, I., "The Production of Rain by a Chain Reaction in Cumulus Clouds at Temperatures above Freezing." *J. Meteor.*, 5: 175-192 (1948).
35. —— and Blodgett, K. B., *A Mathematical Investigation of Water Droplet Trajectories*. U. S. Army Air Forces Technical Report No. 5418, 1946.
36. Laws, J. O., and Parsons, D. A., "The Relation of Raindrop Size to Intensity." *Trans. Amer. geophys. Un.*, Pt. 2, pp. 452-460 (1943).
37. Lewis, W., "A Flight Investigation of the Meteorological Conditions Conducive to the Formation of Ice on Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1393, 34 pp. (1947).
38. —— Kline, D. B., and Steinmetz, C. P., "A Further Investigation of the Meteorological Conditions Conducive to Aircraft Icing." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1424, 18 pp. (1947).
39. Lewis, W., and Hoecker, W. H., Jr., "Observations of Icing Conditions Encountered During 1948." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1904, 43 pp. (1949).
40. Nakaya, U., Toda, Y., and Maruyama, S., "Further Experiments on the Artificial Production of Snow Crystals." *J. Fac. Sci. Hokkaido Univ.*, Ser. 2, 2: 13-57 (1938).
41. Nakaya, U., Sato, I., and Sekido, Y., "Preliminary Experiments on the Artificial Production of Snow Crystals." *J. Fac. Sci. Hokkaido Univ.*, Ser. 2, 2: 1-11 (1938).
42. Nakaya, U., and Terada, T., Jr., "Simultaneous Observations of the Mass, Falling Velocity and Form of Individual Snow Crystals." *J. Fac. Sci. Hokkaido Univ.*, Ser. 2, 1: 191-200 (1935).
43. Owens, J. S., "Condensation of Water from the Air upon Hygroscopic Crystals." *Proc. roy. Soc.*, (A) 110: 738-752 (1926).
44. Palmer, H. P., "Natural Ice-Particle Nuclei." *Quart. J. R. meteor. Soc.*, 75: 15-22 (1949).
45. Rau, W., "Gefriervorgänge des Wassers bei tiefen Temperaturen." *Schr. dtsch. Akad. Luft.*, 8: 65-84 (1944).
46. Schaefer, V. J., "The Formation of Ice Crystals in the Laboratory and the Atmosphere." *Chem. Rev.*, 44: 291-320 (1949).
47. Schumann, T. E. W., "The Theory of Hailstone Formation." *Quart. J. R. meteor. Soc.*, 64: 3-17 (1938).
48. Shaw, J. H., Sutherland, G. B. B. M., and Wormell, T. W., "Nitrous Oxide in the Earth's Atmosphere." *Phys. Rev.*, 74: 978-979 (1948).
49. Simpson, G. C., "Sea-Salt and Condensation Nuclei." *Quart. J. R. meteor. Soc.*, 67: 163-169 (1941).
50. Thomson, Sir W., "On the Equilibrium of Vapour at a Curved Surface of Liquid." *Phil. Mag.*, Ser. 4, 42: 448-452 (1871).
51. U.S. Air Force, Air Materiel Command, *Harvard-Mount Washington Icing Research Report 1946-1947*. Air Force Tech. Rep. No. 5676, 802 pp., Wright-Patterson Air Force Base, Dayton, Ohio, June 22, 1948.

52. Volmer, M., *Kinetik der Phasenbildung.* Leipzig, T. Steinkopff, 1939.
53. Vonnegut, B., "Nucleation of Supercooled Water Clouds by Silver Iodide Smokes." *Chem. Rev.*, 44: 277–289 (1949).
54. Wegener, A., *Thermodynamik der Atmosphäre.* Leipzig, J. A. Barth, 1911.
55. Weickmann, H., "Die Eisphase in der Atmosphäre." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 6, 54 SS. (1949).
56. Wigand, A., "Über die Natur der Kondensationskerne in der Atmosphäre insbesondere über die Kernwirkung von Staub und Rauch." *Meteor. Z.*, 30: 10–18 (1913).
57. Woodcock, A. H., and Gifford, Mary M., "Sampling Atmospheric Sea-Salt Nuclei over the Ocean." *J. mar. Res.*, 8: 177–197 (1949).
58. Wright, H. L., "Atmospheric Opacity at Valentia." *Quart. J. R. meteor. Soc.*, 66: 66–77 (1940).

NUCLEI OF ATMOSPHERIC CONDENSATION*

By CHRISTIAN JUNGE

Meteorological Institute for Northwestern Germany

The Condensation Effect of the Nuclei

According to Wall [38], the condensation effect of a nucleus can be compared to that of a droplet of pure water on which, because of the surface curvature, condensation will take place only when there is a definite amount of supersaturation (Thomson's equation). The following distinctions between different types of nuclei can then be made:

1. Particles insoluble in water and *unwettable*. These may serve, as tests have shown [20], as nuclei of condensation. They require (when spherical) a larger amount of supersaturation than do droplets of pure water. Their importance for the natural aerosol is slight.

2. Particles insoluble in water, but *wettable*. According to the existing humidity and to their degree of wettability, these particles are surrounded by one or more molecular layers [40] of adsorbed water. They begin to condense at, or somewhat below, the value given by Thomson's equation. If these particles are of irregular or flaky structure, condensation begins before saturation, either in cavities (as capillary condensation) or in porous nuclear substances (as absorption of water).

3. Droplets of solutions. These form an important group of nuclei. Owing to the dissolved substance, the degree of supersaturation necessary for condensation falls to as low a value as ⅓ of that indicated by Thomson's formula [25].

Between particles of types 2 and 3 there are many intermediate types, because, as a result of coagulation, many nuclei will contain both soluble and insoluble matter (mixed nuclei). Figure 1 shows the relationship between the radius of the nuclei and the amount of supersaturation necessary for condensation to occur. It is evident that, in the presence of condensation nuclei, the range of supersaturation necessary for the formation of clouds extends from 0 to about 20 per cent.

The processes of condensation on nuclei at temperatures below 0C, which have recently been the object of thorough research, will be mentioned only briefly here, because they represent a transition into the field of cloud physics [27]. Weickmann [41] found that the nuclei act fundamentally as condensation nuclei for the *liquid* phase even at temperatures below freezing, that is, only at saturation with respect to *water* does condensation take place in the form of droplets, some of which may subsequently freeze. Sublimation nuclei, in the sense of Bergeron-Findeisen, which may form ice particles already at supersaturation with respect to ice, seem to exist only in negligible quantities, if at all.

According to Lafargue [22], the droplets initially formed freeze at about −41C in the size range between 1 μ and 20 μ, rather independently of the presence of dissolved substances. Heverly [13] found that this spontaneous freezing point rises to −16C for drops of 0.4 mm diameter and then remains constant for drops up to about one mm diameter, likewise largely independently of the source of the water. However, according to Weickmann, the presence of certain solid particles, so-called freezing nuclei, appears to modify these processes by raising the freezing point. This modification depends on the size and nature of the freezing nuclei.

Size, Number of Nuclei, and Methods of Measurement

If we disregard the small ions (radius $r \approx 10^{-7}$ cm), which are not of interest here, the size range of condensation nuclei extends approximately from $r = 4 \times 10^{-7}$ to 10^{-4} cm (Fig. 1). Although all types of particles may be active as condensation nuclei in a nuclei counter (see next paragraph and [20]), only those nuclei which require the lowest degree of supersaturation (*i.e.*, haze droplets and large nuclei droplets) are active in the actual atmosphere. It is within this group of particles —recently re-examined by Dessens [8, 9] and Woodcock and Gifford [43]—that we find the real meteorological condensation nuclei. By what method, now, can we measure these particles which range in size over three orders of magnitude?

All nuclei, whether hygroscopic or even unwettable, with radii ranging from about 4×10^{-7} to 2×10^{-5} cm, are counted in *nuclei counters*, so called after Aitken [3, 23]. The lower limit of this range is determined by the ratio of expansion; however, it appears that the values of supersaturation computed from this ratio are substantially too high [17]. The upper limit—which is highly uncertain—can be established with a certain degree of probability by assuming that larger nuclei droplets grow so rapidly in the saturated air of the counting chamber that they are probably precipitated before the actual measurement (Fig. 3). This seems to be corroborated by the photographs of condensation nuclei taken with an electron microscope, as described by Linke [26]. For these photographs, the nuclei were precipitated on an object screen by condensation in the nuclei counter and show an upper limit at $r = 10^{-5}$ cm, which corresponds to droplets of solution of $r \approx 2 \times 10^{-5}$ cm (Fig. 1). However, the aerosol, which was taken from a room, undoubtedly contained larger particles. It is possible that these limits are subject to variations when different types of counters, or even different instruments of the same type are used; this would explain the discrepancies recently found by Israël and Krestan [17] when they made comparative measurements with different nuclei counters.

While the nuclei counter gives only the number of nuclei, the method developed mainly by Israël [18]

* Translated from the original German.

permits the measurement of the size distribution of these particles. However, such an "ion spectrum" is not equivalent to the "nuclei spectrum," because:

1. The proportion of nuclei with an electrical charge depends upon their size.

2. Ions have multiple charges when the number of nuclei is small and their size large (see pp. 187 f.).

It is therefore not permissible to infer the number of particles from the number of electrical unit charges. Figure 1 gives examples of such ion spectra with the terminology introduced by Israël. The influence of a counter is correctly computed from the expansion [17]. Observations of the fall velocity of the nuclei [23] as well as of the optical properties of haze [35, 46, 48] establish only mean nuclei sizes. Photographs taken with the electron microscope appear to be most promising. Here, however, it must be considered that during the exposure all the water and, to some extent, also the solid substances evaporate because of the high vacuum in the microscope and the heat generated by the electron beam. For this reason, the residues that are photographed represent only minimal sizes.

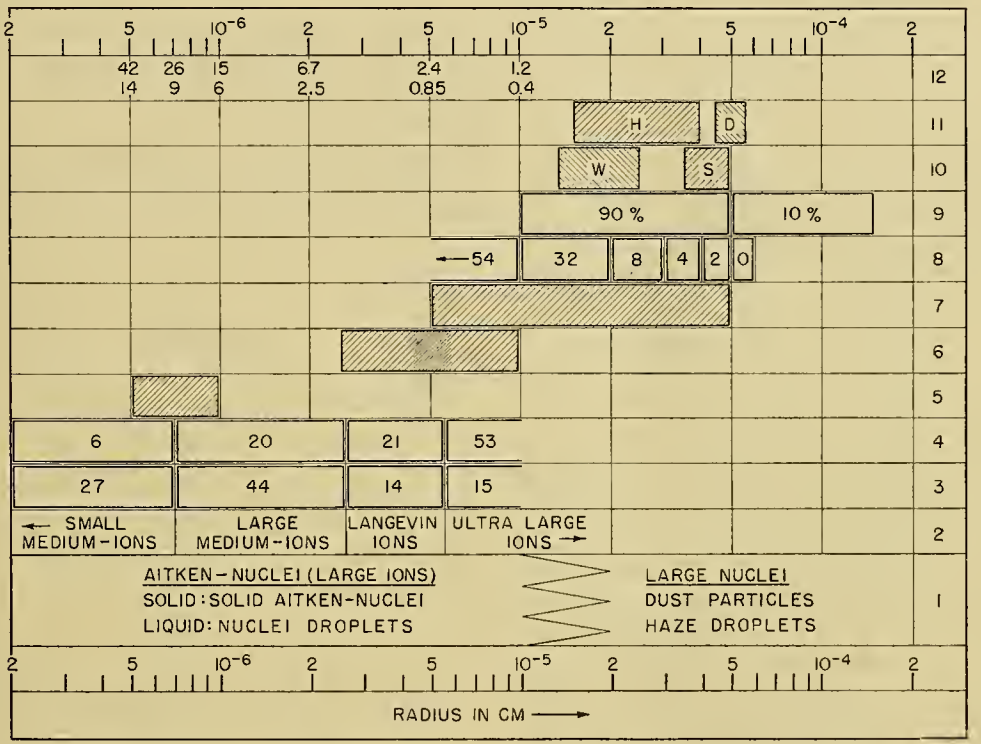


FIG. 1.—A survey of the size range of atmospheric condensation nuclei.

1. Notation used in the text for the various types of nuclei.
2. Classification of ions according to Israël [18].
3. Distribution of the ions in per cent of the total as found in Bad Gastein [18].
4. Distribution of the ions in per cent of the total as found in Frankfort-on-the-Main [18].
5. Size range of newly formed gas-flame ions [19].
6. Size range of the Aitken-nuclei according to photographs taken with an electron microscope [26].
7. Size range of soot, crystals of cigarette smoke, etc., according to photographs taken with an electron microscope [41].
8. Per cent distribution of haze droplets as found by Dessens [8].
9. Size distribution of atmospheric dust [10].
10. Predominant sizes of the dust particles in winter (W) and summer (S) [33].
11. Size of haze droplets (H) and dust particles (D) of preferential optical effectiveness, according to Siedentopf [35].
12. Supersaturation values necessary for the condensation on pure water droplets (above) and droplets of solutions (below) in per cent relative humidity.

large city is clearly shown by the high content of large ions. The upper limit of resolution of the ion spectra is at $r \approx 10^{-5}$ cm.

There have been other attempts to determine the size distribution of nuclei. For example, by measuring the supersaturation necessary for condensation [19], it is possible to infer the "size spectrum of the nuclei" from the "supersaturation spectrum," under the assumption that, as with droplets of solution, there is an unequivocal relationship between nuclei size and supersaturation and that the supersaturation in the nuclei

The range of particles up to about $r = 2 \times 10^{-5}$ cm may be designated here as *Aitken-nuclei*. For the time being there is no method available for determining the proportion of nuclei droplets and solid nuclei in this size range. However, from the growth of these particles with increasing humidity (see pp. 184 f.) we may infer that there is a preponderance of nuclei droplets.

For reports on the results of nuclei counts we may refer to the excellent monographs by Landsberg [23] and Burckhardt and Flohn [3]. The nuclei concentrations vary within a wide range; absence of nuclei, how-

ever, has seldom been found. Tables I and II may serve as a guide.

The large nuclei (those with radii greater than about 1 to 2 $\times$ 10^{-5} cm) are above the resolution limits of the optical microscope. In this category only solid particles have been measured so far, in the well-known dust counters of Owens and Zeiss. It was only recently that Dessens, and Woodcock and Gifford succeeded in observing the haze droplets, thus closing a gap in our methods of observation. Dessens [9] moved fine spider threads through the air, thus capturing the haze droplets quantitatively. It is uncertain whether or not solid particles can be collected in this way. Woodcock and Gifford [43] deposited haze droplets on small glass plates and showed that over the ocean practically all these droplets consist of *NaCl*-solutions. Dust and haze-droplet counters enable us to determine both the total number of suspensions and their size distribution (Fig. 1).

Table I. Horizontal Distribution of Aitken-Nuclei [23]

Location	Number of		Nuclei per cc				
	localities	observations	Average	Mean maximum	Mean minimum	Absolute maximum	Absolute minimum
Large city	28	2500	147,000	379,000	49,100	4,000,000	3,500
Town	15	4700	34,300	114,000	5,900	400,000	620
Open country	25	3500	9,500	66,500	1,050	336,000	180
Ocean	21	600	940	4,680	840	39,800	2

Table II. Mean Vertical Distribution of Aitken-Nuclei (Data from 28 Balloon Ascensions [23])

Altitude (km)	0–0.5	0.5–1	1–2	2–3	3–4	4–5	5
Nuclei per cc	22,300	11,000	2,500	780	340	170	80

The dust concentrations show—particularly in industrial areas—a close correlation with the Aitken-nuclei concentrations [44], which is not surprising since both types of particle often have the same source. The concentration of the dust particles and haze droplets over land amounts to about 10–200 cm^{-3}, that of the haze droplets over the ocean to about 1–10 cm^{-3} [43].

From the foregoing discussion we can conclude that it is not yet possible to measure the nuclei spectrum in its entirety; so far there is only the substitute of counting nuclei, ions, dust, and haze droplets simultaneously.

Form and Physical State of Nuclei

Photographs of Aitken-nuclei taken on the stage of an electron microscope show irregular clusters of molecules interspersed with a few hexagonal crystal outlines [26] that can be attributed to hexagonal or cubic crystals. Soot, for example, consists of loosely connected platelets; cigarette smoke, of small crystals [41]. It is likely, however, that these are essentially the boiled-down residues of nuclei droplets.

Previous examinations of samples collected with dust counters have revealed a similar picture. These samples showed, in addition to various crystalline shapes (*e.g.*, cubic forms) and some organic substances, a preponderance of irregular particles. However, according to recent investigations by the author, these dust particles, when deposited on suitable surfaces, consist to a considerable extent of droplets. This is to be expected from Dessens' observations. The proportion of droplets varies greatly and increases with humidity. Many of these droplets contain both soluble and insoluble substances. It is conceivable that the Aitken-nuclei have a similar composition, since the mixed nuclei among them constitute an even larger portion owing to the greater coagulation, and all intermediate types between pure water droplets and solid particles covered with a hygroscopic film probably exist.

Important clues to the physical state of nuclei may be deduced from the study of their growth with increasing humidity. Figure 2 shows a few typical results

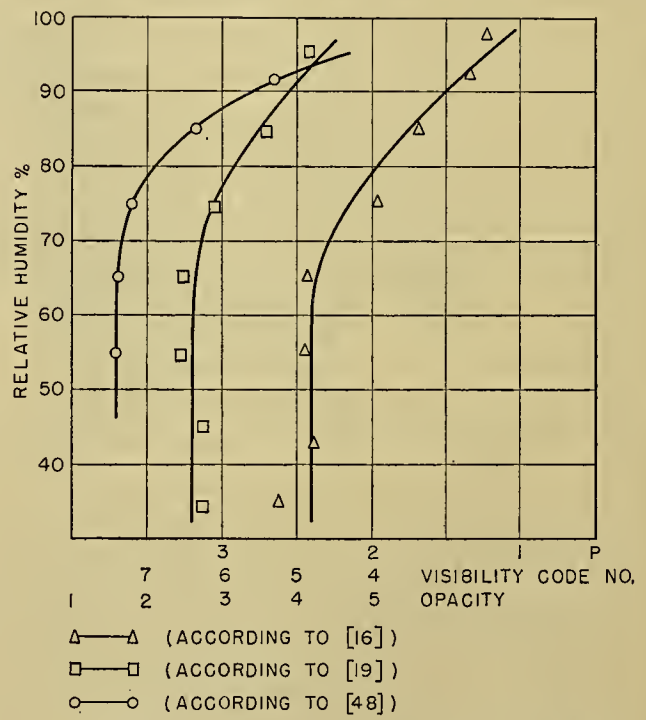


Fig. 2.—Effect of the growth of nuclei with increasing relative humidity on the opacity (circles), on the visibility (squares), and on the ratio $P = \frac{\text{total number of nuclei}}{\text{total number of large ions}}$ (triangles).

of observations of various quantities related to natural aerosols. The dependence of visibility on relative humidity (see pp. 187 f.) reflects primarily the growth of the large nuclei ($r > 10^{-5}$ cm) with humidity, whereas the dependence of P on humidity (see pp. 187 f.) chiefly illustrates the growth of nuclei in the size range below $r = 10^{-5}$ cm. A few isolated, direct size measurements lead to the same conclusions (Fig. 3, curve 1). From these results we see that as a general rule, the

growth of the nuclei, regardless of their size and origin, starts only after a relative humidity of approximately 70 per cent has been attained, and that this growth accelerates as the humidity approaches saturation value.

The interpretation of this principle of growth is not devoid of difficulties. A comparison of the curves of growth of different-sized solution droplets, computed in the usual way [25], shows reasonably good agreement above a relative humidity of 70 per cent. Wright [48] attributes the deviation at humidities below this value to the crystallization of *NaCl* which he believes to be predominant. However, Simpson [37] and Wall [39] correctly point out that in the case of crystallization the size of the droplets (Fig. 3, curve 8*a*) is bound to decrease in a discontinuous manner until all the water is evaporated. Dessens [8] has been able to observe this process directly; he noted that the solution droplets may show a supersaturation which is greater the smaller the droplets, and that the ensuing violent condensation sometimes results even in a shattering of the crystal. The renewed dissolution of these crystals will then take place at higher degrees of humidity. There is, however, no evidence for such abrupt size changes of the nuclei at intermediate humidities.

On the other hand, there is a fairly close approximation to the observations with mixed nuclei, that is, droplets of solution with a more or less large content of insoluble substances. Figure 3 gives a few computed curves of growth with hatched markings indicating the radii of the solid portions of the nuclei. Below a relative humidity of 70 per cent, the solid nucleus is covered by only a thin solution film, with the radius remaining almost constant.

Wall [40] pointed out that insoluble but wettable surfaces may adsorb multimolecular layers of water at relatively high humidities, a process that may become significant considering the original smallness of the nuclei. An example of such a growth of a solid particle having $r = 10^{-6}$ cm and the surface properties of glass, is given by curve 3 in Fig. 3. Comparison with the solution droplets of curve 2 shows that growth by adsorption becomes decisive for sizes below that value. Evidently at these dimensions the borderline between solution and adsorption becomes confused [39]. This is in accord with the observation [19] that insoluble but flaky nuclei require a higher degree of humidity for their first condensation than for their second; thus they seem to disintegrate at the first condensation, and the finely divided substance goes, so to speak, "into solution."

As yet, no attention has been paid to the influence of

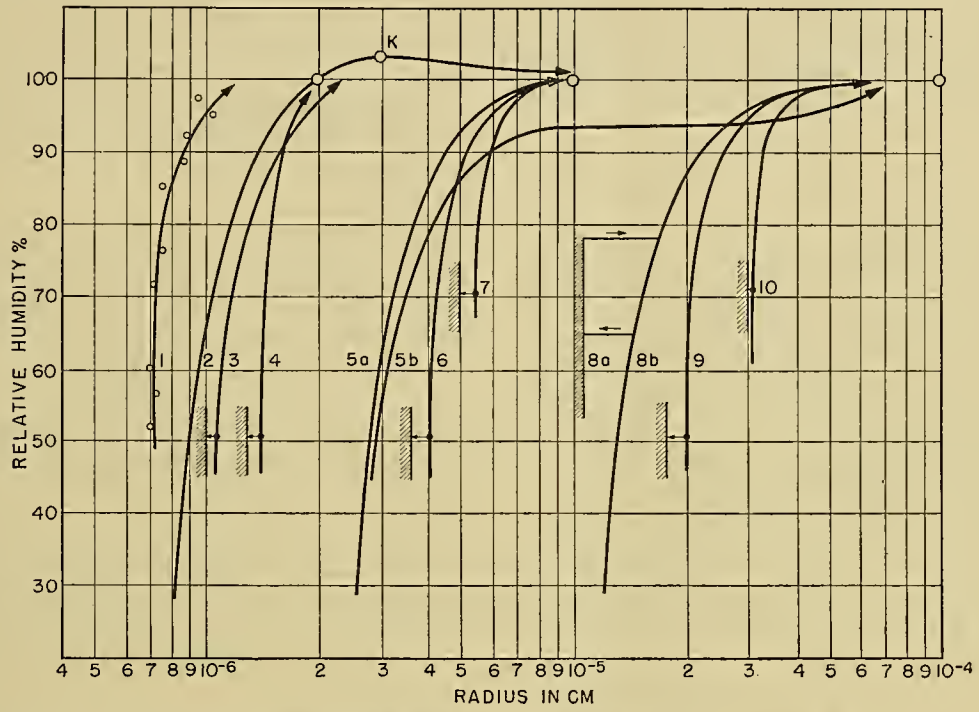


FIG. 3.—Curves relating the growth of various nuclei to humidity.

1. Measured values of the growth of newly formed gas-flame ions [19].
2, 5*a*, 8*b*. Calculated curves of the growth of pure droplets of solutions whose radii at 100 per cent relative humidity have values of 2×10^{-6}, 10^{-5}, and 10^{-4} cm, respectively.
3. Growth by adsorption of a solid nucleus with the surface properties of glass and $r = 10^{-6}$ cm [40].
4, 6, 7, 9 and 10. Curves for the growth of mixed nuclei whose radii at 100 per cent relative humidity are successively 2×10^{-6}, 10^{-5}, and 10^{-4} cm and whose solid portions have the radii indicated by the hatched lines.
5*b*. Modification of the growth curve 5*a* in the presence of 250×10^{-6} g m^{-3} of *HCl* and a nuclear count of 10^4 cm^{-3}. *HCl* was chosen, since we have precise physical-chemical data for it.
8*a*. Crystallization (with decreasing humidity) and solution (with increasing humidity) of a *NaCl* nucleus (schematic).
K. Point of condensation, boundary between the growth of the nucleus, and droplet condensation.

stable compounds, such as HCl, HNO_3, and H_2SO_4, on the growth of nuclei. In the presence of traces of such compounds there is equilibrium between the phase dissolved in the nuclei and the gaseous phase. If gases, such as SO_2 and NH_3, that form less stable solutions are involved, the dissolved part is minute (because of the smallness of the nuclei) so that practically everything goes into the gaseous phase. With increasing humidity a point will be reached (depending on the quantity and type of the substance) at which there is a sudden growth of nuclei. This growth will continue until the vapor pressure of the gas traces decreases perceptibly (Fig. 3, curve 5*b*). This process is likely to be of importance in the formation of dry fogs in industrial areas.

Formation of Nuclei in Smoke and During Gas Reactions

The formation of nuclei in smoke is engendered by the growth of the smallest molecular particles by sublimation, and by condensation of matter having a very high boiling point. Because of the high original nuclei concentrations, these particles continue to grow by coagulation, causing the formation of mixed nuclei. Among the important examples of this type of nuclei are soot particles and the tarry and hygroscopic particles present in most combustion effluents.

Gas reactions are likewise significant as a source of nuclei. Often they may become effective only at a great distance from the point where the gas traces originate, after mixing with other gas traces or through photochemical processes. According to Rothmund [32], nuclei are always produced if the products of such reactions are water-soluble and if the humidity of the air is sufficiently high. During the most varied reactions, he noted the formation of haze droplets of $r \approx 5 \times 10^{-5}$ cm. Details of the formation mechanism of these nuclei, particularly with reference to the significance of humidity, as well as the establishment of an equilibrium between the gaseous participants in the reaction and the reaction product (nuclei number, size), are not yet known.

The concept of such equilibriums between the gaseous phase and the "nuclei phase" permits these two conclusions:

1. Many traces of matter will be found in the gaseous phase and also as condensation nuclei. This is corroborated by the fact that in many instances the amount of matter appears to be too high to be explained by plausible numbers and sizes of nuclei. For example, Cauer [5] was unable to establish a relationship between the concentration of Aitken-nuclei and traces of matter (NH_4^+, NO_2^-, SO_4^{--}, and Cl^--contents). It must be pointed out, however, that *quantitatively* the *large* nuclei, that is, the haze droplets and dust particles, may (in spite of their small number) prevail over the Aitken-nuclei, so that a more exact statement will have to wait until the entire nuclei spectrum is known.

2. The change in external conditions (*e.g.*, in the humidity or the concentration of the gaseous phase) may cause the number of nuclei to change; the nuclei concentration in a given mass of air would thus not represent a constant quantity. It is, for example, not unlikely that nuclei whose substances are only faintly volatile (such as oils and tars as well as the acids HNO_3, H_2SO_4, and HCl) slowly evaporate if mixed with pure air. In areas which are far removed from human settlements and industry (*e.g.*, mountains, oceans), one may not expect to find aerosols other than those consisting of solutions of nonvolatile salts or solid matter. In this connection mention should be made of an observation by Schlarb [34], according to which pressure changes in a climatic chamber between ½ and 1½ atmospheres were found to result in a fluctuation of the number of nuclei in a proportion of about 1:50!

In the following paragraphs a few meteorologically important gas reactions will be mentioned briefly [4, 23, 29, 32].

Ozone, O_3, while not in itself nucleogenic, enters into reaction with many other traces of matter, such as nitric oxides or SO_2, thereby producing nuclei [31]; H_2O_2, which, however, according to Cauer [4], is found only close to flames, shows a similar behavior.

Oxides of nitrogen, NO and NO_2, originate simultaneously with O_3 during electric discharges (thunderstorms) and as a result of the ultraviolet solar radiation [29], as well as in all incandescent processes and in flames, according to Coste and Wright [7]. These oxides are very effective in creating nuclei.

Sulfur dioxide, SO_2, which develops in all combustion processes, forms nuclei in combination with O_3 or, according to Aitken [2], through photochemical processes in combination with gas traces (formation of dry fogs *after* sunrise). These nuclei are probably composed of H_2SO_3 and H_2SO_4.

Chlorine ions, Cl^-, according to Cauer [6], escape from the surface of the sea as well as from droplets of ocean spray, and form HCl-nuclei photochemically. This explains why a separate determination of the magnesium and chlorine contents of the air gives mass proportions which are entirely different from those found in sea water. Part of the chlorine content exists in the form of salts.

Ammonium ions, NH_4^+, result from all processes of combustion and from the decomposition of organic substances and, when reacting with the acids mentioned above, lead to the formation of nuclei.

Iodine, I_2, is dispersed into the air along the Atlantic Coast of Europe mainly by industrial charring of seaweed for the extraction of iodine [4].

Some typical values, indicated by Cauer [4], will stress the significance of these traces (values are in 10^{-6} g m^{-3}).

Ozone:	Tatra 30, Jungfraujoch 10;
Nitrite:	Tatra and Silesia 0.2;
Sulfite:	Berlin 200, Silesia occasionally 3, Tatra, faint traces, but very rarely;
Chloride:	Brittany 7, on the sea near Kiel 149, Silesia 32, Tatra 70;
Ammonia:	Silesia 17, Brittany 21;
Iodine:	Average for Central Europe before November 1, 1933, 0.6; thereafter 0.05 (be-

cause of the crisis in the iodine industries).

Mechanical Sources of Nuclei

In contrast to the particles treated in the foregoing section, which grow from molecular dimensions, the dust particles and the droplets of sea spray are produced by mechanical separation. Simpson [36] points out that, according to Gibbs, the limit of these separations is of the order of $r = 5 \times 10^{-6}$ cm as a result of the action of cohesion- and surface-forces; usually, however, the particles are found to be substantially greater. Thus Köhler [21] found in his spray experiments with sea water that the greatest number of droplets had dimensions ranging between $r = 1.5$ and 3.5×10^{-5} cm, that is, the same size as the haze droplets obtained by Dessens.

As a result of the thorough investigations by Woodcock and Gifford [43] it appears certain that over the ocean the large nuclei largely consist of sea water sprayed by storms and surf and are transported by austausch to great heights in maritime air masses. The chloride content of hoarfrost and precipitation [47], the observed cubic crystals sublimated from haze droplets over the continent [1, 8], and the observation by Cauer [4, 6] of considerable traces of *Mg* and *Cl* ions during influx of maritime air masses into the Tatra mountains lead to the conclusion that, far into the continents, the haze droplets among the large nuclei represent a more or less significant portion of maritime aerosol. Although it is certain that these nuclei are often numerically insignificant as compared to the nuclei concentrations over densely populated continents, they nevertheless play a role in the analysis of traces of substances because of their size.

It is still uncertain to what extent the Aitken-nuclei over the oceans consist also of seawater spray [11, 37, 46, 47, 48]; it is certain only that the oceans are not very productive sources of nuclei in point of numbers [3, 23] (Table I), so that the numerical proportion of *NaCl*-Aitken-nuclei is probably small over land.

One important source of nuclei is the particles that, together with dust, are raised by storms over land. Substantial amounts of matter can thus be carried considerable distances, as is indicated by the Sahara-dust falls occasionally observed in Europe [12].

In conclusion of the discussion on the sources of atmospheric nuclei, the estimates of the relative contributions by various sources to the total nuclei content of the air, first given by Lettau [24], may be reproduced. If the total production of the nuclei is set to equal unity, the proportions indicated below are obtained, attributable as follows:

Steppe and forest fires	0.4
Detachment from the soil surface	0.3
Combustion products from man-made fires	0.2
Volcanic activity	0.1
Detachment from the sea surface and from extra-terrestrial sources	negligible amounts

The indicated values can be taken only as very rough approximations.

Electrostatic Charges of Nuclei

The electrical state of an aerosol is conveniently characterized by the ratio

$$P = \frac{\text{total number of nuclei}}{\text{sum of positive and negative ions}}.$$

The combination of small ions and *uncharged* nuclei, which is, among other things, considered as proportional to the surface area of the nuclei, is increased in the case of *charged* nuclei as a result of the attraction of the electric charges which is independent of the size of the nuclei. It follows that a greater percentage of the larger nuclei must be charged. Wright [45] derived the values

Table III. Relationship between P and the Size of Nuclei (*after Wright* [45])

Radius of nucleus (cm $\times 10^{-6}$)	0.0	1.0	2.0	4.0	6.0	∞
P	∞	10.0	3.3	2.0	1.7	1.5

given in Table III. The quantity P can be measured by the following two methods:

1. Through the determination of the total nuclei count and the nuclei count after the elimination of the large ions (both found with a nuclei counter).

2. Through the determination of the total nuclei count and of the number of elementary charges (with an ion counter).

From the difference of the two measurements [14, 15], we obtain the number of multiple charges on the nuclei

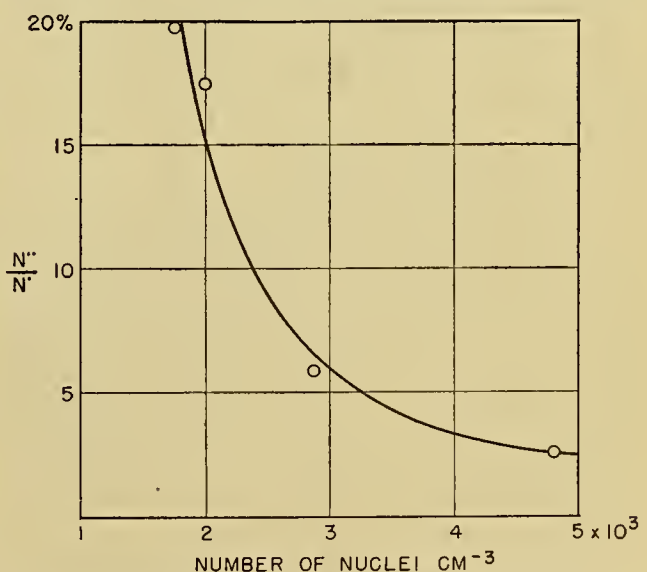


Fig. 4.—Percentage of doubly charged ions as a function of the number of nuclei [14, 15].

(Fig. 4). With nuclei counts below 5000 cm^{-3} a considerable proportion of nuclei are doubly charged.

From extensive observations by Israël [16] and others the following relationships involving P are obtained:

1. There is an increase of P with increasing nuclei count, varying locally; this increase becomes more pro-

nounced when the second method is used because of the multiple charges on the nuclei. Thus in a larger nuclei count, a smaller fraction of the nuclei are charged, because only a limited number of small ions are available. As yet, there has been no quantitative explanation of this.

2. There is a decrease of P with increasing humidity (Fig. 2). The nuclei, whose growth with humidity has a marked influence on P, have, according to Table III, radii of less than about 6×10^{-6} cm. Thus the behavior of P reflects the growth of particles of this size.

The quantity P gives us a good example of how careful we must be when we want to interpret statistically determined relationships between the properties of natural aerosols and meteorological quantities. If, for example, we arrange measurements of the nuclei concentrations and P values according to visibility, as Israël did in Frankfort, we obtain an increase in P values and a decrease of nuclei count with an increase of visibility. According to the relationship (1) described above we would have expected a decrease in P. However, it is probable that, on the average, with increasing visibility, both the number and size of the nuclei decrease, and the size appears, according to Table III, to have a dominant influence on P.

Israël further points out that individual measurements of P show considerable fluctuations which are especially large in the vicinity of cities, that is, of the sources of nuclei [16]. He attributes this phenomenon to the slow establishment of the ionization equilibrium.

From the preceding survey it is apparent that the study of the ionization equilibrium permits us to reach important conclusions concerning the aerosol. However, this possibility will be most fully realized only when we can take into account in the investigations the influence on P of the size distribution in the nuclei spectrum.

The Optical Effects of Nuclei

Primarily effective for the scattering of light, and consequently the turbidity of the atmosphere, are the particles of the order of magnitude of the wave lengths of light or larger, that is, those with radii of approximately $r \geqslant 4 \times 10^{-5}$ cm. If the radius of the particle decreases, the optical effects diminish very rapidly. Thus, according to our method of notation, only the large nuclei, and not the Aitken-nuclei, are significant for the turbidity, unless the Aitken-nuclei are particularly numerous, or the atmosphere is extremely clear.

Wright [46] used these facts in order to reach conclusions concerning the composition of the aerosol, through extensive investigation of the visibility conditions at some English stations. He found that there were approximately 1000 table-salt nuclei with a radius of 2×10^{-5} cm in a cubic centimeter. This nuclei concentration varies relatively little with time and location and constitutes the maritime portion of the aerosol. Over land we have additional solid particles (such as soot) which can be detected with a dust counter. These can become so numerous, especially during wintertime and in densely populated areas, that their optical effects are dominant. The many small nuclei of combustion (Aitken–nuclei) have hardly any effect on visibility.

Dessens [8] found good proof for these conclusions. He measured the attenuation coefficients for various wave lengths and also calculated them from the simultaneous determinations of the size distribution of the haze droplets, according to the formulas of Stratton and Houghton. The agreement between observed and computed values is good and at the same time shows that his method of measurement yields a quantitatively correct sample of the haze droplets.

Recently, and independently, Siedentopf [35] and others have deduced the optical effects of the aerosol from Mie's theory on a much more inclusive and exact basis. Siedentopf bases his calculations on an atmosphere in which the visibility is 20 km, and attempts to find the composition of the aerosol which would correspond most closely to the dependence of the absorption on the wave length, as well as to the scattering function (dependence of the scattered light on the angle of the incident ray). With pure dust aerosols (completely opaque, index of refraction $n = \infty$) or with pure droplet aerosols ($n = 4/3$), no agreement could be reached. This was possible only with "mixed aerosol," composed of particles of both types in the proportion of 20 dust particles per cubic centimeter to 500 haze droplets per cubic centimeter, with an average radius from 2 to 3×10^{-5} cm. The addition of several thousand Aitken–nuclei with $r = 5 \times 10^{-6}$ cm has no effect on the results. These findings are in good agreement with the results of both Dessens and Wright, a fact which seems to indicate that these haze droplets, found far inland, are essentially of maritime origin.

It is therefore clear why previous investigations yielded no definite relationship between the Aitken–nuclei count and visibility (according to [23]). The situation is different, however, for dust particles, as was shown by Rötschke [33].

The effect on visibility of the growth of haze droplets with humidity has already been discussed. Wright [46, 47] has established a similar relationship between humidity and visibility when the optical effect of the dust particles predominated. This can be explained by postulating either a considerable accumulation of such particles during very humid weather conditions or a coagulative effect of humidity.

Processes of Nuclei Destruction

Aside from condensation itself, the following processes essentially diminish the number of nuclei:

1. Sedimentation.
2. Diffusion (adhesion to surfaces of all kinds).
3. Coagulation.

In all these processes, the mobility of the particles plays a role. The mobility of particles that are smaller than, or comparable to, the length of the mean free path of the air molecules (approximately 10^{-5} cm), is

$$B = \frac{1 + Al/r}{6\pi\eta r},$$

where B = mobility (sec g^{-1}),
l = free path length (cm),
r = radius of particle (cm),
A = numerical factor of 0.9,
η = viscosity of air (g cm^{-1} sec^{-1}).
The viscosity η is practically independent of the atmospheric pressure, so that with decreasing pressure, B increases with l.

1. The fall velocity of the particles is

$$V = BMg,$$

where M = mass of the particles (g),
g = gravitational acceleration (cm sec^{-2}).
Despite the small fall velocity, the sedimentation process that takes place continuously and everywhere in the immediate vicinity of the earth's surface is probably important for the nuclei balance [24]. The same holds true for the charged particles settling in the electric field of the earth.

2. The diffusion coefficient that is valid for the adhesion of particles to surfaces of all kinds is

$$D = \frac{RT}{N} \cdot B,$$

where R = gas constant (erg deg^{-1} mol^{-1}),
T = absolute temperature (deg),
N = Loschmidt number.
The diffusion coefficient D greatly increases with decreasing radius of the particle and has been used for the determination of the average particle size [28]. In contrast to sedimentation, adhesion affects especially the small particles. In closed spaces or vessels, the influence o diffusion increases with increasing ratio of surface to volume [30].

Diffusion is effectual in forests and grassland with their very large surface areas (filter effect) and in the clouds, in which the decreased nuclei concentration [23] is probably caused not so much by consumption during condensation as by adhesion to the droplets. For this reason, analyses of traces of substances in hoarfrost and other condensation products in combination with measurements of droplet number and size do not permit any conclusions regarding the mass of the individual nuclei.

3. As regards coagulation, reference is made to the exemplary investigations of Whytlaw-Gray and Patterson [42]. The decrease in the number of particles n per unit time and volume in homogeneous aerosols is

$$\frac{dn}{dt} = -\frac{8\pi RT}{N} Brn^2 = -\frac{4RT}{3\eta N}\left(1 + A\frac{l}{r}\right) n^2,$$

that is, the small particles coagulate more rapidly. The numerical results given in Table IV clearly show that coagulation is much more effective for the originally very small combustion nuclei that are highly concentrated in city air, etc., than for the coarser and less concentrated dust and ocean spray aerosols. Mixed nuclei are, therefore, to be expected especially from industrial and similar nuclei sources. The laws of coagulation, experimentally confirmed by Whytlaw-Gray, become more complicated in natural aerosols because of the broad distribution of sizes and occasional electrostatic charges. For example, if two particle sizes are present, the greater coagulation rate of the smaller nuclei is further increased by adhesion to the larger nuclei. Therefore, aged aerosols contain only a small percentage of small nuclei.

Table IV. The Decrease in the Number of Particles per Hour in Homogeneous Aerosols

Nuclei number (cm^{-3})	Radius (cm)		
	10^{-6}	10^{-5}	10^{-4}
10^3	6.86×10^0	1.66×10^{-1}	1.13×10^{-1}
10^4	6.86×10^2	1.66×10^1	1.13×10^1
10^5	6.86×10^4	1.66×10^3	1.13×10^3

Since the large particles grow more rapidly with increased humidity than do the small ones, a slight increase in coagulation rate with humidity is to be expected. This was confirmed by Landsberg [23] who found a decrease in the number of nuclei in rooms when the humidity was increased. Effenberger [10] also found by simultaneous nuclei and dust counts in the open air that, with increasing humidity, the nuclei concentration decreases while the dust content increases. However, in this case it must be noted that there is a possible selective effect of the synoptic situation. Moreover, the sedimentation rate in the dust counters is greater with higher humidity because of the growth of the dust particles.

Concluding Remarks

Because of lack of space it has been impossible to include a compilation of the results of the many works dealing with nuclei counts. However, such a compilation is not really necessary, since the reader can refer to exhaustive monographs [3, 23], and no new information has been obtained on the subject. In the various sections we have touched upon those relationships, determined by such investigations, of the nuclei count, etc., to other meteorological elements that were of interest from the point of view of the physics of nuclei. In many cases these relationships have only a limited value in explaining physical processes, since the correlations generally overlap. For this reason, it has been suggested by various investigators that stress should be put on additional laboratory research work. A number of open questions touched upon in the sections above could thereby be clarified.

In conclusion some of the areas of study that seem especially important to the author are summarized:

1. The growth of nuclei, especially at humidities of less than 70 per cent, and the associated problem concerning the physical structure of nuclei (mixed nuclei, supersaturation of solution, and crystallization).

2. The chemical composition of the aerosol. Parallel measurements of the traces of substances and of the number of nuclei, dust particles, and haze droplets

should be made. The proportion of traces of gaseous materials should be determined.

3. The performance of haze-droplet counts on a larger basis to determine the contribution of maritime aerosol (sea-salt nuclei).

4. The behavior of substances of low volatility, and the influence of equilibrium and humidity in gas reactions on the concentration and size of nuclei.

5. Effect of humidity on coagulation.

However, the success of pertinent research projects is contingent upon the development of a new method for the measurement of the entire spectrum of nuclei. At the present time this measurement is possible only approximately through simultaneous counts of the nuclei, ions, dust particles, and haze droplets. We should, however, investigate exactly how the individual ranges of measurement overlap. Here, the electron microscope probably offers new possibilities.

REFERENCES

1. Advisory Committee on Atmospheric Pollution. London, 1922 (see pp. 34–42); 1923 (see pp. 33–46); 1924 (see pp. 35–44).
2. Aitken, J., "The Sun as a Fog Producer." *Proc. roy. Soc. Edinb.*, 32: 183–215 (1912).
3. Burckhardt, H., und Flohn, H., *Die atmosphärischen Kondensations-Kerne in ihrer physikalischen, meteorologischen und bioklimatischen Bedeutung.* Berlin, J. Springer, 1939.
4. Cauer, H., "Chemie der Atmosphäre" in *FIAT Rev. of German Sci., 1939–1946, Meteorology and Physics of the Atmosphere*, R. Mügge, senior ed. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948. (See pp. 277–291)
5. —— und Cauer, G., "Studium über den Chemismus der Nebelkerne in Oberschreiberhau." *Der Balneologe*, 8: 345–353 (1941).
6. —— "Das Magnesiumchlorid der Nebelkerne." *Der Balneologe*, 9: 301–309 (1942).
7. Coste, J. H., and Wright, H. L., "The Nature of the Nucleus in Hygroscopic Droplets." *Phil. Mag.*, 20: 209–234 (1935).
8. Dessens, H., "Les noyaux de condensation de l'atmosphère." *La Météor.*, pp. 321–327 (1947).
9. —— "Un compteur de particules atmosphériques." *La Météor.*, pp. 328–340 (1947).
10. Effenberger, E. F., "Kern- und Staubuntersuchungen am Collmberg." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 12: 305–359 (1940).
11. Findeisen, W., "Die Kondensations-Kerne." *Beitr. Phys. frei. Atmos.*, 25: 220–232 (1939).
12. Glawion, H., "Staub und Staubfälle in Arosa." *Beitr. Phys. frei. Atmos.*, 25: 1–43 (1939).
13. Heverly, J. R., "Supercooling and Crystallization." *Trans. Amer. geophys. Un.*, 30: 205–210 (1949).
14. Hogg, A. R., "Atmospheric Electric Observations." *Beitr. Geophys.*, 41: 1–31 (1934).
15. —— "Some Observations of the Average Life of Small Ions and Atmospheric Ionization Equilibria." *Beitr. Geophys.*, 41: 32–55 (1934).
16. Israël, H., "Ionen und Kerne; eine kritische Studie." *Beitr. Geophys.*, 57: 261–282 (1941).
17. —— und Krestan, M., "Zur Methodik der luftelektrischen Messungen: Die Zählung der Kondensationskerne." *Beitr. Geophys.*, 58: 73–94 (1942).
18. Israël, H., und Schulz, L., "Über die Grössenverteilung der atmosphärischen Ionen." *Meteor. Z.*, 49: 226–233 (1932).
19. Junge, C., "Übersättigungsmessungen an atmosphärischen Kondensationskernen." *Beitr. Geophys.*, 46: 108–129 (1935).
20. —— "Zur Frage der Kernwirksamkeit des Staubes." *Meteor. Z.*, 53: 186–188 (1936).
21. Köhler, H., "An Experimental Investigation or Sea-Water Nuclei." *Nova Acta Soc. Sci. upsal.*, Ser. 14, Vol. 12, No. 6 (1941).
22. Lafargue, C., "Sur la congélation des gouttelettes d'eau vers −41°C." *C. R. Acad. Sci., Paris*, 230: 2022–2024 (1950).
23. Landsberg, H., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, (Supp.) 3: 155–252 (1938).
24. Lettau, H., "Versuch einer Bilanz im Kondensationskern-Haushalt der Troposphäre im Durchschnitt für die ganze Erdoberfläche." *Ann. Hydrogr., Berl.*, 67: 551–559 (1939).
25. Linke, F., "Das atmosphärische Aerosol," *Handbuch der Geophysik*, Bd. 8, Kap. 2, SS. 14–27. Berlin-Zehlendorf, Gebr. Borntraeger, 1942.
26. —— "Kondensationskerne im Elektronenmikroskop sichtbar gemacht." *Naturwissenschaften*, 31: 230–231 (1943).
27. Möller, F., "Thermodynamik und Physik der Wolken" in *FIAT Rev. of German Sci., 1939–1946, Meteorology and Physics of the Atmosphere*, R. Mügge, senior ed., pp. 83–114. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948.
28. Nolan, J. J., and Guerrini, V. H., "The Determination of the Mass and Size of the Atmospheric Condensation Nuclei." *Trans. Faraday Soc.*, 32: 1175–1181 (1936).
29. Pringal, E., "Über den wesentlichen Einfluss von Spuren nitroser Gase auf die Kondensation von Wasserdampf." *Ann. Physik*, (4) 26: 727–750 (1908).
30. Rauscher, H., "Zählungen von Kondensationskernen in geschlossenen Gefässen." *Anz. Akad. Wiss. Wien.*, 70: 198–200 (1933).
31. Regener, E., "Das atmosphärische Ozon" in *FIAT Rev. of German Sci., 1939–1946, Geophysics*, II, J. Bartels, senior ed. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948. (See pp. 297–307)
32. Rothmund, V., "Über das Auftreten von Nebeln bei chemischen Reaktionen." *Mh. Chem.*, 39: 571–601 (1918).
33. Rötschke, M., "Untersuchungen über die Meteorologie der Staubatmosphäre." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 11: 1–78 (1938).
34. Schlarb, G., "Untersuchungen über Kondensations-Kerne und Leichtionen in künstlich klimatisierten Räumen." *Biokl. Beibl.*, 7: 86–105 (1940).
35. Siedentopf, H., "Zur Optik des atmosphärischen Dunstes." *Z. Meteor.*, 1: 417–422 (1947).
36. Simpson, G. C., "On the Formation of Cloud and Rain." *Quart. J. R. meteor. Soc.*, 67: 99–133 (1941).
37. —— "Sea-Salt and Condensation Nuclei." *Quart. J. R. meteor. Soc.*, 67: 163–169 (1941).
38. Wall, E., "Einfaches Schema der atmosphärischen Eiskeimbildung." *Meteor. Z.*, 59: 177–183 (1942).
39. —— "Die Eiskeimbildung in Lösungskernen." *Meteor. Z.*, 60: 94–104 (1943).
40. —— "Zur Physik der Wasserdampfkondensation an Kernen." *Z. angew. Meteor.*, 59: 106–125 (1942).

41. Weickmann, H., "Die Eisphase in der Atmosphäre." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 6, 54 SS. (1949).
42. Whytlaw-Gray, R., and Patterson, H. S., *Smoke: A Study of Aerial Disperse Systems*. London, E. Arnold & Co., 1932. (See pp. 57–72)
43. Woodcock, A. H., and Gifford, M. M., "Sampling Atmospheric Sea-Salt Nuclei over the Ocean." *J. mar. Res.*, 8: 177–197 (1949).
44. Wright, H. L., "Observations of Smoke Particles and Condensation Nuclei at Kew Observatory." *Geophys. Mem.*, Vol. 6, No. 57 (1932).
45. —— "The Association Between Relative Humidity and the Ratio of the Number of Large Ions to the Total Number of Nuclei." *Terr. Magn. atmos. Elect.*, 39: 277–280 (1934).
46. —— "Atmospheric Opacity: A Study of Visibility Observations in the British Isles." *Quart. J. R. meteor. Soc.*, 65: 411–442 (1939).
47. —— "Sea-Salt Nuclei." *Quart. J. R. meteor. Soc.*, 66: 3–11 (1940); "The Origin of Sea-Salt Nuclei." *Ibid.*, 11–12 (1940).
48. —— "Atmospheric Opacity at Valentia." *Quart. J. R. meteor. Soc.*, 66: 66–67 (1940).

THE PHYSICS OF ICE CLOUDS AND MIXED CLOUDS

By F. H. LUDLAM

Imperial College of Science and Technology and Meteorological Office, London

THE APPEARANCE OF ICE CLOUDS

It has long been recognised that clouds of ice particles have a characteristically fibrous appearance very different from that of droplet clouds, which look "solid" and have more sharply defined edges, even when dissipating. Small, isolated trails of snow or soft hail, known as *Fallstreifen* (fall-streaks) or *virga*, which fall from medium-level clouds, also possess this fibrous structure, and when they become separated from their parent clouds and are seen in a favourable light exactly resemble forms of true cirrus. A fibrous structure on a coarser scale can also be seen in rain beneath shower clouds.

It is therefore natural to interpret the fibrous appearance as the result of the high falling-speed of unusually large particles falling in trails from small condensation-regions, and to consider typical isolated cirrus clouds as a form of precipitation. Wegener [21] regarded cirrus uncinus as *Fallstreifen*, and attributed their development to the absence of efficient ice nuclei, those available acting only at a high supersaturation so that the ice particles then formed grow quickly and acquire a large falling-speed. There have been other attempts to explain the form of these clouds, but careful observation supports the *Fallstreifen* theory, and in recent years investigations concerning the properties of ice-particle nuclei have completely explained why ice clouds so often contain much larger particles than droplet clouds, and have led to a better understanding of the processes attending the formation and decay of both pure *ice-clouds* and *mixed clouds* (containing both ice and liquid water).

THE BEHAVIOUR OF ICE NUCLEI

For some time it has been thought that the nuclei effective in producing ice crystals are probably solid, water-insoluble particles, distinct from the hygroscopic nuclei which readily allow the formation of droplets. Wegener [21] and Findeisen [5] suggested that quartz dusts, carried up from the ground, might be the atmospheric ice nuclei. The prevalence of clouds of liquid water at temperatures down to about −10C was ascribed to the inefficiency of the ice nuclei or to their absence. The nuclei were called *sublimation nuclei*, and it was generally believed that the ice crystals were formed by the direct transition of water vapour into the solid state.

This conception has been altered by recent researches. The first important contributions to the subject were by Krastanow [10, 11], who outlined the theory of condensation according to the principles of statistical physics, and showed that the direct deposition of vapour into ice upon such solid nuclei as occur in the atmosphere is to be expected very rarely and only at temperatures approaching −70C.[1] He suggested that ice particles are usually produced secondarily by the thermodynamically easier path through the freezing of droplets containing these nuclei, and claimed that such infected droplets exist because condensation can also proceed upon solid, insoluble (but "wettable") particles as well as upon hygroscopic nuclei. Accordingly, it appeared that crystals usually arise as a result of the freezing of supercooled drops (and perhaps also by the freezing of the microscopic film of adsorbed water which grows on solid "wettable" particles at vapour pressures approaching the saturated vapour pressure over liquid water, followed by direct condensation from the vapour). For any such particle there are successively lower ranges of temperature in which it may act as a nucleus for a supercooled drop, for a supercooled drop which freezes, and for an ice crystal formed by direct condensation. Krastanow was concerned primarily with estimating these ranges, but his work contained the important implication that atmospheric crystal nuclei would become active only at approximate saturation over liquid water—formerly it was believed that they would allow direct condensation at small ice-supersaturations. At the temperatures of the cirrus levels, however, water-saturation corresponds to relative humidities of 150 per cent or more over ice.

Some support for this view was provided by the expansion-chamber experiments of Regener [16] in which droplet fogs were formed at temperatures down to at least −50C, crystals also appearing when high expansion ratios were used. Weickmann [22], observing the condensation products upon nuclei supported on a metal face chilled to about −40C, gave a complete confirmation, finding that no crystals could be produced at low ice-supersaturations, that a few formed under prolonged high ice-supersaturations, and that relatively abundant formation occurred in the neighbourhood of water-saturation. He also showed that particles of the common minerals (including quartz) are not especially active as ice nuclei, but concluded that atmospheric ice nuclei are small insoluble solids about 1 μ or less in size, which, from their mode of action, he calls *freezing nuclei*.

It is interesting that by careful search particles have now been found which appear to act as true sublimation nuclei (allowing the formation of ice crystals direct from the vapour at temperatures below −10C and at

1. A more detailed account of these references and some subsequent ones has already been given by the present writer [12].

small ice-supersaturations). Fournier d'Albe [9] produced these particles by evaporating crystals which had been grown at −41C in the presence of solution droplets of cadmium iodide. It seems likely that the particles remaining consisted of cadmium iodide bearing microfilms of water which retained an icelike structure such as cannot be imposed upon films adsorbed from the vapour, even by cooling to temperatures in the neighbourhood of −70C. This suggests that the only sublimation nuclei are in fact particles of ice, and raises the possibility that after the evaporation of ice clouds at low temperatures some freezing nuclei may be left in a condition which enables them to act as sublimation nuclei in a future cloud formation. However, nuclei having the efficiency of those made by Fournier d'Albe cannot be expected to occur naturally in the atmosphere, where even the poor ice nuclei available are greatly outnumbered by nuclei capable of aiding the formation of liquid droplets. This scarcity of ice nuclei, which may be even more marked in the high troposphere than near the ground, is an important factor which in association with the speed of atmospheric condensation processes results in the frequent appearance of droplets during cloud formation at temperatures down to −50C or even lower. If the humidity is increased very slowly some ice particles will form some time before water-saturation is reached; their subsequent growth may then remove enough vapour to prevent a further increase in humidity, or the process responsible for the rise of humidity may cease. In this way very thin pure ice clouds may arise without any droplet formation. Their occurrence must be very infrequent in view of the great scarcity of ice nuclei which act before water-saturation is almost reached, and the slowness and duration of the condensation mechanism which would be necessary, but the structureless veils of cirro-nebula and the "diamond-dust" ice-mists probably are such clouds.

Evidently, in experiments designed to discover the critical temperatures at which droplet formation is entirely replaced by crystal formation in the presence of foreign nuclei, particular attention must be paid to the speed of the condensation process. It is clear that the violent expansions commonly used in cloud chambers may produce predominantly droplet clouds at temperatures down to −40C and even lower, and it is also clear why different workers, using various rates of expansion, have found different values for the critical temperatures.

The expansion-chamber experiments which most faithfully reproduced natural conditions are those of Findeisen and Schulz [8], who used an unusually large chamber (volume 2 m^3) to increase the sensitivity to the rare but still meteorologically important ice nuclei, and who controlled the rate of expansion to values corresponding to vigorous atmospheric convection processes. Their apparatus was not portable, and so only the surface air (at Prague) was examined for ice nuclei. The number of such nuclei in this polluted air may have been much larger than is usual in the air of the upper troposphere, but even so in the chamber-clouds the droplets still outnumbered the crystals at temperatures around −40C. Down to this temperature the clouds formed in the chamber by uninterrupted expansions were initially pure water-clouds or mixed clouds, but some pure ice-clouds were produced with great difficulty by ceasing the expansion just before droplet-formation was expected. The numbers of nuclei active at various temperatures composed an ice-nucleus "spectrum," which changed somewhat from day to day and with the rate of expansion, but which possessed the general form shown in Fig. 1. With another apparatus, in which the speed of the condensation process was much less, Schulz [17] has recorded results which do not differ materially from those given in this figure. As the temperature falls below about −32C there is a sudden great increase in the number of active nuclei, which Findeisen and Schulz consider due to the introduction of a second class of nuclei. These appear to be the nuclei responsible for the first appearance of crystals (in droplet fogs) in the smaller chamber used by Fournier d'Albe [9], and which are now referred to as the *32-nuclei*. Findeisen and Schulz apparently did not continue their experiments to temperatures below −40C, and so did not find a second critical temperature of about −41C below which Fournier d'Albe observed relatively dense ice-fogs in which the presence of any droplets could not certainly be established. Fournier d'Albe, however, was able to confirm that these ice fogs formed only when the air was cooled near to, if not below, the dew point (at saturation with respect to liquid water). The reason for this critical temperature of −41C and the part it plays in more moderate expansions remain unknown. More detailed examination may show this critical temperature to be no more sharply defined than the threshold temperature at which the 32-nuclei become active.

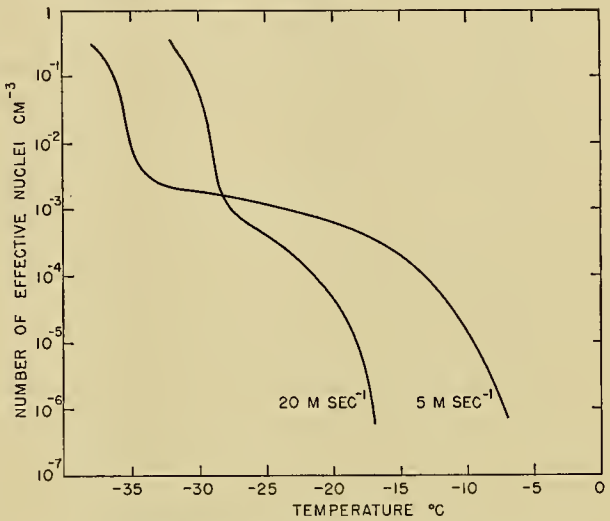


FIG. 1.—Average ice-nucleus "spectrum" for samples of surface air at Prague, showing the numbers of nuclei active at various temperatures during the expansion of the air at rates corresponding to vertical speeds in the atmosphere of 5 m sec^{-1} and 20 m sec^{-1}. (*After Findeisen and Schulz.*)

Very little is known about the number of ice nuclei occurring in the upper troposphere. Palmer [14] has

used an expansion chamber in an aircraft on four occasions and found that the 32-nuclei were absent (or very scarce) above the lowest inversion or haze-top. This suggests that these nuclei at least are derived from the earth's surface, and that at the cirrus levels the ice-nuclei count is likely to be far smaller than that observed at the surface by Findeisen and Schulz [8]. It is therefore to be expected that, in the atmosphere, clouds formed by the more vigorous ascending motions will initially contain liquid water even at very low temperatures, and indeed this has been observed in several ways. For example, R.A.F. aircraft have reported serious icing at temperatures down to below −40C [13], Weickmann [24] has photographed droplets (with crystals) at −50C over the Alps in the lenticular Moazagotl cloud, and aircraft wing-icing has been observed at −58C in a condensation trail formed at the airscrew [1]. The limit of supercooling, at which crystals grow spontaneously in liquid water without the aid of foreign nuclei, is not known. It has been roughly estimated theoretically at about −70C or below, and Rau [15] appeared to have found it experimentally at −72C, but Cwilong [4] has since thrown doubt on this result. It would certainly seem that at all temperatures likely to occur in the troposphere ice crystals always require special nuclei for their formation, and that these nuclei are always actively available at low enough temperatures, since stable supercooled clouds have apparently not been observed at temperatures below about −35C [24] and are rarely observed at temperatures below about −15C.

Before leaving the discussion of nuclei it may be said that further information is needed on the behaviour of hygroscopic nuclei at low temperatures, both when they are pure and when they are contaminated by freezing nuclei. Wall [20] has considered the matter theoretically, and some experiments on the freezing of droplets of common-salt solution are described by Weickmann [24]. From these latter experiments it appears that salt crystals can act as ice nuclei at temperatures below about −30C, but in spite of the decreasing solubility of salt with falling temperature it is unlikely that solid salt crystals can occur at high humidities unless they arise in contaminated solution-nuclei. It also seems that freezing nuclei may have their efficiency considerably reduced in the presence of a concentrated salt solution, so that such nuclei might begin to act only when the solution becomes sufficiently diluted by condensation. Mason, at the Imperial College of Science and Technology (in work unpublished), has attributed the critical temperature of −41C mentioned above to the freezing of solution droplets in which the ions formed by the dissociation of the electrolyte act as nuclei.

Meteorologically, the most important features of the recent studies on ice nuclei are the recognition of the rarity of ice nuclei and the establishment of the fact that ordinarily they allow the formation of ice crystals only near water-saturation, and therefore in the presence of high ice-supersaturations. Present problems concerning ice nuclei may be divided into two classes.

1. Problems related to individual nuclei: What is the mechanism of nucleation? and, What are the properties which determine the efficiency of an ice nucleus? Such questions may be answered by examining the behaviour of small numbers of artificial nuclei, or perhaps even a single nucleus, under humidities increased at carefully controlled rates. Dew-plate techniques are likely to be more convenient than those using expansion chambers, in which the control of temperature and humidity is very difficult, but investigations must first be made to determine whether the presence of a supporting surface interferes with the behaviour of the nuclei. If it can be established that certain unwettable surfaces are sufficiently indifferent, this method could readily be used to examine the efficiency of particular nuclei when they are covered by films of adsorbed vapour, and when they are immersed in droplets of both pure water and salt solutions. It should be possible to determine whether solutions of salts or of gases (*e.g.*, NO, SO_2, and NH_3) are in any circumstances able to act as ice nuclei when not contaminated with insoluble solid particles.

2. Problems of a more meteorological nature: What are the effective ice nuclei in the atmosphere? and, What are their numbers at different levels in various air masses? To observe the numbers of ice nuclei occurring naturally a very sensitive counter is required, capable of indicating ice-nucleus concentrations as low as 1 m^{-3}. The high-pressure expansion apparatus described by Schulz [17] was designed (but never used) as a portable nucleus counter; in this apparatus the air was first compressed to a pressure of about twelve atmospheres so that, on controlled expansion to about one atmosphere, low temperatures could be reached without the necessity of a preliminary cooling of the expansion chamber, while the concentration of ice nuclei was not reduced below that occurring in the atmosphere. However, elaborate precautions are needed to prevent the introduction of artificial nuclei during the compression process, and it is to be hoped that some much simpler counting apparatus may be practicable.

THE DEVELOPMENT OF THE ICE PHASE IN NATURAL CLOUDS

Cumulus and Cumulonimbus. Cumulus clouds are composed of air which has risen from near the ground, and conditions within them are likely to resemble closely those produced in the expansion chamber of Findeisen and Schulz [8]. Accordingly, ice-crystal formation does not begin in these clouds until the temperature has fallen several degrees below 0C, depending to some extent on the rate of ascent of air into the cloud base (see Fig. 1). Since the crystals are in a considerable supersaturation they grow more rapidly than the droplets and after a period of perhaps a few minutes reach a size and falling-speed such that growth by coagulation with droplets in their path begins to exceed that due to the condensation of vapour. Rough calculations show that this stage is reached when the ice particles have reached diameters of about 100 μ and falling-speeds of about $\frac{1}{4}$ m sec^{-1} and when (relative to the surrounding air) they will have fallen rather less than 30 m. Beyond this stage growth by coagulation rapidly accelerates,

both the volume swept by a particle and its rate of fall quickly increasing, and the particle becomes a precipitation element. However, it is important to note that it takes an appreciable time for the growth to reach the critical stage: if the cloud begins to dissolve in the meantime the ice particles will not develop into precipitation elements, but will disappear with the droplets. Observation shows that the summits of cumulus clouds are constantly and rapidly replaced from below, fresh cloud masses rising only to diverge, sink, and evaporate, so that the life of each particle near the cloud tops is very brief. It can therefore readily be imagined that cumulus, rising some distance above the level at which the ice nuclei first act in significant numbers, have summits containing ice crystals which fail to reach the critical size which would enable them to sink into the bulk of the cloud and continue their growth, thus eventually becoming hail or rain. The magnitude of this effect is difficult to estimate, but it is likely to be most marked in clouds of low liquid-water content, in which the rate of growth of the ice particles is least. In general the liquid-water content of cumulus varies directly with the temperature of the cloud base. This may account for the failure of cold-weather cumulus, with summit temperatures as low as −20C, to produce showers. Such clouds may be wholly below the temperature at which the ice nuclei first act, and yet show no external trace of the presence of ice particles. Valuable data for this problem could easily be obtained by observing the base- and summit-level temperatures of convection clouds (noting those clouds which produce precipitation) and by attempting to detect the presence of small ice particles during flight through the supercooled tops of cumulus.

When the critical stage is passed and the ice particles have begun their growth as precipitation elements, the cumulus becomes transformed into the cumulonimbus. If the cloud ceases to grow soon after the critical stage is reached, the total number of ice particles formed remains very small, and slowly dissipating patches of water cloud may persist for a time near the summits, all the ice particles having fallen out. Usually, however, the top of the cumulonimbus becomes an ice cloud of considerable density, and it seems that the number of crystals finally formed must be substantially increased by a mechanism which Findeisen has begun to investigate [6, 7]. During the growth or evaporation of an ice particle, air flowing past it carries away minute charged fragments of ice (radius about 20 μ). These ice "splinters" are produced only by particles which are aggregates of small crystals, and are not formed at the surfaces of simple crystals or at glassy layers of ice. Little is known about the manner in which these splinters and their electrical charges arise, but Findeisen regards this process as the predominant source of thunderstorm electricity, and it must be of great importance in all ice clouds and mixed clouds as a means of ice-crystal reproduction without the aid of special ice nuclei. According to Findeisen's experiments [7], the rate of formation of ice splinters during the direct condensation of vapour into ice amounts to about 5 sec^{-1} from an ice surface of area 50 cm^2. At temperatures down to about −30C the number of ice particles growing upon ice nuclei is not likely to exceed 10^3 m^{-3}. If one-tenth of these have reached a radius of 10^{-1} cm, after about twelve minutes they alone will have produced as many ice crystals by the splintering process as were formed on the ice nuclei. During the growth of ice particles by coagulation with supercooled droplets, however, the rate of splinter formation is increased perhaps a thousand fold, and it therefore becomes clear that this process is of dominating importance in the ice-nucleus economy of all mixed clouds. The large numbers of crystals of insignificant falling-speed, and therefore of small size, which compose the extensive anvils of mature cumulonimbus are probably almost all produced without the aid of special nuclei. The splinter formation may be essential to the development of cumulonimbus over the oceans in polar air masses whose content of ice nuclei is likely to be extremely small.

Here it may be remarked that as yet little attempt has been made to work out in detail the growth conditions of cumuliform clouds and the modifications introduced by the appearance of the ice phase and the development of precipitation. Again it will be necessary to consider the speed of the condensation and coagulation processes. The rate of ascent of air in vigorous cumulus is of the order of 10 to 20 m sec^{-1}, greatly in excess of the speeds assumed by recent writers (*e.g.* Austin [2]), who have developed the concept of entrainment of air from the environment into the rising currents. The present writer doubts whether this materially affects the growth of large clouds, and believes that the simple parcel method of representing convection may still prove the most useful, but that it needs modification in the light of important factors which have been ignored so far, such as (1) the short life of individual clouds and the effect of their continual dissipation on the condition of the environment, (2) the effect of evaporation and mixing at the edges of small clouds, and (3) the effect of the rate of heating at the bottom of the convective layer. It is doubtful, in view of the predominant condensation of liquid water in rapidly ascending air, whether the latent heat of fusion plays any part in the development of cumulonimbus. The writer has shown (in work not yet published) that in dense clouds the larger ice particles pick up supercooled water at a rate which prevents its complete freezing, so that the precipitation elements have wet surfaces and a temperature not much below 0C. A proportion of the latent heat of fusion is thus used in warming the precipitation and is lost to the cloud. On the other hand, the accumulation of precipitation elements in the upper part of the cloud leads to the mechanical destruction of the updraught, and with the release of the precipitation a powerful downcurrent is initiated [3]. There is therefore reason to believe that in aerological work the development of cumulonimbus may best be followed on thermodynamic diagrams constructed wholly with respect to liquid water. Such diagrams are also appropriate in considering the formation of high ice-clouds, but it is advisable also to include lines showing the saturation

mixing-ratios over ice when the subsequent growth of these clouds is to be examined.

Medium- and High-Level Clouds. Medium- and high-level clouds are formed in air which has not recently been near the ground and in which the ice nuclei are likely to be especially rare. Nevertheless it appears that the number of active nuclei always increases as the temperature falls. In general, apparently stable supercooled clouds are encountered at temperatures between 0C and about −10C. At somewhat lower temperatures supercooled clouds commonly survive the appearance of ice particles, which quickly grow and fall out as *Fallstreifen*, while at temperatures below about −30C droplet clouds become increasingly rare. Usually they appear only fleetingly before becoming rapidly transformed into pure ice-clouds.

Clouds of the altocumulus type are often very shallow and have very small liquid-water contents, and since the air in the cloudlets is constantly replaced, any ice crystals present will grow slowly and may for some time produce no effect which would reveal their existence to an observer on the ground. However, the ice crystals will be slower to evaporate than the droplets, and it is conceivable that in a shallow layer of convective circulations they may be drawn back into cloudlets a number of times, increasing in size and number all the while (because of splinter formation). It may be in this way that parts of altocumulus sheets are transformed into thin ice clouds some time after their formation; the ice particles frequently fall out of the damp layer in which the cloud had formed and evaporate in the drier air beneath, so that the final effect of the action of the ice nuclei is the dissipation of the cloud. This is the typical result of the introduction of crystal nuclei into shallow, slightly supercooled clouds. On some occasions altocumulus appear to become infected with ice particles over a very limited area, a hole developing in the cloud layer over a patch of *Fallstreifen* [18]. This may be due to the irregular distribution of the ice nuclei, or may be the result of splinter formation following the action of a very few ice nuclei in a particularly large cloudlet with lower summit-temperatures than are reached in the other cloudlets. It is common for some cloudlets of a high altocumulus to produce *Fallstreifen*, and it is characteristically the larger cloudlets which are so affected, not only because of the lower summit-temperatures at which more ice nuclei act, but also because in the larger cloudlets the growth of ice particles is favoured by the higher water content, the greater depth of cloud, and the stronger upcurrent, which prevents early precipitation.

While falling through the ice-supersaturated layer beneath the base of the parent cloud, *Fallstreifen* particles continue their growth. In general, the depth of this layer and the degree of supersaturation in its upper part increase with fall of temperature. High, delicate altocumulus clouds may disappear in producing faint short trails of ice particles, which then continue to grow while descending some thousands of feet, finally becoming dense ice clouds.

Clearly no fundamental distinction need be made between ice clouds originating in this way and cirrus forming from a condensation region at lower temperatures with no trace at all of a parent cloud containing liquid water. At temperatures below about −40C the ice nuclei seem to be significantly more abundant; water droplets have a very short life and are produced only in clouds of convective or lenticular type in which updraughts of the order of a few meters per second occur. Nevertheless, the number of ice particles found in cirrus clouds by Weickmann [24] was on the average only about $\frac{1}{6}$ cm^{-3} (droplet clouds usually contain several hundreds of droplets per cubic centimetre), and it is evident that the crystals are so few that most may reach a size at which they have an appreciable falling-speed (about $\frac{1}{2}$ m sec^{-1}). Weickmann [23] photographed the particles of high clouds and obtained striking evidence of the large size of cirrus particles, which typically consisted of radiating tufts of long prisms containing hollows and air-enclosures. This crystal form is an indication of rapid growth at high supersaturation, although the cause of the grouping into tufts is unknown. The crystals of cirrostratus were smaller and had simpler "complete" shapes. For this reason cirrostratus clouds produce halos much more readily than cirrus, in which halo details are rather rare.

In the initial stages of their formation cirrus clouds are not typically fibrous. Frequently the first details to appear have a patterned or granulated structure resembling that of cirrocumulus, which slowly coarsens and fuses while thin *Fallstreifen* develop, or they may contain waves and ripples. With steep lapse rates, shreds of clouds very much like fractocumulus occur; these sometimes develop into true cumuliform clouds several thousands of feet deep, but more often they spread out and become less dense, gradually becoming transformed into patches of *Fallstreifen*. Bright, iridescent colours may be seen in such newly formed clouds when they are very near the sun, and there seems no reason to doubt that initially they contain at least some water droplets. On the other hand, the typical *Fallstreifen* forms of cirrus, although they often occur with shreds of cloud near their summits which are not truly fibrous and look like the vestiges of a parent cloud, do not always develop from such shallow initial forms. It is likely that these clouds often arise when especially suitable ice nuclei just forestall a more general condensation, so that a very few crystals are able to reach an unusual size and falling-speed. Weickmann [24] has remarked on the extraordinarily small number of large particles (diameter up to about 400 μ) which are found in the trails of cirrus uncinus.

The shape of *Fallstreifen* trails is determined by the wind shear at their level and the falling-speed of their particles. At first the trails are usually almost vertical, but near their lower ends they become almost horizontal as the falling-speed of the evaporating particles dwindles and a decrease of wind causes them to lag more and more behind the condensation region. A simple hook-shape (uncinus) can thus arise with only a gradual wind-shear, but some cirrus trails indicate a very rapid

change of wind with height, suggesting that frontal surfaces at these heights may be more sharply defined than is usually supposed.

There is a noticeable tendency for the lowermost parts of *Fallstreifen*, even in the cirrus levels, to become mammillated, presumably owing to the chilling of the air in the trail by the evaporation of the ice particles. The *Fallstreifen* path probably marks a stream of descending air whose motion is initiated by the drag of the particles and is sustained by this chilling. The downdraughts in precipitation are very pronounced in cumulonimbus and they play an important part in the development of these clouds by spreading out near the ground and acting like scoops to help fresh cloud growth near the shower borders [3]. A similar process on a smaller scale, may occur near *Fallstreifen* trails and lead to lines of cloud arranged along the direction of the wind shear.

THE RELATION OF ICE CLOUDS TO CONDENSATION MECHANISMS

High-level clouds are produced by the large-scale lifting of air masses, especially at frontal surfaces, but it is unusual for the clouds to appear as uniform sheets. Thus the first clouds to appear ahead of a warm front are characteristically isolated cirrus containing uncinus forms, and the overcast of cirrostratus which follows often contains *Fallstreifen* or is of very irregular density. Schwerdtfeger [19] attempted to show that isolated cirrus are the result of cooling in the high troposphere with the production of a convective layer, whereas cirrostratus is produced by the lifting of stable air masses over frontal surfaces. It would be possible to examine this view more critically now that frequent upper-air soundings are available. Isolated cirrus clouds could arise in uniformly lifted air because of the irregular distribution of especially suitable ice nuclei or as the result of small disturbances in the general air flow. Waved or rippled detail and lenticular-like patches of cloud are often seen in cirrus systems, and it is likely that orographic disturbances frequently reach the cirrus levels and trigger the formation of clouds. Whereas lenticular clouds consisting largely of droplets remain stationary at the crests of standing waves, similar clouds containing many crystals could survive descent in the troughs and continue their growth in ice-supersaturated layers.

The growth of individual ice clouds in deep supersaturated layers may extend over a few hours before general decay sets in, and during this time the clouds may be carried hundreds of miles from their birthplace. The air may remain supersaturated and the growth be maintained even though the mechanism which lifted the air has ceased. Thus, the mere presence of ice clouds, whose decay is correspondingly protracted, by no means indicates an active condensation mechanism, and in this respect they differ from droplet clouds, which evaporate within a few minutes of the cessation of the condensation mechanism. *Only young ice clouds accompany active condensation mechanisms; they may always be recognised by the presence of cirrocumulus (droplet cloud) or patches of granular and flocciform cloudlets.* Beautifully arranged delicate *Fallstreifen* forms are also young ice clouds—as they age they degenerate into diffuse streaks or fibres of lifeless appearance, usually described as cirrus filosus.

As far as the writer knows, no methods of forecasting the extent and thickness of ice clouds are in use other than the making of estimates based on recent measurements or the relating of the clouds to frontal systems in accordance with textbook models. It is unlikely that much more can be done at present, as forecasting high cloud formation is essentially a matter of forecasting the vertical displacement of air at high levels, about which very little is known. More accurate measurements of humidity than are now made at these heights would probably also be required. A further difficulty is that abundant cirrus clouds often occur far from fronts and appear to have their origin in disturbances existing entirely above the surface layers. Thus cirrus systems may be seen to move with shallow cold "pools" and troughs which are clearly shown on 300-mb charts (but not on 500-mb charts). The occurrence of cirrus well ahead of cold fronts is also unexplained, but its presence seems related to the movement of the front, and its disappearance is symptomatic of the retardation of the front. However, there are few cirrus systems which are not highly organised, perhaps containing great parallel bands of cloud or long sharply defined clearing-edges and with *Fallstreifen* indicating systematically changing wind shears, so that it is certain that the mechanisms producing the systems have coherent structures which may be forecast once they are understood. It is unfortunate that the sharpness of the boundaries of cirrus systems is so often masked on synoptic charts by the reporting of clouds near the horizon which are at a great distance—these reports might profitably be made only by isolated stations.

CONCLUSION

It is hoped that many important problems of cloud physics will have been sufficiently indicated in the paragraphs above. It may be remarked that one class of problems concerns the microphysics of cloud particles, and another the macrophysics of the occurrence and growth of clouds, and that usually the two are very intimately related, especially at temperatures below 0C. Recently, substantial progress has been made with the microphysics, but careful attention must now also be directed to the study of the formation and growth of clouds in relation to air movements. Here valuable information may be expected from the researches on atmospheric dynamics which are being pursued so vigorously.

REFERENCES

1. Aanensen, C. J. M., "Unusual Condensation Trails." *Meteor. Mag.*, 77: 17–18 (1948).
2. Austin, J. M., "A Note on Cumulus Growth in a Non-saturated Environment." *J. Meteor.* 5: 103–107 (1948).
3. Byers, H. R., and Braham, R. R., Jr., "Thunderstorm Structure and Circulation." *J. Meteor.*, 5: 71–86 (1948).
4. Cwilong, B. M., "Observations on the Incidence of Super-

cooled Water in Expansion Chambers and on Cooled Solid Surfaces." *J. Glaciol.*, 1: 53–57 (1947).

5. Findeisen, W., "Die kolloidmeteorologischen Vorgänge bei der Niederschlagsbildung." *Meteor. Z.*, 55: 121–133 (1938).
6. —— "Über die Entstehung der Gewitterelektrizität." *Meteor. Z.*, 57: 201–215 (1940).
7. —— und Findeisen, E., "Untersuchungen über die Eissplitterbildung an Reifschichten." *Meteor. Z.*, 60: 145–154 (1943).
8. Findeisen, W., und Schulz, G., "Experimentelle Untersuchungen über die atmosphärische Eisteilchenbildung, I." *Forsch. ErfahrBer. Reichs. Wetterd.*, Reihe A, Nr. 27, Berlin (1944).
9. Fournier d'Albe, E. M., "Some Experiments on the Condensation of Water Vapour at Temperatures Below 0°C." *Quart. J. R. meteor. Soc.*, 75: 1–14 (1949).
10. Krastanow, L., "Über die Bildung der unterkühlten Wassertropfen und der Eiskristalle in der freien Atmosphäre." *Meteor. Z.*, 57: 357–371 (1940).
11. —— "Beitrag zur Theorie der Tropfen- und Kristallbildung in der Atmosphäre." *Meteor Z.*, 58: 37–45 (1941).
12. Ludlam, F. H., "The Forms of Ice-Clouds." *Quart. J. R. meteor. Soc.*, 74: 39–56 (1948).
13. Meteorological Office, Air Ministry, London, "Airframe Icing at Low Temperature." *Mem. meteor. Off., Lond.*, 494 (1946).
14. Palmer, H. P., "Natural Ice-Particle Nuclei." *Quart. J. R. meteor. Soc.*, 75: 15–22 (1949).
15. Rau, W., "Gefriervorgänge des Wassers bei tiefen Temperaturen." *Schr. dtsch. Akad. Luftfahrtforsch.*, 8: 65–84 (1944).
16. Regener, E., "Versuche über die Kondensation und Sublimation des Wasserdampfes bei tiefen Temperaturen." *Schr. dtsch. Akad. Luftfahrtforsch.*, 37: 15–24 (1941).
17. Schulz, G., "Die Arbeiten und Forschungsergebnisse der Wolkenforschungsstelle des Reichsamtes für Wetterdienst in Prag." *Ber. dtsch. Wetterd., U. S.-Zone*, Nr. 1, SS. 3–12 (1947).
18. Schumacher, C., "Beobachtungen an einer Altokumulusdecke." *Z. angew. Meteor.*, 57: 214–220 (1940).
19. Schwerdtfeger, W., "Über die hohen Wolken." *Wiss. Abh. D. R. Reich. Wetterd.*, Bd. 5, Nr. 1, 34 SS., Berlin (1938–39).
20. Wall, E., "Die Eiskeimbildung in Lösungskernen." *Meteor. Z.*, 60: 94–104 (1943).
21. Wegener, A., *Thermodynamik der Atmosphäre.* Leipzig, J. A. Barth, 1911.
22. Weickmann, H., "Experimentelle Untersuchungen zur Bildung von Eis und Wasser an Keimen bei tiefen Temperaturen." *Zent. wiss. Ber. Luftfahrtforsch., Forsch. Ber.* Nr. 1730, Berlin-Adlershof (1942).
23. —— "Formen und Bildung atmosphärische Eiskristalle." *Beitr. Phys. frei. Atmos.*, 28: 12–52 (1945).
24. —— "Die Eisphase in der Atmosphäre." *Ber. dtsch. Wetterd., U. S.-Zone*, Nr. 6, 54 SS. (1949). (Wölkenrode Report and Translation No. 716, Feb. 1947, Library Translation No. 273. Ministry of Supply, Sept. 1948.)

THERMODYNAMICS OF CLOUDS*

By FRITZ MÖLLER

Gutenberg University at Mainz

OBSERVATIONAL BASIS

The radiosonde epoch, in which—from an aerological point of view—we are now living, is not favorable to a thermodynamical examination of clouds. The great age for thermodynamics and aerology was about twenty or twenty-five years ago, when Shaw [37], Stüve [40], Robitzsch [31, 32], Rossby [33], and others after them [47] computed and used their thermodynamic diagrams to make a careful examination of each sounding of the free atmosphere. Today it is still customary to plot each ascent on such diagrams, but only in rare cases is there a thorough analysis of the stability, the available energy, etc. There are several reasons for this:

1. The observational data are incomparably more numerous. Where formerly one had five ascents a day, one now has one or two hundred. Therefore, each sounding cannot be studied with as much attention and care. The analysis is made for synoptic purposes in most cases and is therefore exclusively dynamic in character. In consequence of this:

2. The information which is transmitted is arranged to permit a quick, dynamic-synoptic representation, in other words, the drawing of isobaric surfaces. In the German weather service, thermodynamic measures of energy have been calculated and transmitted for every ascent for about ten years. These measures have not been adopted internationally, but their use would facilitate synoptic-thermodynamic investigations.

3. We have reason to assume that the instruments now in use, the radiosondes as well as the instruments designed for use with aircraft, do not represent the temperature distribution properly. Twenty years ago, Lautner [19] used sensitive resistance thermometers and found, in flying through a subsidence inversion, a discontinuous temperature increase upward, in other words, a really ideal inversion. It is very likely that most inversions have the same form, but that, because of the lag of the instruments, this can never be established. In the writer's opinion, an examination of the tropopause by a trained meteorological observer, who would test it with thermoelements or resistance thermometers, would yield many surprises.

4. The most important obstacle to the successful development of cloud physics is related to this problem of inadequate instrumentation: The radiosonde furnishes only an inflexible, unmodifiable cross section of the atmosphere; the observer in an airplane is much more efficacious for cloud investigations. He can supplement the vertical pressure, temperature, and humidity curves by what he sees: the form and character of the clouds, vertical stratification, horizontal extent, spatial density, changes in the course of time, etc. He can circle or penetrate an interesting cloud, taking along various special instruments or even an entire laboratory; this cannot be done with radiosondes except with tremendous difficulty. For this reason, airplane ascents and weather reconnaissance flights with meteorologically well-trained observers furnish invaluable observational material. Furthermore, such flights are of indirect value, because, in the process of observing, striking new problems always occur which are much harder to detect when one is merely working at one's desk. Observation flights will always be necessary for special problems; soundings alone will never suffice.

PREREQUISITES FOR THE IDEAL MOIST-ADIABATIC PROCESS

The basic problem around which the thermodynamic investigation of clouds must be centered is the moist-adiabatic process. This holds true for stratus as well as for cumulus clouds. Besides this macrophysical problem, there are microphysical processes whose effect is by no means restricted to invisibly small elements. This group of processes will be treated in a later section of this article. The classical theory of the moist-adiabatic process has recently been supplemented by various considerations which take into account the influence of the environment of the vertically moving particle. In this connection the slice method and the concept of lateral entrainment may be mentioned here. Further additions to the theory will be required in order to describe fully the basic process of the thermal reaction of a cloud. Nevertheless, the simpler method still furnishes some interesting viewpoints which are worthy of consideration.

The prerequisites for the ideal moist-adiabatic process are (1) the heat of condensation is exchanged so rapidly between condensing water vapor or evaporating water and the air that no temperature differences will develop; (2) there is no transfer of heat between the parcel and the environment; and (3) there is no transfer of mass, either water or air, between the parcel and the environment.

Heat Exchange between the Cloud Elements and the Air. Findeisen [10] has carried out a numerical investigation of the heat exchange between droplets and the air. He has come to the conclusion that under normal cloud conditions the temperature difference between the air and the droplets is not much greater than 0.2C. However, for clouds of low density, that is, those with a very small number of droplets or crystals per unit volume (as in the upper parts of altostratus or cirrus), the distances along which heat is transferred through diffusion or conduction are so large that temperature differences may reach several degrees. The vertical movement of the air can then proceed almost

* Translated from the original German.

dry-adiabatically, or the parcel may even be cooled by radiational heat loss (see below, also p. 203).

Heat Conduction in the Interior of the Cloud. Cloud elements can be heated either by conduction or by radiation. Conduction inside a cloud can be neglected entirely if the cloud has a homogeneous horizontal temperature distribution and a moist-adiabatic or at least a uniform lapse rate of temperature. It is known through the investigations of Peppler [27] that nimbostratus clouds frequently have a "laminated" structure, in other words, they are characterized by numerous small superimposed inversions and isothermal layers. If this observation holds true for a closed nimbostratus mass and not for the dissolving region to the rear of it (as seen from the ground these two conditions cannot be distinguished), then the turbulent heat conduction in the vertical can lead to deviations from the moist-adiabatic process in the interior of the cloud. However, no deviation from a smooth temperature distribution is to be expected in the interior of a cumulus cloud.

Evaporation and Heat Conduction in the Environment. The edge of the cloud is a mixing zone between the cloud itself and the environment, where, no doubt, there are considerable deviations from the ideal moist-adiabatic process as a result of the turbulent heat conduction across the horizontal and vertical cloud boundaries. Evaporation of the droplets into the neighboring dry air and the increased water-vapor content of the air lead to a loss of heat from the border zones of the cloud, which then become colder than the environment. This process can be of importance on the edges of cumulus clouds and on stratus cloud surfaces which lie under inversions. This fact was pointed out by von Bezold [4] and later by Robitzsch.

Radiation Processes. The heat transfer by radiation from and to the cloud is just as important as is heat conduction. In general, the clouds absorb only a small part of the sun's radiation. About 50 to 70 per cent of the incoming radiation is reflected from the upper cloud surfaces [14]; most of the remaining 30 to 50 per cent is transmitted and only a very small portion is absorbed. This absorption is unimportant in comparison with the other thermal processes. However, it is the only process which shows only a minor decrease from the edge to the interior of the cloud; although insignificant, the absorption is rather uniformly distributed throughout the entire cloud mass [1].

The long-wave radiation, however, acts in an entirely different way. The surface of a cloud can be considered as a black body for the wave-length region from 4 to 100 μ. Even very thin layers of water vapor and carbon dioxide in the surrounding air absorb almost totally at most of these wave lengths; however, between 8 and 13 μ, even very thick air layers with an abundance of water vapor and carbon dioxide have almost no absorption. Therefore, the clouds can produce strong surface emissions in this spectral range and lose much heat. On the undersurface of the cloud layer where the radiation of droplets is opposed by that of the usually warmer earth's surface acting as a black body, the clouds receive more heat than they radiate and therefore become heated. Numerical values for this process can be found in the article on long-wave radiation in this Compendium.[1] In any case, these processes do not penetrate very deeply into the cloud mass; they influence only a very thin surface layer, whose thickness can be estimated as about 10 m in the case of a cloud with high water content. Although these processes are characteristic of the surfaces of stratocumulus or cumulus clouds, the processes in thin altostratus or cirrus are entirely different. Here, the individual heat-radiating particles, the ice crystals or water droplets, are so far apart from each other that radiation exchange with the environment can take place even through thick cloud layers. The heat loss upward and the heat gain from below therefore take place rather uniformly in the entire mass of thin, veil-like clouds. Here, a clear-cut, moist-adiabatic process that can be defined thermodynamically is no longer possible, because the heat balance of the cloud is considerably influenced by radiation. In the lower layers of the troposphere, where the clouds have a relatively high water content, the rule holds true that the moist-adiabatic process is undisturbed by radiation into some distance from the cloud's edge. In the same way, other surface processes (such as evaporation and "outer" heat conduction on clouds) cannot have any effect within extensive cloud masses or in smaller clouds characterized by large vertical velocities. Thin and scattered clouds depart even further from the moist-adiabatic process. The effect of radiation upon various cloud types will be treated below.

Mixing Processes. It seems to be obvious that an air parcel which is lifted moist-adiabatically does not undergo any mixing with the surroundings, and one is easily inclined to take this for granted. However, there is no proof of this. In cumulus, the entire cloud mass is not uniformly lifted; the parcels in the upper portion of the cloud are retarded, whereas in the lower portion new masses move up and penetrate the higher layer. It is not at all certain in this process that the subsequent parcels from below move along the same moist-adiabatics. Probably, their equivalent potential temperature is higher, and the mixing of the masses disturbs the thermodynamics.

Furthermore, it must be considered that the requirement of the absence of mixing also concerns the liquid and the solid constituents of a cloud. In the normal pseudo-adiabatic process we assume that all liquid water drops out as soon as it is produced. This, however, would mean a loss of mass; the ejected mass of liquid water takes its part of the entropy with it, that is, the process is no longer isentropic. Similarly, one must also consider that in a precipitating cloud the crystals and the growing water droplets are falling relative to the air in the cloud. Relatively speaking, they thus enter the individual ascending parcels of moist air from above, collect some cloud droplets, and then leave the air parcel. All in all, the process is by no means isentropic.

1. Consult "Long-Wave Radiation" by F. Möller, pp. 34–44.

THE MOIST-ADIABATIC PROCESSES

Ascending Motions. Despite the influence of the processes just mentioned, the classical theory of the moist-adiabatic still represents the basic process upon which all refinements must be made. The fundamental concept has been checked by Schnaidt [36] and Bleeker [6] in careful theoretical investigations. Schnaidt no longer distinguishes between the obsolete rain, snow, and hail stages. He defines, according to Dinkelacker's suggestion [9], (1) the cloud-adiabatic, in which all condensed water is retained in the ascending mass and is carried along, (2) the general pseudo-adiabatic, in which part of the water drops out, and part is carried along, and (3) the special pseudo-adiabatic, in which all the water drops out. All three stages can take place with condensation or sublimation. The first and third processes do not occur in nature. The normal process is the second one which, however, cannot be theoretically evaluated, since no numerical estimate can be made of the amount of precipitation elements which drop out or are carried along. The temperature curve for the three cases is different. For the same pressure decrease, the temperature decrease is smaller for the cloud-adiabatic than for the pseudo-adiabatic. The case which can be handled most easily from a numerical standpoint does not permit a physical interpretation: In the pertinent equation, the specific heat of the mixture of air, water, and water vapor contains a term cM (where c is the specific heat of the water, and M is the entire content of water plus water vapor); this term is set equal to zero. The resulting equation can be integrated as a whole and is the basis for the diagrams of Rossby and Shaw; however, the trend of the curve yields a still more rapid temperature decrease with decreasing pressure than do the cloud- and pseudo-adiabatics. Dinkelacker has called this the "main-adiabatic." Distinguishing these cases and their computation should not be considered mathematical sophistry; the selection of the proper adiabatic as the basis for a graphical determination of instability criteria or for the precise computation of the lateral entrainment can become very important.

Descending Motions. It must be noted that the moist-adiabatic is also being followed by air parcels that contain liquid water and have a vertical, *descending* motion [26]. One example of this process is the descending motion which the falling precipitation establishes below a thundercloud. As the air descends moist-adiabatically to the ground, it loses part of its charge of water droplets by evaporation; the temperature difference between this air and the dry-adiabatically stratified air surrounding the thunderstorm becomes increasingly negative toward the ground. The instability thus released is, to a large extent, responsible for the kinetic energy of the squall that is, so to speak, squeezed out from the region of falling precipitation. Suckstorff [41, 42] has studied these processes very thoroughly.

Wagner [43], in a careful study, has recently drawn attention to a second phenomenon, the formation of mammatus clouds. If the top of a thundercloud takes the form of an anvil, then this part of the cloud air is in thermal equilibrium with the dry air underneath it; it is therefore neither heavier nor lighter than the latter. Through a renewed swelling of neighboring cumulus tops, the anvil is quite often forced to descend, particularly in the rear of the thunderstorm. The layers of dry air below the anvil, and the cloud's air, are descending simultaneously, the former dry-adiabatically, the latter moist-adiabatically. Therefore, a temperature discontinuity develops, with cold air above warm air, in other words, a region of instability with an almost horizontal boundary surface. The mammatus pouches form either through horizontal differences of temperature or water content or through dynamic perturbations of the boundary surface. As they penetrate deeper into the dry air underneath, they become increasingly colder than their environment. Under certain conditions they can even detach themselves from the upper anvil clouds. Similar extended layers of altostratus mammatus are frequently observed on the north side of low-pressure areas (in central Europe, for instance, to the west of *Vb*-lows which are moving from south to north), where the warm air masses have ascended above the lower air and where divergence in the horizontal flow may cause a subsidence of the entire stratified mass of air.

EXTENSION OF THE THEORY OF CONVECTIVE CLOUDS

In the last decade, two fundamental lines of investigation have been followed with the purpose of widening the theory of the dynamics of convective clouds. On the one hand, we have the theories of Bjerknes [5] and Petterssen [28] dealing with the descent and dynamic heating of the air masses surrounding cumulus clouds; on the other hand, we have investigations into the entrainment of the surrounding air in the rising movement of cumulus, an approach developed particularly by Austin [2]. Both ideas may be based on the observation that an indication of marked "moist-instability" from the sounding is not always accompanied by a strong formation of convective clouds. On the contrary, even with large temperature lapse rates convection is often missing or only very feebly developed. Both theories, therefore, lead to a higher instability lapse rate than does the parcel method.

The Slice Method. The simple theory of moist-adiabatic change does not take into account the fact that an ascending air parcel is surrounded by air which must descend in order to preserve continuity of mass. This becomes quite clear when we consider that the exact theory does not describe the temperature change with respect to height, but with respect to pressure; in other words, it may also hold true for air in the chamber of an air pump. Actually, the descending air masses in the environment of a convective cloud are warmed so that the temperature excess, which the ascending air will take on because of the released heat of condensation, becomes smaller. This decrease in temperature difference is insignificant because the temperature excess is so large when the ascending cloud columns are small, and a slow descent takes place over the large spaces between the columns. However, the

temperature excess becomes considerably smaller with an increase in the downward displacement of air or an increase in the quotient M'/M where the ascending and descending masses of air are M' and M, respectively. These ideas of Bjerknes and Petterssen have already been adopted in a number of textbooks, therefore a detailed derivation is unnecessary here. It ought to be borne in mind, however, that the smaller the quotient M'/M, the more valid are the instability criteria of the parcel method according to Shaw, Normand, or Stüve. Furthermore, in this case the "moist–instability" is much greater with the same lapse rates, and a much smaller lapse rate in the surrounding air can still be considered as an unstable stratification. As the ratio M'/M becomes larger, it becomes increasingly imperative to take into account Petterssen's slice method, since the energy associated with the instability, which is decisive for the possible vertical velocities, becomes so much smaller, and a greater lapse rate corresponds to neutral equilibrium.

Petterssen and collaborators [29] have been able to demonstrate the usefulness of the slice method by an excellent statistical investigation; according to this new method the maximum possible height of convective clouds is much smaller than it is according to the parcel method. In fact, the heights attained by the tops of cumulus clouds have been found to be in good agreement with the heights predicted by the slice method. The parcel method yields no clues concerning the amount of cloudiness, whereas the slice method provides some information which corresponds to the actual cloudiness, although the correspondence displays considerable scattering. Nevertheless, it should not be overlooked that in other respects the parcel method yields useful information. For instance, it answers questions such as how high the condensation level is, and whether or not cumulus clouds are to be expected at all. However, the degree of instability is less important here than is the stability of the stratification which has to be overcome initially.

Here, too, there seems to be a minor defect in the slice method. In establishing the criteria, it is assumed that the mass of air ascending per unit time equals the descending mass of air. This means that on the average no vertical component of motion will result over a sufficiently large horizontal area. It has already been proved, through the valuable investigations of Calwagen [7], that the most favorable condition for the occurrence of local summer showers is the existence of an old front, believed dissipated, which sometimes will be steered back into the region under observation. Such old fronts are simply minor convergence zones in the horizontal air flow; their presence requires the simultaneous existence of weak vertical motion over the given region according to the continuity of mass. Petterssen [29] implicitly comes to the same conclusion when he shows that 50 per cent of the cases of weak convection are coupled with cyclonic curvature or wind shear, whereas this holds true in 85–95 per cent of the cases of stronger convection. This, however, means convergence, although it may be only weakly developed. Then, however, the consideration of a resultant upward component (in other words, the prevalence of ascending masses M' over the descending masses M) leads, in turn, to an increase in the instability lapse rate, so that the parcel method yields a more accurate measure than does the slice method.

A further point should be taken into account when considering an isolated convective cloud. In the derivation of the slice method, the assumption must be made that a cloud tower has already penetrated the whole layer, and that the ascent and descent of the air masses take place uniformly throughout the entire layer. However, a growing cumulus cloud pierces vertically into the layer, so that the downward displacement and heating of the air around the intruding head are relatively slight, while around the lower portions of the cloud these processes are intensified. The parcel method is therefore more applicable to the conditions at the cloud top than to the conditions in the cloud column beneath. Perhaps this effect is the explanation of the detachment frequently observed in rapidly developing castellatus towers which ascend without a supply of air from below —much in the manner of free balloons.

The slice method makes the following particularly valuable contribution: When the ascending and descending particles are subject to the same changes of state, it predicts a released energy of instability that is significantly *larger* than would be indicated by the parcel method. This is the case in the dry convection below the condensation level, the violence of which can thus be explained. It is likewise true for the rapidly ascending masses in the interior of a cumulonimbus which give rise to vertical squalls. The theory of the formation of waterspouts and tornadoes (for which Koschmieder [17] considered the unstable spouting up of cloud particles within the cumulonimbus to be necessary) thus finds a welcome support. The difference between the two types of instability can be seen directly from cloud observations. In time-lapse motion pictures, taken by Mügge [24], one can recognize that the growth of cumulonimbus does not occur in the form of a symmetrical and uniform ascent. A cloud tower shoots up quickly and calms down; immediately a second one shoots up by its side, partly piercing the old one which, in turn, descends and may evaporate in its thinner portions. The vertical motion of the new tower ceases only when it has reached the height of the top of the old tower and finally comes to rest, whereupon another tower builds up. This rapid ascent and quick succession of the individual protuberances (which has also been demonstrated in the Thunderstorm Project) is caused by moist-adiabatic ascent within an environment that descends moist-adiabatically. The slow lifting of the top levels of the entire multiple cloud structure manifests the restricted energy of moist-adiabatic ascents in a dry environment.

The Lateral Entrainment of Air. A second and older assumption by Calwagen [7] would indicate that cumulus convection becomes more difficult when the air to be penetrated is particularly dry. The synoptic investigator who is familiar with this explains it as the result of

subsidence causing the dry air and the divergence of the horizontal flow. The statistics of Petterssen and his collaborators, mentioned above, also confirm this. It is, therefore, surprising to note that, according to the most recent investigations, this hindrance to convection is explained by a lateral entrainment of dry air masses in the ascending cloud. This theory is already well developed [2] and is treated in a separate article in this Compendium.[2] However, neither the ordinary observations of a cumulus cloud nor the processes visible in time-lapse motion pictures lead to the conclusion that a cumulus cloud is fed in any other way than by the entrance of air masses through its base. Indeed the cauliflower-like forms of cumulus have been considered as proof that no mixing with the surrounding air takes place. At best, only the barrier layers that are pierced by cumulonimbus clouds and surround the cloud towers like collars as, for example, a wreath of stratocumulus cumulogenitus, can be considered as locations of substantial feeding of the cloud by lateral entrainment. This takes place only at discrete heights, and not continuously at all heights.

Theoretical computations based on observational material show the lateral supply of mass and the liquid water content as a function of height. The second quantity, which can be measured from an airplane, should serve as a criterion for testing this theory [39]. However, another point must also be considered: The theory has taken into account only the parcel method of the moist-adiabatic and has supplemented it by the consideration of the lateral entrainment. This leads to a temperature lapse rate within the cloud which is greater than the moist-adiabatic and which approaches the dry-adiabatic. We come to the result already presented above by the slice method of convection. Perhaps a combination of the two methods would lead to different ideas about lateral entrainment in cumulus clouds. However, a simultaneous test of the two theories would require a very extensive observational program.

Cellular Convection. Perhaps an entirely different method of studying the instability conditions of cumulus humilis may be necessary. The time-lapse motion pictures by Mügge [24] mentioned above indicate that a cumulus humilis is not a single swelling structure. The individual parts of a cumulus are in constant motion. In the vertical wind shear the cloud mass, which appears calm to the eye, is, in fact, rolling; condensing cloud masses ascend continuously in the rear, pass the summit of the cumulus, and then dissolve while descending in front. The cumulus is not to be considered as the top cover of a chimney of ascending warm air, but rather as the upper portion of a rolling ball or cylinder. If this picture is correct, then the laws of cellular convection [20] must control the formation of such convective structures [38]. In this case, the existence of convection currents or cells depends not only on the lapse rate's exceeding a certain value, but also on the vertical thickness of the layer in which the convection takes place. Furthermore, it depends on the viscosity and heat conductivity of the air (thus on the austausch coefficient) and, finally, on the horizontal dimensions of the various convection cells. The theoretical concepts of cellular convection, which have been successfully applied to the explanation of stratocumulus and altocumulus structure [21, 44], must then also be valid for cumulus humilis [45]. A complicated picture would result: The laws of cellular convection apply to the distribution of clouds, the parcel method is applicable to the condensation level, and the slice method applies to the vertical extent of the clouds. In fact, clouds of vertical development are formed in layers characterized by uniform convection, because Petterssen and his collaborators [29] found that in a region of cumulus the vertical wind shear and shift are particularly small while directly above the summit of a cumulus cloud the wind shear is four times larger than at lower levels. This fact seems to justify the use of the cellular method. However, as is quite often the case in meteorology, it is problematic whether the small wind shear is the cause of cumulus formation, or whether, inversely, the existence of strong convective mixing is responsible for the absence of a large variation of wind with height.

Radiational Influences on Certain Cloud Forms. The influence of long-wave radiation upon cloud surfaces has already been mentioned above. This influence is most important for stratocumulus and altocumulus. As a consequence of the heat supplied at the cloud base and the heat lost at the top, parcels of air in the lower portion of the cloud are heated, evaporate their liquid water, and rise, while particles near the top cool and descend. Thus the heating and cooling establishes an internally driven convection cell [22], which, according to Bénard, operates in conjunction with the cellular convection to break up the cloud layers into individual globules. This internal convection transports heat from the bottom to the top of the cloud; for this reason, Raethjen has called stratocumulus clouds "the thermodynamic singularity in the vertical flux of radiation" [30].

In the tropics, the effect of radiation on high cloud layers (when there are no low or middle clouds below) becomes so great that above approximately 14 km clouds can no longer exist. Even a broken cloud cover would receive so much energy from the warm surface of the earth that it would be subject to evaporation and dissolution in a short time [23]. The fact that the upper limit of cirrus clouds in the tropics is found 3–4 km below the tropopause can probably be explained by this process alone.

The effect of radiation on thunderclouds has not been so clearly determined. The summit of a high-towered cumulonimbus loses much heat by radiation. In daytime this cooling is probably cancelled in part by radiation from the sun, but toward evening or during the night the emission from the cloud tops prevails. It is possible that because of this emission a kind of cold convection from above will be initiated in the remnants of thunderclouds, which might lead to a revival of earlier thunderstorm activity in the late evening or

2. Consult "Cumulus Convection and Entrainment" by J. M. Austin, pp. 694–701.

during the night. In addition to this process there may be still other causes for the formation of the nocturnal thunderstorm maximum.

It is much more difficult to try to explain in this manner the nocturnal shower maximum over the oceans. In this case, no thunderclouds that could be reactivated are left over from the daytime. On the contrary, synoptic investigations and weather-reconnaissance flights made during World War II revealed that, during the day, there were no clouds over the ocean, whereas at night the region was densely covered with severe thunderstorms. This condition recurred for several days in succession. The synoptic situation, that is, the presence of a cold air dome, was favorable for convective clouds in these cases, but the initiation of the cloud formation is hardly attributable to long-wave radiation. One is much more inclined to ascribe it to a remnant of the diurnal temperature variation which, over the thermally passive ocean surface, would produce a diurnal instability variation that is opposite to that observed over the thermally active surface of the continent.

THERMAL PROCESSES ON CLOUD ELEMENTS

The cloud elements are the actual vehicles of thermal changes. As the droplets or crystals fall in the gravitational field or mix with neighboring air parcels containing a different concentration of cloud elements, they do not remain in the volume of air in which they formed. Thus, they are no longer in equilibrium with their environment; evaporation will occur if the new environment is drier, condensation if the environment is supersaturated; likewise, freezing or melting of the cloud elements may take place.

Evaporation of Cloud Droplets. Evaporation of cloud elements takes place on the surface of clouds. It has already been pointed out in the papers by von Bezold [4] and later by Robitzsch that the lower temperatures which are frequently observed when flying into cumulus clouds [16] may be caused by evaporation of cloud droplets into the drier environment. This process was studied quantitatively by Findeisen [12], who found that cloud droplets in an environment of 90 per cent relative humidity have a lifetime of only 2½ seconds before they evaporate. The conditions for evaporation seem to be especially favorable at the edges of cumulus clouds. In the strong wind shear that exists between the ascending cloud and the descending environment, violent turbulence and mixing of cloud air with the environment takes place. It has been observed quite frequently from gliders and airplanes that the turbulence is much stronger in the peripheral than in the interior parts of the cloud. However, photographic measurements of the ascending velocity of cumulus clouds [35] seem to show that the ascending motion in the outer portions is only slightly retarded, and that the air directly adjacent to the cloud is also lifted. The sinking occurs only some distance away from the cloud. The model proposed by Christians [8] concerning motion in cumulus clouds and comparison with the hydrodynamic jet stream according to Schmidt [34] support this assumption. Then, however, the relative humidity in the immediate vicinity of the ascending cloud tower cannot be very low, and the evaporation as well as the resultant cooling will therefore be insignificant.

Still less significant is the evaporation over the dome of an ascending cumulus cloud, for here the surrounding air is strongly lifted, and its relative humidity is increased. This can even lead to the well-known phenomenon of caps (*pileus*) which stay for a short while over the swelling cumulus head separated from it by a thin, cloud-free region, until fusion takes place. However, there may be a positive temperature difference between the moist-adiabatically ascending cloud top and the dry-adiabatically lifted air above it, a difference which favors evaporation. This, however, is opposed to the results of Petterssen and his collaborators [29] concerning the heights of cumulus tops. These heights are given approximately by the height at which the moist-adiabatic lapse rate becomes equal to the vertical temperature gradient of the surrounding air. Therefore, no large temperature differences and, in turn, no significant evaporation can be expected. On a non-swelling cloud, these temperature differences would be equalized by austausch processes or even reversed by radiation.

Condensation and Sublimation on Cloud Elements. The opposite process, the condensation of water vapor on cloud elements that fall from their original air volume, occurs everywhere in the cloud. However, until now, this process has been given little attention in theoretical studies of the structure of clouds and the lapse rate within them. According to the computation by Findeisen [11], condensation contributes very little to the formation of precipitation, particularly in comparison with the essentially much more effective "chain reaction" process of Langmuir's theory [18]; nevertheless, it cannot be neglected in the thermodynamics of clouds. Condensation can only occur when the falling cloud elements are colder than their new environment. This causes the saturated water vapor to become supersaturated with respect to the cooler droplets. The amount of this supersaturation and its dependence on the temperature difference has been given, for example, by Harrison [15]. The temperature difference itself, however, depends on the heat transfer between the air and the falling water droplets; this transfer is, in turn, influenced by the condensation process. It is probably for this reason that condensation is rather ineffective in the growth of cloud elements.

The sublimation of water vapor on ice crystals or small graupel pellets is much more intense, since cloud air which is saturated with respect to water is considerably supersaturated with respect to the falling solid particle, depending on the temperature. In this case, the supersaturation is so great that the difference in temperature of the crystals with respect to their surroundings can be neglected. The thermal reactions have been discussed by Harrison [15]. The downward increase of the fall velocity and diameter of the droplets and the increase in vapor content of the air cause more heat of condensation or sublimation to be released in the

lower parts of the clouds than in the higher parts. For the numerical computation of these processes on falling snow crystals it would be valuable to consider the investigations of Wall [46] and the indices of crystal growth introduced by him. The greater heat release in the lower cloud portions increases the vertical temperature gradient and augments the instability in all clouds which already contain falling particles of precipitation. Now, the slice method shows that when air ascends or descends moist-adiabatically within the cloud, the released instabilities and vertical motions are considerably larger than would be expected according to the parcel method. Therefore, these comparatively minor amounts of released heat have a certain importance for the thermodynamics and mechanics of the cumulonimbus as well as for the upgliding nimbostratus. A numerical calculation of all these processes would be very desirable.

Droplet Accretion and Graupel Formation. A quantitative evaluation of the thermal effects of accretion of small cloud droplets on falling ice crystals and hail grains has not been made. Only the accretion on falling rain droplets has been computed in the theory of "chain reaction" processes in convective clouds by Langmuir [18]. We will not go into any further details here. However, the thermal consequences will be stressed again. It is evident that the heat of condensation does not play a role in this accretion. In the formation of precipitation, accretion on falling droplets is more important than condensation, but it is of no significance in the thermal processes. This is similarly true of the accretion of supercooled droplets on ice crystals (formation of graupel). In this process only the heat of fusion is released, whereas the amount of heat released in the sublimation of water vapor on ice crystals is 8½ times greater per unit mass. The accreted masses, however, are considerably larger in the process of graupel formation than in sublimation or the formation of rime; therefore the quantities of heat involved may be comparable. Here, too, the effects of the growth and the increasing fall velocity of the droplets, etc., are such that larger amounts of the heat of fusion are released in the lower portions of the cloud, thus causing an additional instability [15].

This becomes particularly important for the intermittent formation of a thundercloud. It has already been emphasized that a cumulonimbus does not form all at once, but that only one cloud tower at a time builds upward, spreads to an anvil (*incus*), glaciates, and stops growing; immediately following and quite close to the old tower, a new one rises. At first this contains only supercooled droplets, but soon it mushrooms into the existing anvil consisting of ice crystals, thus providing the best possible natural seeding of the ascending cloud. Consequently, conditions are favorable for the precipitation of large drops, graupel, and hail; for whenever one cloud tower or cell has fulfilled its task of producing precipitation, the next one swells up, ready to catch the falling ice particles and carry them upward again. Thus the processes of graupel formation and the release of heat of fusion causes additional instability in each subsequent tower which enables it to ascend higher than the preceding one. Also noteworthy is the interaction between the upper glaciated part of a nimbostratus cloud, which effects the seeding, and the lower part, which contains a large supply of liquid water [25]. A separation of the two cloud parts from each other may, under certain conditions, occur at mountain barriers when the lower current (and thus the cloud) is held back, while the upper current passes over the top of the mountain (Bergeron [3]). This can also happen with warm fronts, where the rain areas will be either very narrow or very wide, depending on the variation of wind with height.

Melting Processes. Somewhat more complicated processes exist in or directly below the freezing level. If the cloud reaches below this level, melting of the falling ice particles will not immediately occur, but rather, intensified condensation and sublimation will take place. Thus, the instability will spread below the freezing level. Only at some lower level will the melting of the falling ice affect the thermal processes. This melting will consume so much heat that the air is considerably cooled and its temperature kept constant at 0C. In this layer an isothermal lapse rate and thus marked stability will develop. This isothermalcy has nothing to do with the hail stage of the old classification of the moist-adiabatic processes (rain, hail, and snow stages), for here we do not deal with temperature changes of an individual quantity of moist air moving either upward or downward vertically, but with air that can remain at rest while precipitation falls through it. A strong instability will develop below this stable layer if the lapse rate was previously adiabatic or less. Findeisen [13] has called attention to this instability and has shown that it frequently causes the formation of scud clouds below the rain cloud proper. At this level, there are squalls and strong vertical turbulence, which permit formation only of fractocumulus clouds, but not of a closed cloud layer [15].

PROBLEMS FOR FUTURE RESEARCH

The discussion in this article has been presented with the principal idea that, fundamentally, we are concerned with disturbances or modifications of the simple moist-adiabatic process. It should be the aim of further scientific investigations to combine all the separate stones of our mosaic into a coherent picture of the heat balance of the clouds, in particular, (1) the heat balance of clouds in general, (2) the heat balance of the individual types of clouds, and (3) the role of the individual cloud types in the heat balance of the entire atmosphere.

REFERENCES

1. Albrecht, F., "Theoretische Untersuchungen über den Strahlungsumsatz in Wolken." *Meteor. Z.*, 50:478–486 (1933).
2. Austin, J. M., and Fleisher, A., "A Thermodynamic Analysis of Cumulus Convection." *J. Meteor.*, 5:240–243 (1948).
3. Bergeron, T., "The Problem of Artificial Control of Rainfall on the Globe. II—The Coastal Orographic Maxima

of Precipitation in Autumn and Winter." *Tellus*, Vol. 1, No. 3, pp. 15–32 (1949).

4. Bezold, W. v., *Gesammelte Abhandlungen*. Braunschweig, F. Vieweg, 1906.
5. Bjerknes, J., "Saturated-Adiabatic Ascent of Air through Dry-Adiabatically Descending Environment." *Quart. J. R. meteor. Soc.*, 64:325–330 (1938).
6. Bleeker, W., "Beitrag zur Definition von Äquivalent- und Feuchttemperaturen." *Meteor. Z.*, 57:111–115 (1940).
7. Calwagen, E. G., "Zur Diagnose und Prognose lokaler Sommerschauer." *Geofys. Publ.*, Vol. 3, No. 10 (1926).
8. Christians, H., "Zur Dynamik der Cumuluswolke." *Beitr. Phys. frei. Atmos.*, 22:149–161 (1935).
9. Dinkelacker, O., "Die Feuchtadiabate." *Meteor. Z.*, 56:289–297, 484 (1939).
10. Findeisen, W., "Über Wasserdampfübersättigung in Wolken." *Beitr. Phys. frei. Atmos.*, 20:157–173 (1933).
11. —— "Zur Frage der Regentropfenbildung in reinen Wasserwolken." *Meteor. Z.*, 56:365–368 (1939).
12. —— "Das Verdampfen der Wolken- und Regentropfen." *Meteor. Z.*, 56:453–460 (1939).
13. —— "Die Entstehung der 0°-Isothermie und die Fractocumulusbildung unter Nimbostratus." *Meteor. Z.*, 57: 49–54 (1940).
14. Fritz, S., "The Albedo of the Planet Earth and of Clouds." *J. Meteor.*, 6:277–282 (1949).
15. Harrison, L. P., "Lightning Discharges to Aircraft and Associated Meteorological Conditions." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1001 (1946).
16. Kopp, W., "Über die Möglichkeit mit den gewöhnlichen aerologischen Hilfsmitteln die Übersättigung der Wolkenluft zu bestimmen." *Beitr. Phys. frei. Atmos.*, 16: 173–179 (1930).
17. Koschmieder, H., "Über Tromben." *Wiss. Abh. Reich. Wetterd.*, Bd. 6, Nr. 3 (1940).
18. Langmuir, I., "The Production of Rain by a Chain Reaction in Cumulus Clouds at Temperatures above Freezing." *J. Meteor.*, 5:175–192 (1948).
19. Lautner, P., "Mitteilung Nr. 1 über ein photographisch registrierendes Platindraht-Widerstandsthermometer." *ErfahrBer. dtsch. Flugwetterd.*, 7:218–226 (1933).
20. Lettau, H., *Atmosphärische Turbulenz*. Leipzig, Akad. Verlagsges., 1939. (See pp. 196–198)
21. Mal, S., "Forms of Stratified Clouds." *Beitr. Phys. frei. Atmos.*, 17:40–68 (1931).
22. Möller, F., "Labilisierung von Schichtwolken durch Strahlung." *Meteor. Z.*, 60:212–213 (1943).
23. —— "Wirkungen der langwelligen Strahlung in der Atmosphäre." *Meteor. Z.*, 61:264–270 (1944).
24. Mügge, R., "Zeitrafferaufnahmen von Wolken." *Forsch. Fortschr. dtsch. Wiss.*, 16:221–223 (1941).
25. —— "Das Wetter," Hann-Süring, *Lehrbuch der Meteorologie*, 5. Aufl. Leipzig, W. Keller, 1938–1951. (See pp. 777–778)
26. Normand, C. W. B., "Kinetic Energy Liberated in an Unstable Layer." *Quart. J. R. meteor. Soc.*, 64:71–74 (1938).
27. Peppler, W., "Über die Temperaturabnahme in Nimbuswolken." *Meteor. Z.*, 45:161–166 (1928).
28. Petterssen, S., "Contribution to the Theory of Convection." *Geofys. Publ.*, Vol. 12, No. 9 (1939).
29. —— and others, "Convection in Theory and Practice." *Geofys. Publ.*, Vol. 16, No. 10 (1946).
30. Raethjen, P., "Strömungsvorgänge und Wolkenformen." *ErfahrBer. dtsch. Flugwetterd.*, 2. Sonderband, SS. 147–158 (1932).
31. Robitzsch, M., "Die Verwendung der durch aerologische Versuche gewonnenen Feuchtigkeitsdaten zur Diagnose der jeweiligen atmosphärischen Zustände." *Arb. aero. Obs., Lindenberg*, Bd. 16, Heft C (1930).
32. —— "Ein neuer Vordruck für die Auswertung aerologischer Aufstiege." *Beitr. Phys. frei. Atmos.*, 18:228–233 (1932).
33. Rossby, C.-G., "Thermodynamics Applied to Air Mass Analysis." *M.I.T. Meteor. Pap.*, Vol. 1, No. 3 (1932).
34. Schmidt, F. H., "Some Speculations on the Resistance to the Motion of Cumuliform Clouds." *Meded. ned. meteor. Inst.*, (B) Deel 1, Nr. 8 (1947).
35. Schmidt, W., "Messung der Vertikalgeschwindigkeit an Wolken." *Meteor. Z.*, 33:232–235 (1916).
36. Schnaidt, F., "Über die adiabatischen Zustandsänderungen feuchter Luft, die abgeleiteten Temperaturen und den Energievorrat atmosphärischer Schichtungen." *Beitr. Geophys.*, 60:16–138 (1943).
37. Shaw, Sir Napier, and Fahmy, H., "The Energy of Saturated Air in a Natural Environment." *Quart. J. R. meteor. Soc.*, 51:205–228 (1925).
38. Skeib, G., "Betrachtungen zur thermischen Konvektion." *Z. Meteor.*, 1:178–183 (1947).
39. Stommel, H., "Entrainment of Air into a Cumulus Cloud." *J. Meteor.*, 4:91–94 (1947).
40. Stüve, G., "Potentielle und pseudopotentielle Temperatur." *Beitr. Phys. frei. Atmos.*, 13:218–233 (1927).
41. Suckstorff, G. A., "Beiträge zur Dynamik der Regenschauer." *Nachr. Ges. Wiss. Göttingen*, N. F., Bd. 2, Nr. 2 (1936).
42. —— "Kaltlufterzeugung durch Niederschlag." *Meteor. Z.*, 55:287–292 (1938).
43. Wagner, F., "Mammatusform als Anzeichen für Absinkbewegung in Wolkenluft." *Ann. Meteor.*, 1:336–340 (1948).
44. Walker, Sir Gilbert T., "The Forms of Stratified Clouds." *Quart. J. R. meteor. Soc.*, 58:23–30 (1932).
45. —— "Clouds and Cells." *Quart. J. R. meteor. Soc.*, 59:389–396 (1933).
46. Wall, E., "Über die Entstehung der Schneekristalle I." *Wiss. Arb. DMD i. franz. Bes. geb.*, 1:151–179 (1947).
47. Weickmann, L., *Über aerologische Diagrammpapiere*. Int. Meteor. Org.-Int. Aerol. Kommission, Denkschr. Berlin, 1938.

THE FORMATION OF ICE CRYSTALS

By UKICHIRO NAKAYA

Hokkaido University, Sapporo, Japan

When pure liquid water suspended in the air is cooled, it keeps the liquid state till about −35C. Liquid water at temperatures below the freezing point is called supercooled water. The supercooled water is transformed spontaneously into ice at about −35C. The rate of freezing is determined by the rate at which the latent heat liberated by freezing is removed. In the case of freezing of the water supercooled to s degrees centigrade, $s/80$ parts of the volume are transformed instantaneously into ice at the moment when freezing starts, because the latent heat liberated is 80 cal g^{-1}. The whole system, $s/80$ parts of ice plus $(1-s/80)$ parts of water, is warmed up to 0C, and the speed of later freezing is a function of the rate of removal of the latent heat liberated by the subsequent freezing. Altberg [1] studied the freezing of the rapids in Siberia, and found that most of them are supercooled to the order of −0.05C just before freezing commences.

The freezing of still water in nature starts from the surface, as the maximum density of water is at 4C. X-ray studies show that the thin ice plate obtained at the surface of still water is a single crystal, the orientation being such that the principal axis is perpendicular to the water surface. Under favorable conditions very large single crystals are obtained in lake ice. McConnel [10] observed single crystals as large as about one foot in diameter at Lake Davos.

In the case of water in turbulent flow or in a refrigerated vessel, crystallization starts at numerous points in the water and the crystals develop in random directions. The bulk ice thus obtained is composed of a mosaic of ice crystals oriented at random. Thus ordinary ice has a microcrystalline structure and is not a crystal in the usual meaning of the word.

Single crystals of ice are obtained in the most beautiful and complicated manner by the sublimation of water vapor. In nature they are observed as crystalline frost, window hoar, and snow crystals. In this article the descriptions are confined to the ice crystals obtained by the sublimation of water vapor, with special emphasis on the snow crystal.

THE FORMATION OF ICE CRYSTALS BY SUBLIMATION

Crystalline Frost. It is well known that frost has two structural forms: amorphous and crystalline. Amorphous frost is produced when the temperature is only slightly below 0C, or by the deposition of supercooled water droplets. Crystalline frost is formed at lower temperatures by the condensation of water vapor by sublimation on a solid body. The solid body may be a particle of ground snow or other substance. The crystalline frost produced on a new snow surface usually develops in a fern-like shape and is called surface hoar. It is often observed on cold winter mornings after a calm, clear night. Crystalline frosts are frequently observed also on the walls of snow cavities. Cup crystals and feather-like forms are typical of these crystals. They are called depth hoar. Detailed descriptions of these surface and depth hoars have been given by Seligman [13, pp. 46–77].

The crystalline frosts can be classified into five forms: (1) needle, (2) feather-like, (3) plate, (4) cup, and (5) dendritic. From the crystallographic point of view, each of these frost crystals has its corresponding type in the crystals of snow.

1. Needle form. This crystal often grows out from a wall of snow in the form of needles 0.2–0.5 mm in diameter and about 1 cm in length. Microscopic examination shows that it is an assemblage of parallel hexagonal columns. It corresponds to the columnar crystal of snow.

2. Feather-like form. This type is frequently observed. Sometimes it develops to 5–6 cm in length. Under a microscope it is seen to be composed of small hexagonal columns, some columns being attached at right angles to the sides of others. The corresponding snow crystal is also known.

3. Plate crystal. This type is sometimes observed hanging down from the snow ceiling of a cavity and also growing out in the air from an exposed object such as the wall of a wooden box. The form and structure are quite similar to those of one branch of the snow crystal called sector form.

4. Cup crystal. When completely developed this type takes the shape of a whisky glass of hexagonal form. But usually one side is not completed and appears to be rolled up. This is the most common type of depth hoar. The corresponding snow crystal is very rarely observed.

5. Dendritic crystal. Dendritic crystals are often observed, in addition to the surface hoar, among the crystalline frosts developed on the branches of trees when the weather is calm and the temperature is low. This type corresponds to the fern-like crystal of snow, the only difference being that snow lacks the minute internal structure resulting from sublimation. When the temperature is very low, it is sometimes possible to observe a remarkable frost crystal that is very similar both in form and in structure to one branch of a hexagonal crystal of snow.

The correspondence of the habits of snow and frost crystals makes it possible to infer the conditions for the formation of various types of snow crystals. Before the artificial production of snow crystals, this was the only way to study the mechanism of snow-crystal formation.

Window Frost. On crisp winter mornings, in northern countries, various crystals of ice are observed on the

window panes of houses. These ice flowers are of two kinds: window frost and window ice. The former is a product of the sublimation or freezing of minute supercooled droplets; the latter results from the freezing of a water film. Window ice is likely to be formed in a kitchen or bathroom, where the moisture is abundant and the indoor temperature is just below freezing. The excessive water vapor condenses on the pane as a thin water film, and the mosaic of ice crystals obtained by the freezing of this thin film gives window ice.

Window frost is quite different from window ice in its appearance. The most familiar type is shown in Fig. 1. It must be again classified into two kinds: the detached hoar crystal *H* and the sheet frost *S* in Fig. 1. The hoar crystal is a product of sublimation, and is formed by a mechanism similar to that which produces the snow crystals. The sheet frost, which always extends over some area of the glass pane, is a two-dimensional assemblage of minute ice granules that are formed by the freezing of supercooled droplets scattered over the pane.

The mechanism responsible for the formation of hoar crystals can readily be studied by observing the initial stage of artificial hoar formation. Experiments have been performed in a cold chamber laboratory. Several minutes after exposing the plate to supersaturated air, numerous infinitesimal droplets condense on the surface, and the glass plate takes on a blurred appearance. After a little while, many germs[1] of hoar crystals appear at diverse points and begin to grow. As soon as growth starts, the blurred surface around the crystal begins to clear, that is, the minute droplets in that region evaporate. The vapor pressure of supercooled water is higher than that of ice at the same temperature, so the water is evaporated from the droplets and condenses on the hoar crystal, resulting in the growth of the latter at the expense of the former. The process is shown remarkably well by slow-motion pictures, one frame of which is reproduced in Fig. 2.

When water droplets are frozen into ice granules without evaporating, they form sheet frost. Supercooled water itself does not freeze easily, but once it touches ice, it becomes solid almost instantaneously. Thus, if one of the series of droplets freezes for some reason or other, the action is relayed to all the rest of the droplets, which also freeze. This relaying action can be seen under a microscope of high magnification. From one frozen droplet, a thin streamer of ice extends, and the moment it touches the next droplet, the one thus

Fig. 1—Detached hoar crystals and sheet frost (×0.73).

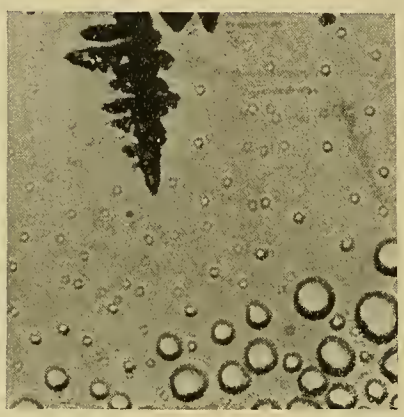


Fig. 2.—Growth of hoar crystals (×90).

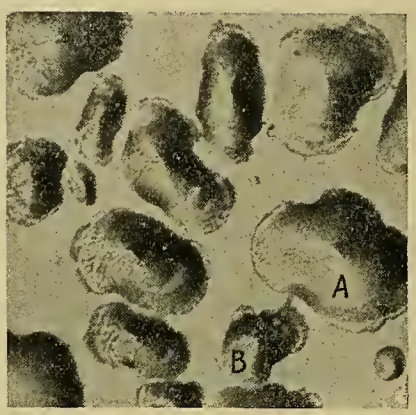


Fig. 3.—Relaying of freezing action in sheet frost (×210).

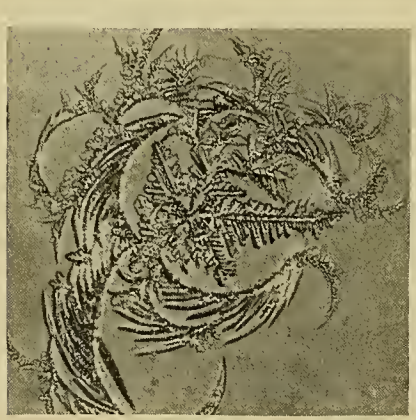


Fig. 4.—Hoar crystal formed on a glass plate covered by an alcohol film (×30).

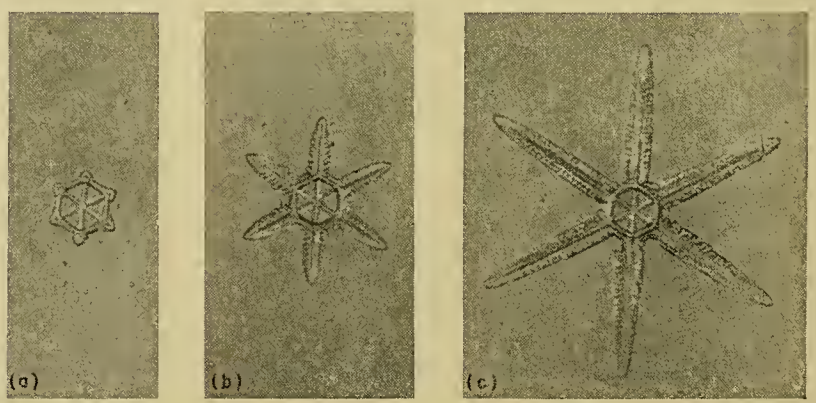


Fig. 5.—Three stages of a hoar crystal developed on a glass plate covered with paraffin film.

1. The term "germ" is used in this article to designate the primitive ice crystal.

touched freezes, while the streamer disappears; the behavior of this streamer which relays the freezing action is beautifully demonstrated by slow-motion pictures, one frame of which is shown in Fig. 3. This frame shows the instant when the droplet marked *A* becomes frozen by a streamer from the already frozen droplet *B*.

There are innumerable variations in the shape of window hoars observable in nature, for example, spiral patterns, odd arabesque designs, snow-like forms, etc. The ordinary glass plate is always covered with an invisible film of some organic substance. It is found that the combination of the effect of this invisible film and the atomic nature of ice gives rise to the variation observed in the patterns. When the glass surface is chemically clean, crystallization takes place very slowly. Even when the crystallization is almost complete, the hoar crystal is very thin and greatly distorted. The effect of an adsorbed film is very well demonstrated by exposing the glass plate to alcohol vapor for a short time, so that the surface is covered with an invisible film of alcohol molecules. Alcohol has a strong affinity with water, and the growth of the ice crystal may be expected to suffer a marked deformation. The results of such an experiment are as might be expected. One example is shown in Fig. 4.

The opposite effect can be observed on a glass plate covered with a thin, invisible film of paraffin wax. Since the water is repelled by the paraffin film, crystallization must be free from the effect of the surface. The glass plate is well cleaned and desiccated, and then exposed to paraffin vapor by being kept in a horizontal position 5 cm above the surface of molten (not boiling) paraffin wax. Under favorable conditions hoar crystals very much like snow crystals can be obtained on the plate. The best example is seen in Fig. 5. The three stages of development of a hoar crystal thus obtained are shown. It is apparent that the window hoar crystal also develops hexagonal symmetry if the effect of the base surface is eliminated.

Snow Crystals. The snow crystal is a solid product of precipitation formed in the atmosphere by sublimation of water vapor on minute solid nuclei. The symmetry of a snow crystal is due to its free development in a suspended state in air. The theory of crystal lattices can explain the symmetry of the angle between the faces, but it cannot touch upon the question of the symmetry of the macroscopic form of a crystal. The extraordinarily symmetrical pattern, sometimes observable in the hexagonal plane crystals of snow, is favored by the rotational motion around the vertical axis, while it is falling in a nearly horizontal position according to hydrodynamic theory.

The formation of a snow crystal is best classified as taking place in two stages: (1) the formation of the germ or initial stage of the crystal, and (2) its subsequent growth into a snow crystal proper. The snow crystal proper is that which we observe on the ground, being several millimeters in dimension. The many varieties in the shape of a crystal are usually discussed with respect to the snow crystal proper, although some varieties are also observable among the germs. The latter are usually very tiny, being a few hundredths of a millimeter in dimension. The well-known experiments on cloud modification by I. Langmuir and V. J. Schaefer are those of transforming the supercooled cloud droplets into the germs. The nature of germs and the conditions for their formation will be found in the article by Schaefer.[2] In this article the description is confined to the snow crystal proper, which henceforth is simply called the snow crystal.

Dobrowolski's book [5] is the most comprehensive study of snow crystals thus far published. The book by Bentley and Humphreys [3] is famous for the vast collection of over three thousand photomicrographs of snow crystals. The crystal appears quite different if the mode of illumination is changed. Transmitted light is usually used. Photography using transmitted light is advantageous in obtaining a clear picture of the boundary and internal structure of the crystal. However, ordinary transmitted light does not show clearly the topography of the surface. An illumination by reflected light increases the beauty of the photograph, outlining the white image clearly against the dark background, but the delicate structure inside the crystal is not revealed. Recently M. Hanajima improved the technique of photomicrography to a remarkable extent by using oblique illumination. He succeeded in taking photomicrographs showing both the internal structure and the slight ruggedness of the crystal surface. The method is shown in Fig. 6.

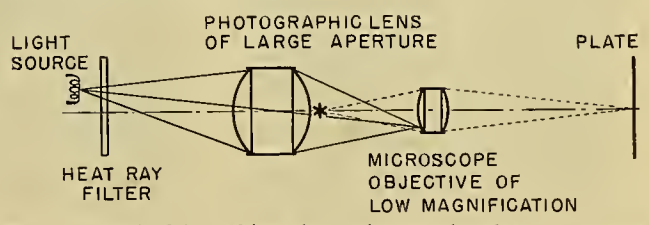


Fig. 6.—Method for taking photomicrographs of snow crystals (*after M. Hanajima*).

Snow crystals of various types photographed by this method of illumination are illustrated in Figs. 7–14.

CLASSIFICATION OF SNOW CRYSTALS

Principles of Classification of Snow Crystals. The famous astronomer Kepler was reputedly the first to point out the hexagonal symmetry of snow crystals. Descartes [4, p. 298] left the first scientific record of snow crystals. He made observations of snow crystals in 1635 at Amsterdam. Hooke, the discoverer of plant cells, gave the first sketches of snow and frost crystals observed through a microscope in his *Micrographia*. Hellmann [9, p. 37] in Berlin and Nordenskjöld [12] in Stockholm independently classified snow crystals into three kinds: planar, columnar, and a combination of the two. The basic idea of this system is retained in the general classification of snow crystals today.

2. Consult "Snow and Its Relationship to Experimental Meteorology" by V. J. Schaefer, pp. 221–234 in this Compendium.

FIG. 7.—Part of a snow flake (×10).

FIG. 8.—Plate with dendritic extensions (×23).

FIG. 9.—Plate with simple extensions (×28).

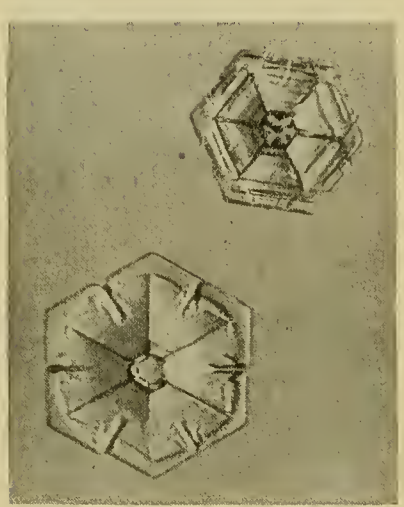


FIG. 10.—Plate and sector form (×28).

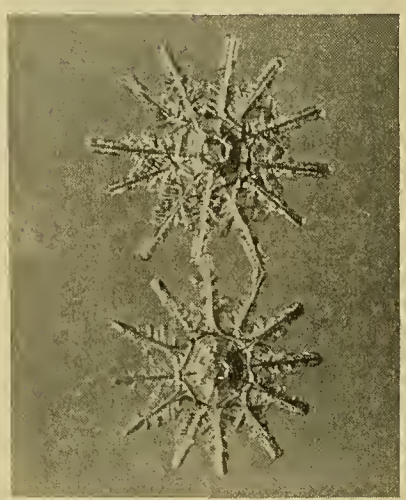


FIG. 11.—Crystal with twelve branches (×15).

FIG. 12.—Combination of bullets (×23).

FIG. 13.—Capped column (×10).

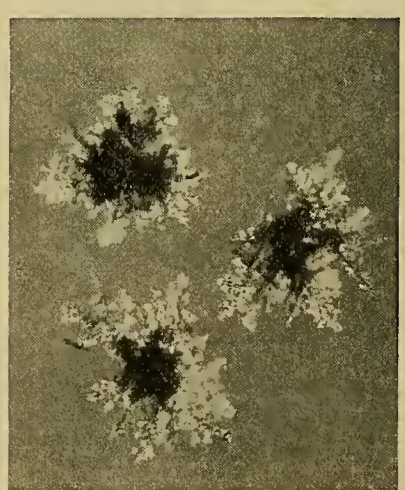


FIG. 14.—Graupel-like snow of radiating type (×9).

TABLE I. GENERAL CLASSIFICATION OF SNOW CRYSTALS

I	*M*	Needle crystal	*1.* Simple needle	*a.* Elementary needle *b.* Bundle of elementary needles
			2. Combination	
II	*C*	Columnar crystal	*1.* Simple column	*a.* Pyramid *b.* Bullet type *c.* Hexagonal column
			2. Combination	*a.* Combination of bullets *b.* Combination of columns
III	*P*	Planar crystal	*1.* Regular crystal developed in one plane	*a.* Simple plate *b.* Branches in sector form *c.* Plate with simple extensions *d.* Broad branches *e.* Simple stellar form *f.* Ordinary dendritic form *g.* Fern-like crystal *h.* Stellar with plates at ends *i.* Plate with dendritic extensions
			2. Crystal with irregular number of branches	*a.* Three-branched crystal *b.* Four-branched crystal *c.* Others
			3. Crystal with twelve branches	*a.* Fern-like *b.* Broad branches
			4. Malformed crystal	Many varieties
			5. Spatial assemblage of plane branches	*a.* Spatial hexagonal type *b.* Radiating type
IV	*CP*	Combination of columnar and planar crystals	*1.* Capped column	*a.* Column with plates *b.* Column with dendritic crystals *c.* Complicated capped column
			2. Bullets with plane crystals	*a.* Bullets with plates *b.* Bullets with dendritic crystals
			3. Irregular assemblage of columns and plates	
V	*S*	Columnar crystal with extended side planes		
VI	*R*	Rimed crystal	*1.* Rimed crystal	
			2. Thick plate	
			3. Graupel-like snow	*a.* Hexagonal type *b.* Lump type
			4. Graupel	*a.* Hexagonal graupel *b.* Lump graupel *c.* Cone-like graupel
VII	*I*	Irregular snow particle	*1.* Ice particle *2.* Rimed particle *3.* Miscellaneous	

The development of photomicrography gave rise to a tendency to attach particular importance to the regular hexagonal crystals of planar type, although in reality spatial and irregular types occur in greater quantity. Before the method of photomicrography was fully developed, crystals with spatial structure were often reported in the literature, but they have been more or less neglected by workers using photomicrography. In

this sense we could say after A. Wegener that the development of photomicrography hindered the study of snow crystals. The author, therefore, has always tried to attach equal importance to every type of crystal actually observed in nature.

General Classification for Scientific Purposes. The general classification presented in Table I is more or less similar to that of Hellmann and Nordenskjöld, but the special feature is the addition of new types and the various forms and structures are observed only slightly less rarely than regular hexagonal ones. They are caused by asymetrical growth of branches, malformation by attachment of nuclei, overlapping of several planes, V-shaped notch in the plate, etc.

5. Spatial hexagonal type, *P5a*. This is composed of a base crystal of dendritic form with many branches attached at various points of the base crystal and extending upwards.

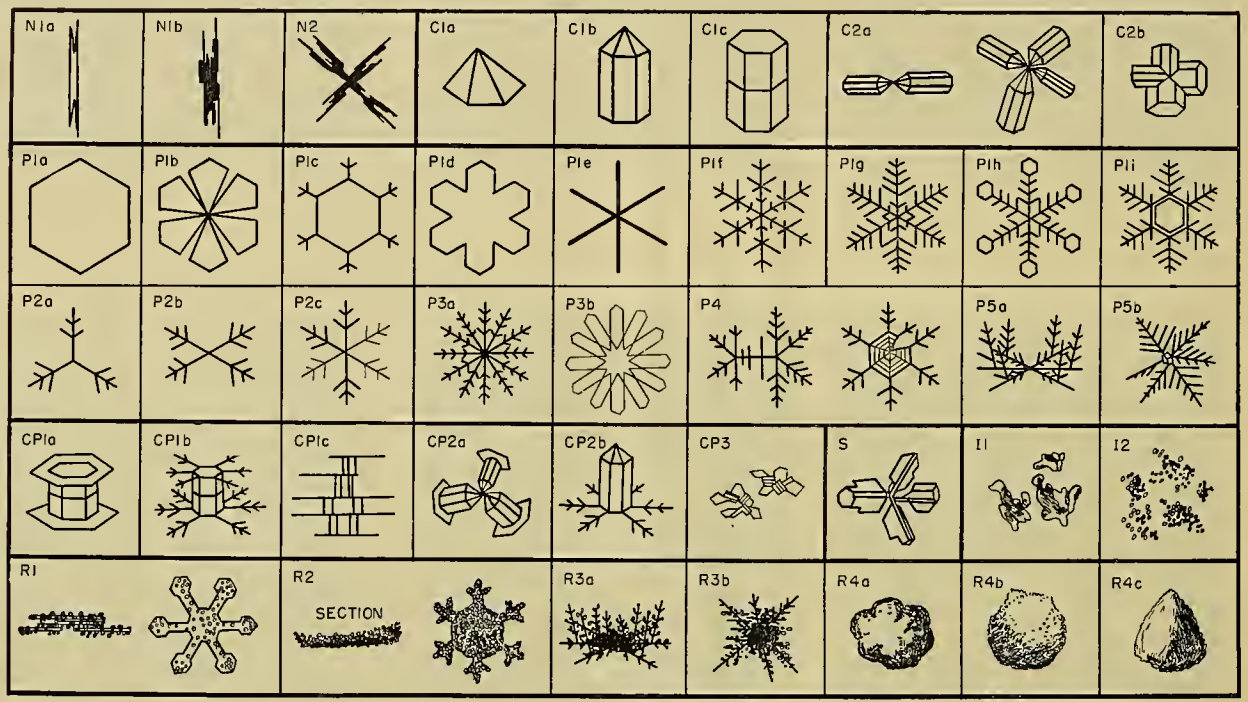


FIG. 15.—Sketches of all types of crystals in the general classification.

alteration in some items of the classification on the basis of crystalline structure. The graupel (snow pellet) is included, because it is an extreme case of the rimed crystal of spatial structure.

The sketches of all types of crystals in the general classification are shown in Fig. 15. A brief explanation will be given for some items in the general classification.

1. Needle crystals, *N*. Considering the results of the experiments on the artificial production of needle crystals, this should be taken as a kind distinct from the elongated column.

2. Crystals with an irregular number of branches, *P2*. Crystals which look as if they had developed from two nuclei or in parallel growth are observed in fairly large numbers. They have the appearance of a twin crystal. This twin crystal can easily be separated by a slight external force, giving the three-branched, four-branched, and other types. Seven pairs of components can be expected, as shown in Fig. 16. All of these component crystals have been actually observed in appreciable quantities.

3. Crystals with twelve branches, *P3*. This type is composed of two overlapping component crystals, and can be separated into ordinary hexagonal crystals.

4. Malformed crystals, *P4*. Malformed crystals of

6. Spatial assemblage of radiating type, *P5b*. This type has dendritic branches radiating in space from the center. The central part is a combination of small sectors or columns. The heavy snowfall in Japan is composed mostly of the spatial crystals, *P5a* and *P5b*.

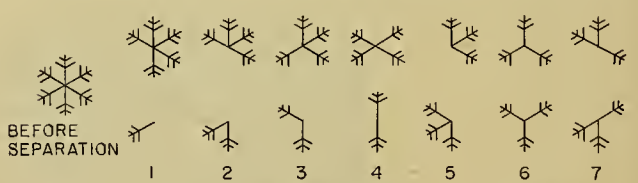


FIG. 16.—Seven modes of separation of a twin crystal.

7. Capped column, *CP1*. This is a crystal composed of a hexagonal column with plane crystals at both ends. We can find some examples of this crystal in the sketches by Descartes.

8. Irregular assemblage of columns and plates, *CP3*. The "flour snow" frequently observed in the Alpine regions is an agglomeration of minute crystals of this type.

9. Columnar crystals with extended side planes, *S*. The structure of this type has not yet been completely clarified, but experiments with artificial snow show that

the plane parts observable in this crystal are the extensions of the side planes of the columns, which form the skeleton of this crystal.

10. Rimed crystal, *R*. Supercooled cloud droplets are sometimes frozen to a snow crystal, and thus a rimed crystal is obtained. When many droplets become attached to a plane crystal, it turns into a thick plate (*R2*), sometimes half a millimeter in thickness. The graupel bearing a trace of hexagonal symmetry (*R4a*) is composed of a spatial hexagonal crystal *P5a* with numerous droplets attached, the graupel-like snow of hexagonal type *R3a* being the intermediate stage. The lump graupel *R4b* is made in a similar manner from the spatial assemblage of radiating type *P5b*, through the stage of graupel-like snow *R3b*. The cone-like form *R4c* is considered to be due to the rotational motion around the vertical axis during its fall.

11. Irregular snow particles, *I*. Snow particles are sometimes observed which do not show any regular crystalline form. There are many varieties: one type looks like an assemblage of pieces of ice (*I1*); another type has many water droplets attached (*I2*), etc.

Practical Classification of Snow Crystals. The general classification described in the preceding paragraph is not adequate for practical purposes. At the Oslo meeting of the International Commission on Snow and Ice in 1948, a committee was appointed to prepare a practical classification of snow, acceptable not only to scientists but also to others interested in snow. The aim was to promote uniformity in the method of describing snow and to simplify the correlation of data obtained by different groups. In 1949, this Committee on Snow Classification proposed a tentative snow classification covering solid precipitation and the deposited snow.

This tentative classification for solid precipitation is considered to be more convenient and adequate for practical purposes than similar ones so far proposed. It is shown in Table II.

In Table II, each class and basic feature of snow is designated by a code symbol with a view toward making the classification international (*i.e.*, independent of language). The description is made very simple. For example, *F1rD1.5* or ⬡ 1.5 means plate crystals with water droplets attached, the average size being 1.5 mm. *F2fwDd* or ❀ *d* means a cluster of stellar crystals partially melted, the average size of the crystals composing the flake being large, or between 2.0 and 3.9 mm.

Table II. Practical Classification of Solid Precipitation

	Code	Graphic symbol	Term	Remarks
Type of particle (*F*)	*1*	⬡	Plates	*P1a, P1b, P1c, P4.*
	2	✳	Stellar crystals	*P1d, P1e, P1f, P1g, P1h, P1i, P2a, P2b, P2c, P3a, P3b, P4.*
	3	▭	Columns	*C1a, C1b, C1c, C2a, C2b.*
	4		Needles	*N1a, N1b, N2.*
	5	⊛	Spatial dendrites	*P5a, P5b.*
	6	☰	Capped columns	*CP1a, CP1b, CP1c, CP2a, CP2b.*
	7	∽	Irregular crystals	*CP3, S, I1, I2, I3.*
	8	○	Graupel (snow pellet)	*R4a, R4b, R4c.*
	9	△	Sleet (ice pellet)	United States Weather Bureau definition; frozen raindrops, fairly small and transparent.
	0	▲	Hail	Solid precipitation formed by the successive freezing of water layers.
Additional characteristics	*m*	✳	Broken	Broken crystals of type *1*, *2*, etc.
	r	✳	Rimed	*R1*, *R2*, *R3*.
	f	❀	Flake	Clusters of crystals of type *1*, *2*, etc.
	w	✳	Wet	Wet or partially melted crystals of type *1*, *2*, etc.
Size of particle (*D*)	*a*	0 –0.49 mm	Very small	The size of particle means the greatest extension of a particle (or average when many are considered) in millimeters. For a cluster of crystals it refers to the average size of the crystals composing the flake.
	b	0.5–0.99	Small	
	c	1.0–1.99	Medium	
	d	2.0–3.99	Large	
	e	4.0 or larger	Very large	

ARTIFICIAL PRODUCTION OF SNOW CRYSTALS

Method for Making Snow Crystals in the Laboratory [11]. The artificial production of snow crystals means the production of frost crystals freely suspended in the air in a cold chamber laboratory. It takes at least from one-half to one hour for the complete development of a snow crystal. A thin filament was used to keep the crystal suspended in air for such a time.

In order to obtain nearly steady convection of water vapor, the apparatus shown in Fig. 17 was designed. Two concentric glass cylinders are held vertically, so that warm water vapor is driven upward inside the inner tube, while the cooled air comes down through the space between the two cylinders. The water in the

reservoir R is warmed electrically. The cover C is a metal sheet and D is a plate of wood. After examining five kinds of filaments, a thin rabbit hair was chosen because it was the most suitable filament to which a few

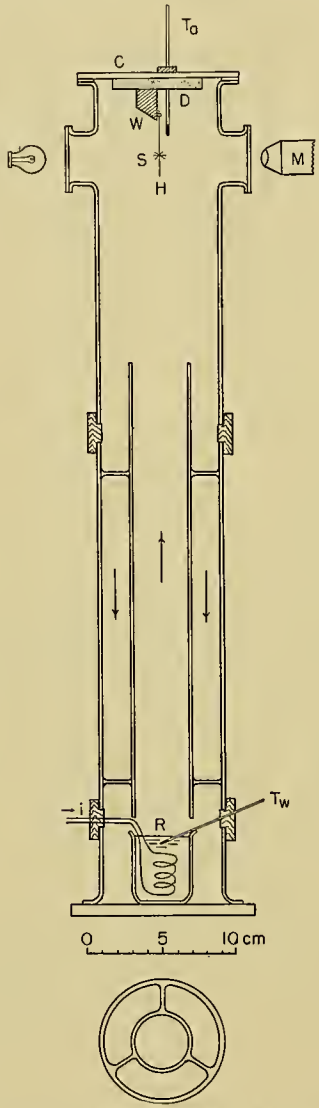


FIG. 17.—Apparatus for making artificial snow.

isolated germs of snow crystals could be attached. The structure of rabbit hair was examined under a microscope of high magnification, and it was found that a few knobs occur at suitable intervals. These knobs seem to serve as the nuclei for snow formation. The whole apparatus is set in a thermostat placed in the cold chamber which is kept below −30C by a refrigerating machine.

The temperature of the water in the reservoir (T_w) is measured. This is a measure of the degree of supersaturation and can be controlled. The temperature of the air (T_a) where the crystal is formed is a function of T_w and the temperature of the thermostat T_t. For a given T_w the air temperature T_a is regulated by changing T_t. Snow crystals have been produced for various combinations of T_a and T_w, and the relationship between the crystal form and the external conditions has been studied. It was found that T_a and T_w are the two elements controlling the form of the crystals.

The mode of air convection in the apparatus was studied by measuring the temperature distribution under operating conditions; one example is shown in Fig. 18. In this figure it is evident that the warm vapor

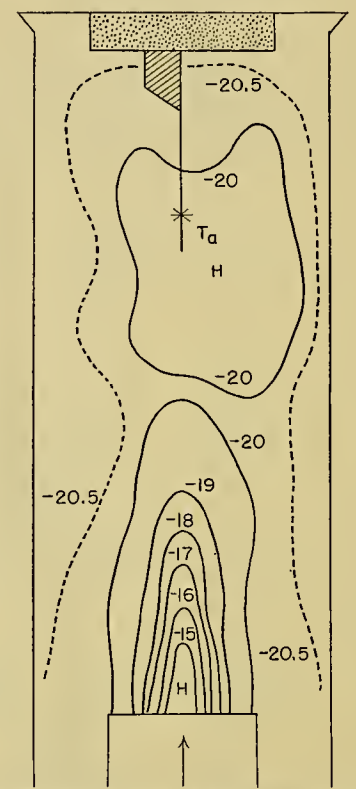


FIG. 18.—Convection of air in the apparatus (temperatures in degrees centigrade); $T_a = -23$C, $T_w = +15$C.

rises through the inner cylinder and is cooled gradually as it ascends from the outlet. The accumulation of warm air in the upper part of the apparatus is shown by the existence of a high temperature region H.

Three values of T_a, T_w, and T_t were recorded during the course of formation of a snow crystal. They were plotted as a function of time and the resulting graph was taken to represent the history of the conditions leading to the production of the crystal in its final form.

The Relation between the Crystal Form and the External Conditions. Inasmuch as T_a and T_w are the factors controlling the crystal form, a given type must be indicated by a point on the (T_a, T_w)-diagram. In fact, it was proved experimentally that a certain type of crystal occupies a certain region in the (T_a, T_w)-diagram. A crystal of complicated form which has been developed through successive stages of various conditions is represented by a succession of arrows in the (T_a, T_w)-diagram.

The relation between the crystal form and T_a and T_w has been examined for 700 crystals by Hanajima [7]. The results are plotted in the (T_a, T_w)-diagram shown in Fig. 19. Crystals are classified into eight types: den-

dritic, sector, plate, spatial assemblage of plates, cup or scroll, needle, irregular needle, and column. Region I in Fig. 19 is the condition for dendritic growth. It means that the necessary condition for dendritic development is that T_a is between −14C and −17C and the supersaturation is above a certain lower limit. Contrary to what has hitherto been believed, the air temperature is an important factor controlling the crystal form.

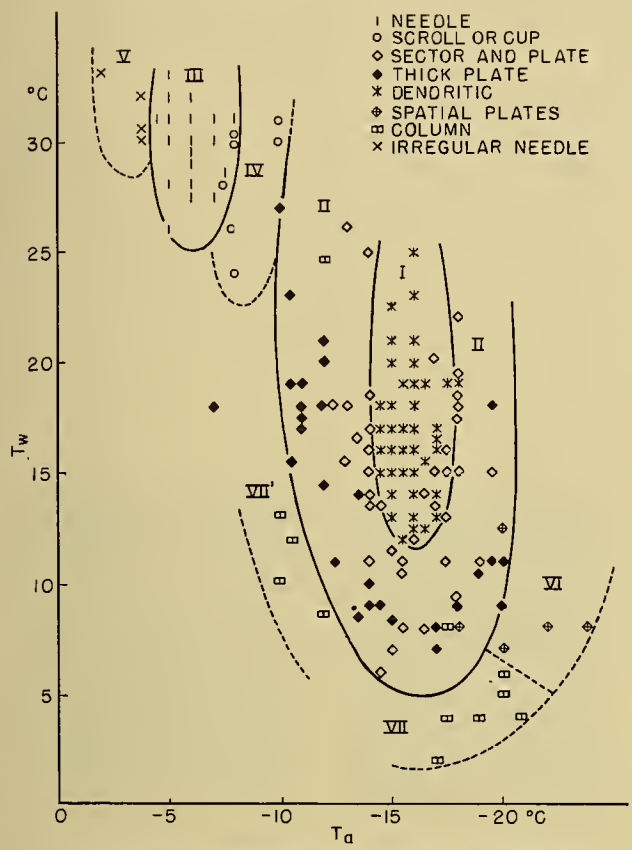


Fig. 19.—Relation between the crystal form and T_a and T_w.

The degree of supersaturation s is taken as the ratio of the amount of vapor and droplets per unit volume (measured) and the amount of saturated vapor at T_a per unit volume (calculated). In other words s is a relative humidity in the range of supersaturation, where humidity is taken to mean the total water content in the atmosphere. The ratio s was measured by a gravimetric method using P_2O_5, and Fig. 19 was transformed into Fig. 20, which represents the crystal form as a function of T_a and s [8].

Figure 20 tells us more clearly that the chief factor controlling the form of the crystal is the air temperature, and that a given type of crystal can be obtained for a wide range of supersaturation. It has generally been believed that the degree of supersaturation determines the form of the crystal, but Fig. 20 shows that this is not the case, except within a certain range of temperature. The transition of the plate form into the dendritic form occurs when the supersaturation exceeds about 110 per cent, provided the temperature lies between −14C and −17C. Above a supersaturation of 140 per cent the crystal is attached with numerous water droplets and becomes a rimed crystal.

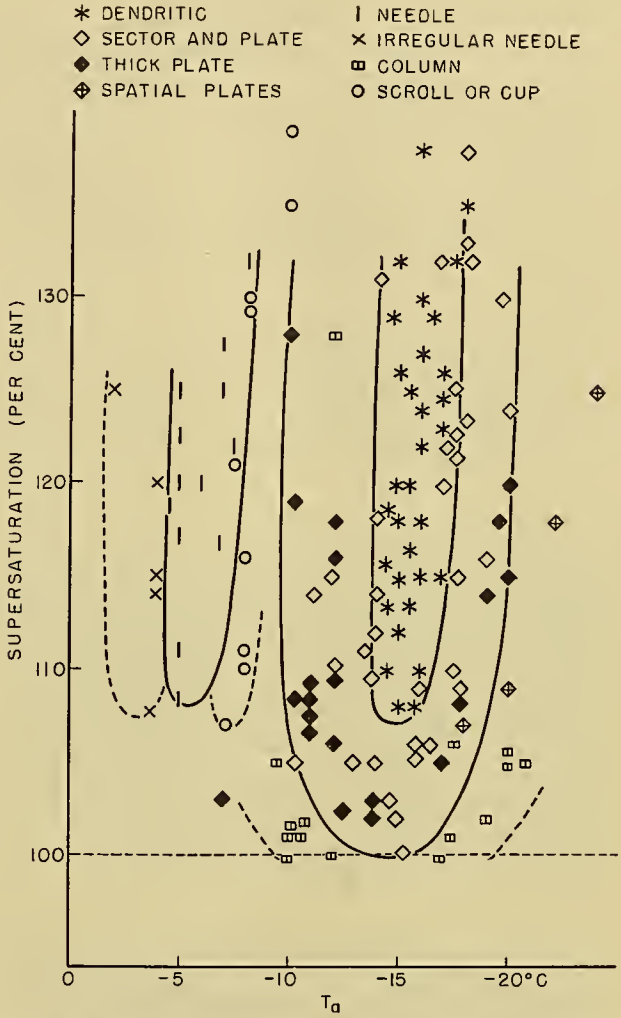


Fig. 20.—Relation between the crystal form and T_a and s.

The Growth of Snow Crystals. The acquisition of a detailed technique made it possible to produce all types of snow crystals almost at will. In the case of natural snow, the germ is first formed in the upper atmosphere as a result of the sublimation of water vapor on a nucleus or by the spontaneous transformation of a supercooled droplet. While falling through the atmospheric layers of various temperatures and degrees of supersaturation, this germ gradually grows into a snow crystal of complicated structure and eventually reaches the ground. In the case of artificial snow, this development can be observed in the course of time since the primitive crystal is recognized at a spot on the rabbit hair.

The course of formation is studied by taking photomicrographs of the growing crystal from outside the apparatus at regular time intervals while keeping the crystal in the apparatus. In order to maintain the crystal at rest against the convection current, the rabbit hair is stretched on a frame of thin glass rods.

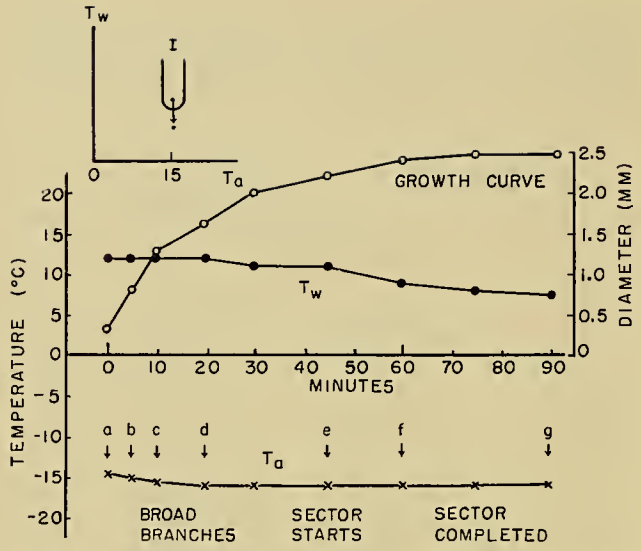


FIG. 21.—Condition of formation of the crystal in Fig. 22.

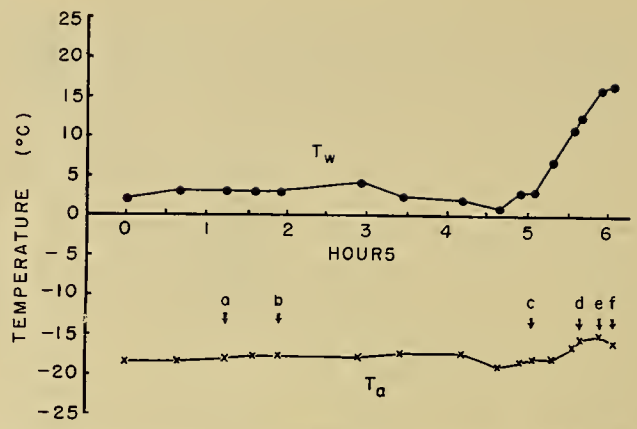


FIG. 23.—Condition of formation of the crystal in Fig. 24.

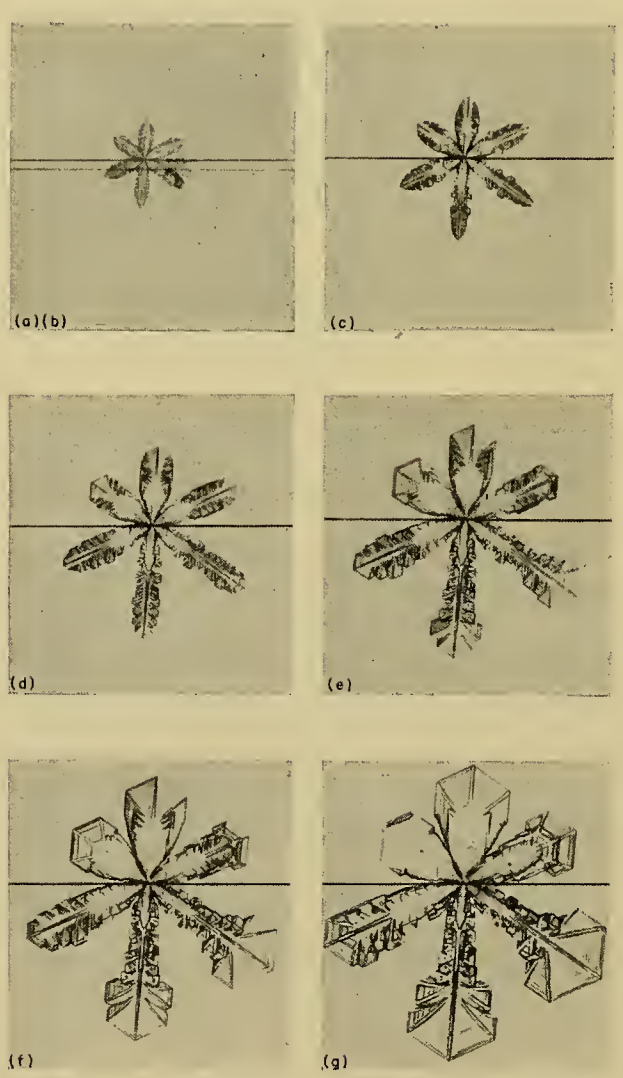


FIG. 22.—Course of formation of a stellar crystal with plates (×14).

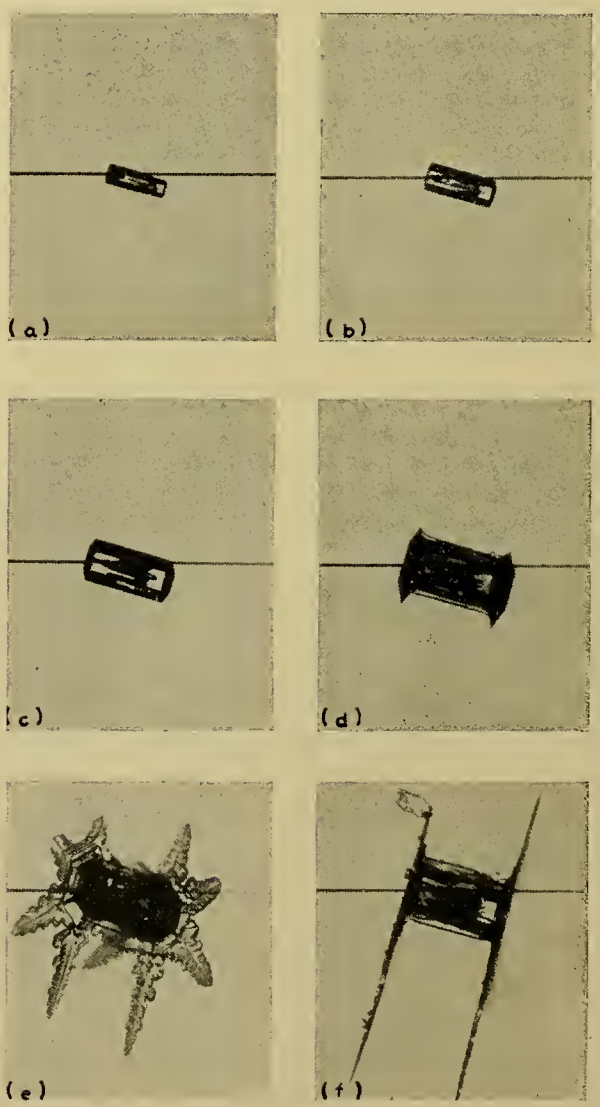


FIG. 24.—Successive stages in the formation of a capped column (×37).

Two examples showing the process of formation are illustrated, together with the condition of formation, in Figs. 21–24. Figure 22 shows the mode of formation of a stellar crystal with plates at the ends of the branches (*P1h* in the general classification). The conditions of formation are given in Fig. 21. The lettered arrows on the T_a curve show the moments at which the photomicrographs of Fig. 22 were taken. During the stages between *a* and *d* the air temperature T_a was kept at nearly −15C and T_w at about +12C; that is, the condition was within the region for dendritic development, although the degree of supersaturation was comparatively low. The crystal became a broad-branched type *P1d*, owing to the small amount of supersaturation. After this stage had been reached, T_w was gradually lowered, while T_a was kept at −16C. This change in condition made the rate of growth gradually smaller. The condition changed to that for plate development, as shown in Fig. 21. Accordingly, the tips of the branches began to widen (*e*), and finally the expected crystal was obtained as shown in *f*.

Figures 23 and 24 show the course of formation of a capped column. This crystal was obtained by first making a column under the condition of low temperature and humidity, and then changing the condition rapidly to that favorable for dendritic development. In the initial stage the column was slender in form. This column thickened without marked longitudinal elongation. In this example it took nearly five hours to attain the dimension observable in the central part of the natural capped column (Fig. 24*c*). Then the temperature of the water in the reservoir was raised rather quickly in order to increase the degree of supersaturation. In this condition a plate began to develop at each end of the column (*d*). The crystal finally developed into a completed, capped column as shown in *f*, through the stage *e*.

These two examples show that any crystal belonging to a certain type or a combination of several types can easily be produced artificially. The method is simple, the only thing to do is to establish the required conditions one after another.

APPLICATION OF SNOW CRYSTAL STUDIES TO METEOROLOGY

Upper-Air Conditions and Snow Crystal Forms. As described in the preceding paragraph, a snow-crystal germ is born in the upper atmosphere and develops to a snow crystal of complicated shape as it is subjected to different conditions while falling through the various layers of the atmosphere. The final form of the snow crystal observed at the earth's surface is an accumulation of the elements produced in the various strata.

Shedd [14], Wegener [15, p. 284], Findeisen [6], and others inferred the mechanism of snow crystal formation by comparing many photographs of snow crystals. Of these theories, Wegener's will be chosen as representative. According to Wegener, a primitive hexagonal crystal is first formed in a sufficiently supersaturated zone of the atmosphere and then the space between the branches is filled with ice while the crystal is falling through a less supersaturated layer, transforming the dendritic crystal into plate form. The next stratum of sufficient supersaturation will add other dendritic appendages. When the crystal comes down through a subjacent layer of less supersaturation, it becomes a larger plate. By the repetition of processes similar to these, many variations occur in the form of the snow crystals.

Wegener's theory was examined in an artificial snow experiment, and the transformation of a dendritic crystal into a plate in a less supersaturated atmosphere was found not to be the actual case, as described in the experiment of Fig. 22. However, the general idea of his theory is correct. Our experiments show that a complicated form of snow crystal is made by adding the characteristic features, corresponding to variations in the external conditions. The simplest example is a plate with dendritic extensions. It is produced when a plate is made in the upper atmosphere and the dendritic branches grow from the corners of the plate while the crystal is falling through lower layers suitable for dendritic development.

By examining the natural snow crystals from this point of view, we can infer to some extent the structure of the upper atmosphere, because the external conditions controlling the form of snow crystals are now made clear.

Comparison of Natural and Artificial Snow Crystals. By comparing natural and artificial snow crystals, the upper-air conditions can be estimated for each of various types of crystals, using the (T_a, T_w)-diagram. Four examples will be described.

Fern-Like Hexagonal Crystal. Natural and artificial crystals of fern-like type are reproduced in Fig. 25*a–b*. The form and structure of the branches are very similar in both cases, although the central part is somewhat different. The course of formation of this fern-like crystal is shown in Fig. 25*c*. During the primitive stage the temperature is lower and the supersaturation is less than the critical value for dendritic formation. Then the initial stage of this crystal is the irregular assemblage of small sectors. From this stage both T_a and T_w are increased so that the condition is most favorable for dendritic development. The crystal rapidly grows in a fern-like form and reaches the stage shown in Fig. 25*b* in about 15 min. The rate of growth is very large in this case. The rate of fall is about 1 km hr^{-1} for crystals of this type. Thus we can infer that there must be an atmospheric layer with ample moisture and a temperature of about −15C near the earth's surface, when the crystal of the type shown in Fig. 25*a* is observed at the earth's surface. The thickness of this layer will be about ¼ km.

Stellar Crystal with Plates at the Ends of the Branches. The crystal shown in Fig. 26*a* is frequently observed in nature, and is designated as *P1h* in the general classification. The artificially produced crystal shown in Fig. 26*b* certainly belongs in this group. The fine strips in the plate portion are less marked in the case of natural snow, which is probably due to sublimation in the

FIG. 25.—(*a*) Natural, and (*b*) artificial crystals of fern-like type; (*c*) physical conditions during course of formation.

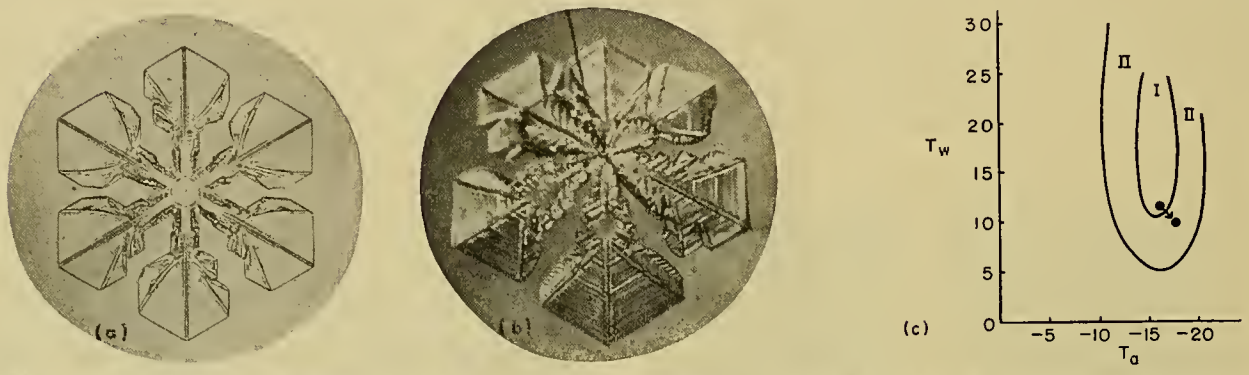


FIG. 26.—(*a*) Natural, and (*b*) artificial crystals of stellar type with plates; (*c*) physical conditions during course of formation.

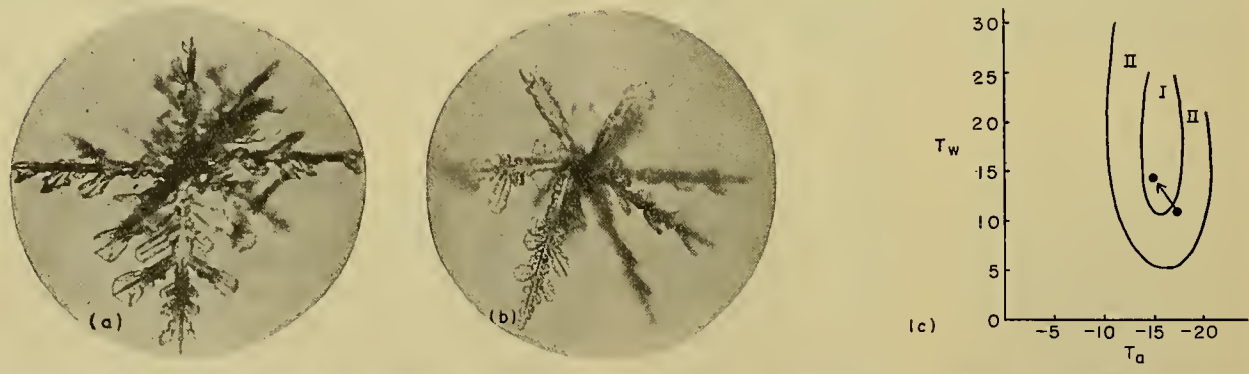


FIG. 27.—(*a*) Natural, and (*b*) artificial crystals of spatial radiating type; (*c*) physical conditions during course of formation.

FIG. 28.—(*a*) Natural, and (*b*) artificial crystals of dendritic type with small plates attached; (*c*) physical conditions during course of formation.

atmospheric layer near the earth's surface. The course of formation is shown in Fig. 26*c*. In the early stage the condition is favorable for stellar development, then both T_a and T_w decrease to just outside Region I, and the crystal is kept under this condition for a considerable time. In general, when a crystal is kept under a nearly constant condition at a point near the lower border of Region I, the ends of branches show the tendency to develop into plate or sector form. This can be explained as the result of a decrease in the vapor supplied per unit area to the crystal surface. The crystal of the form shown in Fig. 26*a* seems to show that a slightly supersaturated, homogeneous layer exists near the earth's surface with a thin layer suitable for dendritic development above it.

Spatial Assemblage of Radiating Type. The conditions favorable to formations of this type are also clear. The early stage is obtained at lower T_a and T_w. In the case of the crystal in Fig. 27*b*, T_a is approximately -20C and T_w is about $+12$C. After the formation of this primitive stage, T_a and T_w are increased to the condition of dendritic development, that is, $T_a = -16$C and $T_w = +15$C. Dendritic branches grow rapidly in space, giving the crystal of Fig. 27*b* in about twenty minutes. We may interpret the natural crystal of Fig. 27*a* as follows: In the upper atmosphere there exists a layer, at a temperature of -20C or lower, in which the crystal is born; this minute crystal falls through a layer characterized by ample moisture and a temperature of nearly -15C. The lower layer will be a few hundred meters in thickness.

Dendritic Crystal with Small Plates Attached. The natural snow reproduced in Fig. 28*a* is a good example of this type. Many small plates are attached in a spatial manner to the base crystal of hexagonal type. An artificial crystal similar to this is shown in Fig. 28*b*. This crystal is produced by a condition just the opposite of that for the preceding case. The base crystal is formed in Region I in about fifteen minutes. Then T_a is gradually decreased to -24C in two hours. The supply of water vapor is also reduced. In this second step many small plates extend out in space from various points of the base crystal. As the rate of growth of these small plates is very small, the development takes nearly two hours in this example. When this type of natural snow is observed, we may expect a thick layer of temperature inversion. The layer near the earth's surface must be at a temperature of nearly -20C and less supersaturated. The thickness of this layer is estimated from the data of falling velocity to be about 2 km, and above this cold layer there exists a warm, well-supersaturated layer at a temperature of nearly -15C. The warm layer may not be as thick as the cold layer, say, only several hundred meters.

The foregoing descriptions demonstrate the possibility of inferring the structure of the lower atmosphere by a synthesis of the examination of crystal forms and a knowledge of artificial snow. In this article we did not consider the question of wind. The horizontal component of wind velocity has no sensible influence upon our argument, but the vertical component and the turbulence will have a strong effect.

CONCLUDING REMARKS

In discussing the phenomenon of ice crystal formation in the atmosphere, the most important factor is the problem of supersaturation. As Bennett said in 1934 [2], "the evidence is merely negative as to whether supersaturation does or does not exist, and positive evidence is urgently required." This question is still left unanswered. It is very unlikely that more water vapor than the critical value of saturation exists in a purely gaseous state in the natural atmosphere. Strictly speaking, the existence of ample supersaturation of water vapor in air cannot be expected, except in some special cases such as at the instant of adiabatic expansion in a Wilson cloud chamber. Furthermore, in this special case the duration of the supersaturation is extremely short, say, 10^{-3} or 10^{-4} sec. Supersaturation observable in the natural atmosphere is considered to mean the existence of minute droplets in saturated air.

In our artificial snow experiment, the supersaturation was defined as the excessive water content in the atmosphere, including both water vapor and minute droplets. Values of supersaturation as high as 120 per cent or 130 per cent, referred to in this article, can thus be understood. We learned that in a rising air current which appeared to be transparent by ordinary illumination, a strong beam of light showed a Tyndall phenomenon. If a glass plate covered with oil film is exposed to the ascending air current for a short time, and then examined under a microscope of high magnification, a great many minute droplets can be observed in the oil film, many of them about 2 μ in diameter, the smaller ones about 1 μ. These minute droplets are not frozen to a snow crystal in the form of droplets, but appear to spread on the crystal surface at the moment when they are brought in contact with the crystal. In helping the growth of crystals, they act as if they were in a gaseous state. The larger droplets, 20–30 μ in diameter, behave in a manner quite different from that of the minute ones. They freeze to the snow crystal as droplets, and give rise to a rimed crystal. High values of supersaturation can be expected in the natural atmosphere, if by the supersaturation is meant an excessive water content, including water vapor and minute droplets of the order of 1–2 μ in diameter. This is not unreasonable, since in the process of condensation these minute droplets act as if they were in a gaseous state. Larger droplets such as ordinary fog particles do not behave in the same manner as do these minute droplets.

Another point is the relation between the air temperature and the crystal form. The apparently new result described in this article may be interpreted as follows: T_a is a factor controlling the rate of heat transfer liberated by the formation of the crystal and this rate of heat transfer determines the form of the crystal. Another explanation is that the vapor pressure difference between ice and supercooled water at the same temperature is a function of temperature, and that this

vapor pressure difference determines the mode of crystal development, that is, the crystal form. Further studies along this line, which is an attempt to introduce the methods of experimental physics into meteorology, may contribute something to this new field of meteorology.

REFERENCES

1. Altberg, V. J., *Twenty Years of Work in the Domain of Underwater Ice Formation (1915–1935)*. Un. géod. géophys. int., Assoc. int. hydro. sci., Bull. 23, 16e assemblée gén. à Edimbourg, septembre 1936. Trans. of the Meeting of the Int. Commissions of Snow and of Glaciers, Riga, 1938. (See pp. 373–407)
2. Bennett, M. G., "The Condensation of Water in the Atmosphere" in *Some Problems of Modern Meteorology*, D. Brunt, ed., pp. 114–125. London, The Royal Meteorological Society, 1934.
3. Bentley, W. A., and Humphreys, W. J., *Snow Crystals*, 1st ed. New York, McGraw, 1931.
4. Descartes, R., *Œuvres*, Tome VI. Paris, L. Cerf, 1902.
5. Dobrowolski, A. B., *Historja Naturalna Lodu*. Warzawa, J. Mianowskiego, 1923.
6. Findeisen, W., "Flugmeteorologische Schneebeobachtungen." *Meteor. Z.*, 56:429–435 (1939).
7. Hanajima, M., "On the Conditions for the Formation of Snow Crystals" (in Japanese). *Kisho shushi* (Magazine of the Meteorological Society of Japan), 20:238–251 (1942).
8. —— "Supplementary to 'On the Conditions for the Formation of Snow Crystals'" (in Japanese). *Kisho shushi* (Magazine of the Meteorological Society of Japan), 22:123–127 (1944).
9. Hellmann, G., *Schneekrystalle*. Berlin, R. Mückenberger, 1893.
10. McConnel, J. C., "The Crystallization of Lake Ice." *Nature*, 39:367 (1889).
11. Nakaya, U., "Artificial Snow." *Quart. J. R. meteor. Soc.*, 64:619–624 (1938).
12. Nordenskjöld, G., "Schneeflocken-Formen." *Meteor. Z.*, 11:346 (1894).
13. Seligman, G., *Snow Structure and Ski Fields*. London, Macmillan, 1930.
14. Shedd, J. C., "The Evolution of the Snow Crystal." *Mon. Wea. Rev. Wash.*, 47:691–694 (1919).
15. Wegener, A., *Thermodynamik der Atmosphäre*, 3. Aufl. Leipzig, J. A. Barth, 1928.

SNOW AND ITS RELATIONSHIP TO EXPERIMENTAL METEOROLOGY

By VINCENT J. SCHAEFER

General Electric Research Laboratory, Schenectady, New York

Snow in its many forms has been the subject of observation, conjecture, and scientific discussion for many centuries. It has long been recognized that a better understanding of the formation of snow in the atmosphere would eventually explain some of the little-known but important meteorological processes related to the development of precipitation.

The occurrence of supercooled clouds in the free atmosphere is one of the most common of meteorological phenomena, even in many parts of the tropics. The importance of such clouds as the source of much heavy precipitation has been pointed out by Bergeron [2]. The differential in vapor pressure between water and ice at all temperatures below 0C permits the rapid growth of snow particles at the expense of the liquid cloud droplets. This process is of basic importance in the formation of snow and is a primary mechanism in the science of experimental meteorology.

Types of Solid Precipitation

Over the years, many attempts [25] have been made to devise a classification system for describing the observed forms of solid precipitation. Most of these have been either too elaborate for easy use or have failed to include important forms.

During the course of snowstorm studies in 1944, an effort was made to devise a simple system which might be used in the field under adverse weather conditions. Revisions of this system were made as extensive field experience demonstrated the need. In the fall of 1949, an effort was made to pool the experience of workers concerned with this problem in Switzerland, Japan, Canada, and the United States. The chart shown in Fig. 1 illustrates the classification decided upon and the code and types proposed for international use. Although subject to further revision, it is believed that the types shown on this chart include most of the basic forms which occur in the atmosphere throughout the world.

As may be expected, there is an almost infinite variation in the forms of the basic types of this solid precipitation. These differences may be so minor as to be visible only under high-power magnification or great enough to be easily seen by the unaided eye. Typical variations in structure and relative size of the plate-type crystal are illustrated in Fig. 2.

Nakaya [16], in his ice-crystal experiments, showed that the crystal habit of snow may be due entirely to the temperature of the environment and the moisture supply available. It is quite likely that most crystals in the free atmosphere grow as they do because of these environmental conditions.

It should be pointed out, however, that the habit of crystals may also be modified by an entirely different mechanism—the blocking of the growth on certain crystal faces by the adsorption of surface-active chemicals [27]. Figure 3 illustrates the effects which may be induced by traces of an impurity in the air where crystals grow. Further research to understand these effects better is under way in the General Electric Research Laboratory.

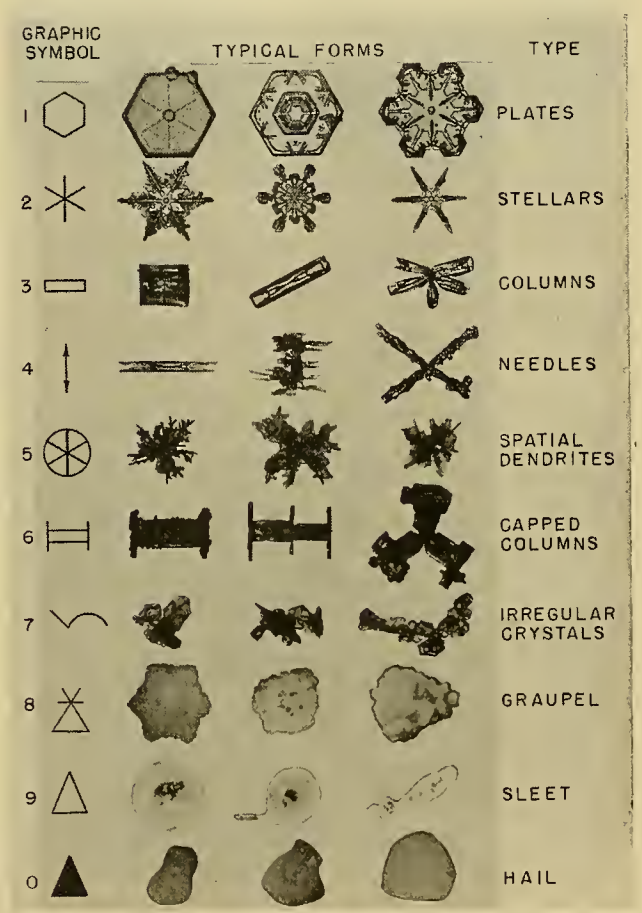


Fig. 1.—Types of solid precipitation.

The Use of Replica Techniques for Studying Snow. In 1941 a method was devised by the writer [18, 19] for making permanent replicas of snow crystals. The technique encases a snow or frost crystal within a thin plastic film which, as it forms, makes an exact three-dimensional impression of the surface features of the crystal. The replica solution, consisting of one to three parts of the synthetic plastic polyvinyl formal dissolved in 100 parts of ethylene dichloride, readily wets an ice surface. By capillarity and surface activity it rapidly covers any ice crystal which comes in contact with it.

The solvent evaporates in five to ten minutes, after which the slide (glass, cardboard, etc.) bearing the samples may be warmed above freezing. Upon melting, the water molecules evaporate through the thin film, leaving a hollow shell which refracts and scatters light

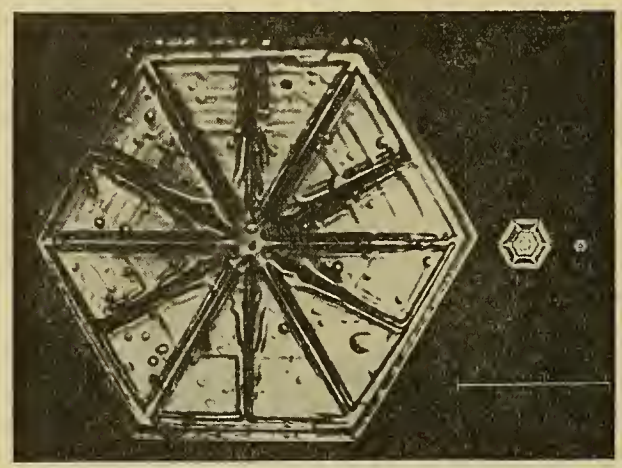


FIG. 2.—Variation in size and structure of hexagonal plates (*a*) from low and middle type clouds, (*b*) from high cirrus type clouds, and (*c*) from very low altitudes, forming in clear air.

in a manner quite similar to the optical properties of the original crystals. Figure 4 shows a typical replica of a stellar crystal.

This technique is also very useful in making surveys of snowstorms since it permits the accumulation during the course of a storm of many samples of crystals for

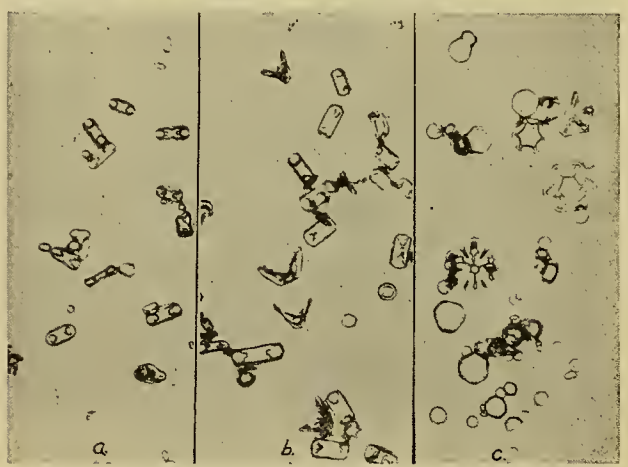


FIG. 3.—The effect of acetone vapor on the crystal habit of snow. (*a*) Effect of 1 molecule of acetone to 10 molecules of water. (*b*) Effect of 1 molecule of acetone to 100 molecules of water. (*c*) Effect of 1 molecule of acetone to 1000 molecules of water.

subsequent study. Samples have now been obtained by this method in most parts of the world as well as at high altitudes in the atmosphere during flight studies with Project Cirrus airplanes. Figure 5 shows some of these replicas.

A more recent technique for making replicas utilizes a plastic spray.[1] Although the solvent used in this spray would not work under normal conditions, since it would tend to dissolve the ice structure, its evaporation is so rapid that satisfactory replicas are obtainable if brief applications are made. This technique works best if the spray container is cooled below 0C. However, if the spraying is carried out in air at temperatures below freezing, enough entrainment of cold air takes place so that good replicas have been made at −10C with the dispenser temperature at 25C.

1. Such as Krylon, made by Foster & Kester, Philadelphia, or Plastic Spray, made by the Bridgeport Brass Co., Bridgeport, Conn.

Solid Precipitation in the Free Atmosphere

Precipitation in the form of snow crystals, graupel, sleet, or hail forms under varied conditions of temperature, humidity, and turbulence, and in the presence of a variety of suitable nuclei to be described later. The

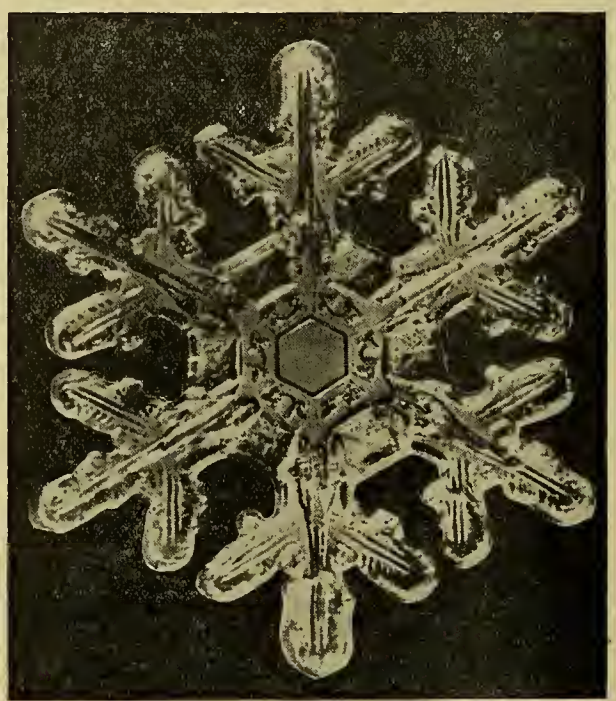


FIG. 4.—Typical replica of a stellar snow crystal.

moisture content of the air in which such precipitation may form at temperatures below 0C range from more than 3 g m^{-3} in supercooled water-droplet clouds to such small amounts that the air contains no visible cloud, although it is supersaturated with respect to ice.

Ice Crystals in Cirrus Clouds. The highest clouds commonly found throughout the world are the cirrus types. Evidence is accumulating suggesting that most clouds of this type form at temperatures below −39C. At this temperature, spontaneous nucleation occurs, that is, foreign particles are not required to initiate the formation of ice crystals.

The simple 22° halo and the more complex optical phenomena of cirrus clouds are generally produced by snow crystals of special form floating with a particular orientation in the air. Since the initial formation of such crystals may take place in clear air having a very low total water content, it is obvious that they must be very small and their growth quite slow. The crystal types common to these high-level clouds are the hex-

agonal plate and column and the irregular or asymmetric crystal.

The "false cirrus" streaming from the tops of very high cumulonimbus clouds, the condensation trails of

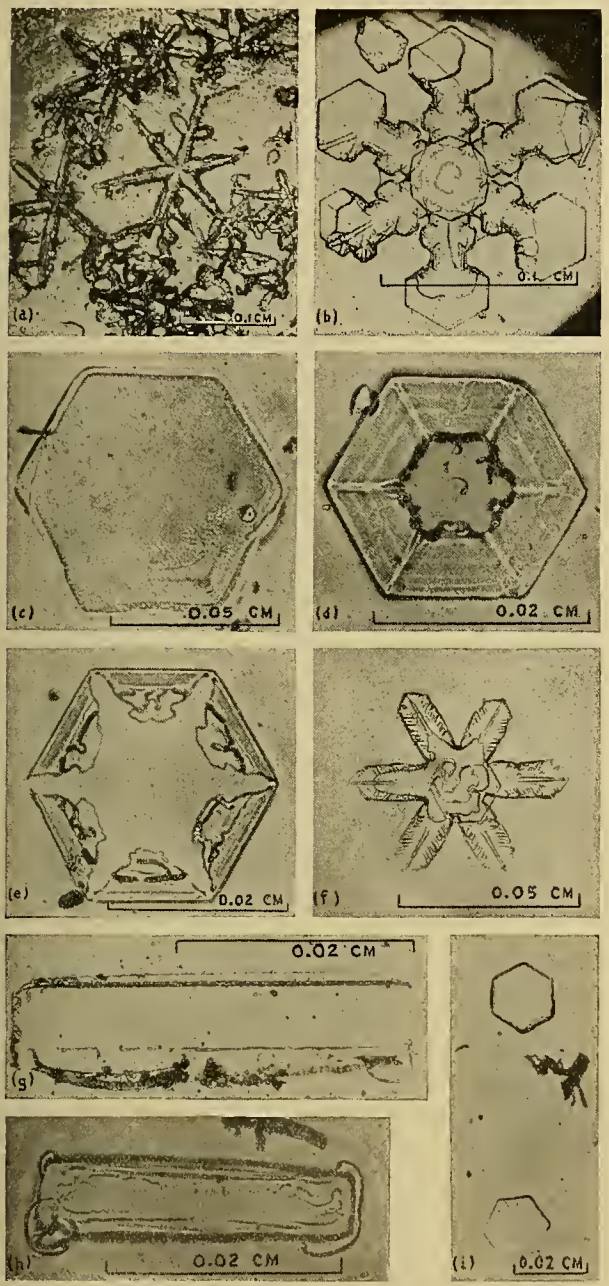


FIG. 5.—Replicas of snow crystals from seeded clouds. (*a*) Stellar forms from cloud seeded with silver iodide. (*b*) Stellar crystal from overseeded cloud after 51 min. (*c*) Plate formed in supersaturated clear air after seeding with dry ice. (*d*) Plate formed in 6 min after dry-ice seeding. (*e*) Plate formed in 10 min after dry-ice seeding. (*f*) Stellar crystal in overseeded area 45 min after dry-ice seeding. (*g*) Hexagonal column found in cirrus cloud at 27,000 ft. (*h*) Hexagonal column found in cirrus cloud at 28,000 ft. (*i*) Plates found in stratus cloud 15 sec after dry-ice seeding.

ice crystals from airplanes, and the ice fogs occurring in polar regions have features similar to cirrus clouds. Their formation is due, apparently, to the fact that the air in which they form is colder than −39C. In the case of condensation trails, this temperature occurs momentarily in the vortices streaming from the wings and propeller tips. When supercooled cloud droplets are present, they become frozen when contacted by the innumerable crystals spontaneously generated in their vicinity.

A snowstorm developing from cirrus clouds often requires from two to six hours for the crystals to reach the ground. A thin ice crystal haze producing a 22° halo thickens until the sun or moon finally appears as though covered by a ground-glass screen. Before reaching the ground, cirrus-type crystals may fall into lower supercooled clouds and become coated with cloud droplets. They sometimes serve as crystallization centers for the formation of stellars, spatial dendrites, or needles.

Solid Precipitation in Stratus Clouds. When air is stabilized by the presence of an inversion, the clouds which form contain considerably smaller quantities of liquid water than do cumulus clouds. Supercooled stratus clouds rarely have a liquid-water content higher than 1 g m^{-3}, while the average content ranges between 0.2 and 0.4 g m^{-3}. Ground fogs may be considered as a special type of stratus cloud, since they have similar properties.

Depending on the properties of supercooled clouds previously mentioned, the precipitation types from stratus clouds may vary from capped columns and stellar crystals to sleet. A deficiency of ice nuclei in stratus clouds is likely to produce a fairly heavy coating of rime on the crystals. When a layer of warmer air overruns a colder stratum, light rain may form in these upper clouds, becoming supercooled as it falls into the lower clouds and reaching the earth as rain or sleet. As sleet forms, an icy shell first coats the raindrop and freezing proceeds toward the center. As the last of the water freezes, the expansion causes the formation of many strange protuberances which modify the symmetry of the icy sphere.

Precipitation in Solid Form from Cumulus Clouds. High values of liquid water occur only in cumulus or thick orographic type clouds. Highest values of condensed cloud water occur when the vertical thickness of the cumulus exceeds 10,000 ft and the base of the clouds is in warm air, that is, warmer than 0C. Moist air forced upward by encounter with a mountain barrier and generally aided by convection induced by insolation often produces orographic clouds with features similar to cumulus. Such clouds may cover the upper slopes of a mountain or may have their bases considerably higher than the summit.

Many precipitation types form within the cold part of cumulus clouds. These include stellars, needles, spatial dendrites, and irregular crystals in addition to graupel and hail. The crystal forms may be expected when suitable ice nuclei occur in abundance, while graupel and hail types are products of a deficiency of such nuclei.

Factors Controlling the Formation of Snow

As suggested in previous sections, and demonstrated in the laboratory by the classic work of Nakaya [16],

solid precipitation types are a product of the variations in the physical nature of their environment. Let us consider the factors which initiate the formation of such particles in the atmosphere. At least five different mechanisms are of importance.

1. *Freezing Nuclei.* It has been shown experimentally [35] that bulk water may be cooled to at least −38.5C. Under carefully controlled laboratory conditions, which avoid seeding of the water by contact with frosted surfaces, water nearly always supercools to at least −5C to −10C [4]. In streams and ponds, however, there are always some freezing nuclei which initiate freezing in water at a temperature of only a few hundredths of a degree below 0C. However, the concentration of such particles seems to be relatively low, rarely exceeding 1×10^3 m^{-3}. The thin discs which form on such nuclei in flowing water are responsible for the formation of frazil ice [33].

Little evidence has been found in our studies to suggest that freezing nuclei are important in the formation of snow crystals in the atmosphere. The centers of snow crystals rarely show them to be formed on a frozen cloud droplet except when the crystals have apparently formed at very low temperatures. The size of cloud droplets, which on the average range between 8 and 25 μ, makes such observations quite feasible. However, the failure to find frozen droplets in the centers of snow crystals may be explained by some recent studies in my laboratory which show that cloud droplets condensed on silver iodide particles in warm air tend to assume the form of perfect hexagonal plates and columns as they freeze at a temperature of −4C.

2. *Sublimation Nuclei.* The most important mechanism responsible for the *initiation* of snow-crystal formation in the lower portion of the free atmosphere involves sublimation nuclei. These are very small air-borne particles upon which ice forms by the condensation of water molecules directly from the vapor phase.

Despite the claims of some workers [6, 40] who believe sublimation nuclei are extremely rare, experiments show that small particles of silicates and minerals common to desert and volcanic sources readily serve as active centers for snow-crystal formation. Figure 6 illustrates the temperature dependence of the effectiveness of different types of such particles when they are placed as an aerosol in cold air having sufficient moisture to be supersaturated with respect to ice.

Sublimation nuclei should not be confused with condensation nuclei. The latter are foreign particles, such as those produced by the evaporation of sea spray [9, 41], the burning of combustible waste, the particulate effluent from industrial processes, and the smoke of forest fires. Water molecules condense on such particles to form liquid droplets. Concentrations of condensation nuclei range from 2×10^8 salt particles per cubic meter in the air over the sea to concentrations of 1×10^{12} particles per cubic meter in the vicinity of industrial regions [11]. While the relative concentration of condensation nuclei is important when related to the stability and persistence of liquid-water clouds, since it is an important factor in determining the particle size, no evidence is known which relates the ordinary condensation nuclei to the formation of ice crystals.

A three-year study of the concentration of sublimation nuclei in air passing over the summit of Mount

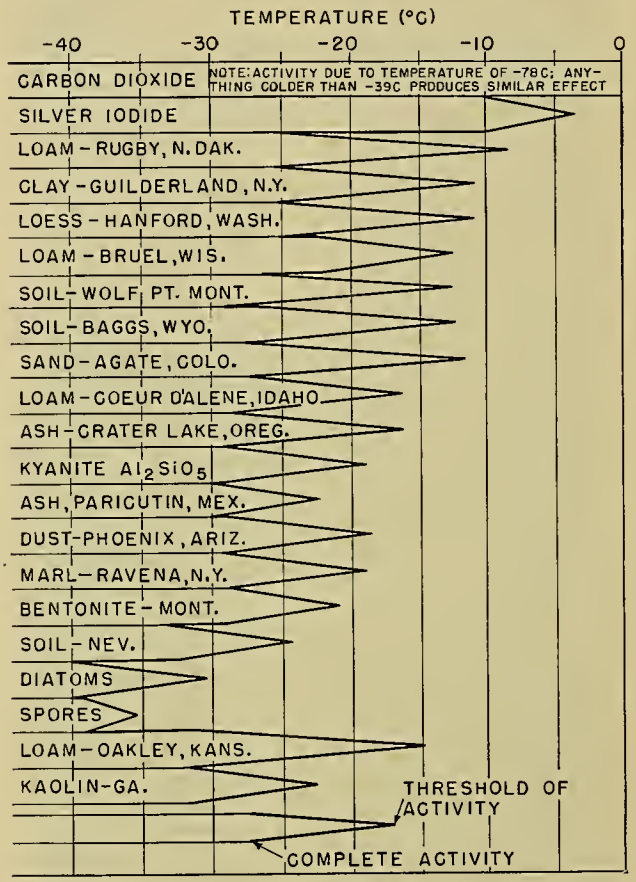


FIG. 6.—The temperature dependence of foreign-particle ice nuclei.

Washington (elevation 6288 ft) using the cold-chamber method [29] showed that the variation in concentration of such nuclei ranged from none observable to ten million per cubic meter [30]. Table I illustrates the range in concentration observed in the 8137 observations

TABLE I. RANGE IN CONCENTRATION OF ICE NUCLEI AT MOUNT WASHINGTON
JANUARY 1, 1948–JANUARY 1, 1950

Approximate number of nuclei per cubic meter of air	Number of observations
1×10^0–5×10^2	4062
5×10^2–5×10^5	3388
5×10^5–1×10^7	687

which were made up to January 1, 1951. Such studies are continuing at Mount Washington and are being supplemented by similar observations at Socorro, New Mexico. A study of the synoptic situation existing when the levels of nuclei were unusually high suggests that they might be dust particles from the Northwest and

the Great Plains which were carried aloft as the air passed over those regions.

The very important conclusion which may be inferred from the Mount Washington studies is that for extended periods the atmosphere contains very low concentrations of suitable nuclei for the formation of ice crystals. Thunderstorms, hail, heavy rain, and other storms which have part of their structure above the 0C isotherm are basically related in their developmental stage to the concentration of ice nuclei in the atmosphere where they form. It follows, therefore, that any modification in the natural concentration of such particles in the atmosphere should exert profound effects on the formation of such storms. This subject will be discussed in greater detail later under the section on experimental meteorology.

The most effective sublimation nucleus known at the present time is silver iodide. This fact was discovered during research activity in the General Electric Research Laboratory [37]. Since silver iodide does not normally occur in the free atmosphere, and since it is much more active than any known natural ice nucleus, far-reaching and anomalous effects may be expected to occur when large numbers of these particles are released in regions where supercooled clouds occur.

As Vonnegut [39] has shown, silver iodide is a very effective nucleus for ice formation because its atomic and crystalline characteristics, including the size of the unit cell, approach the structure of ice within 1 per cent.

While the temperature dependence of the effectiveness of sublimation nuclei is probably related to a number of such parameters, the nature of the surface layers of molecules seems to be of primary importance in determining whether a surface will be *kryophilic* or *kryophobic*.[2]

Typical effects illustrative of these properties are shown in Fig. 7. The kryophilic surface (Fig. 7*a*) consisting of clean glass, is coated with frost crystals whose major, or *c*-axis, is perpendicular to the glass surface. The kryophobic surface (Fig. 7*b*) is glass treated with a molecular layer of a dichlorosilane. Over the relatively small area contacted by frost crystals, the *c*-axis is parallel or inclined away from the treated surface at an acute angle. This surface is so strongly kryophobic that less than 1 per cent of the glass is contacted by the frost crystals.

It is my belief that such properties are inherently related to the effectiveness of, for example, a clay particle (active as an ice nucleus at −15C) and the spore of *Lycoperdon gemmatum* (which does not act as a crystal center until the temperature becomes −36C).

3. *Spontaneous Nuclei.* At high levels in the atmosphere where cirrus clouds form, or at lower levels when the temperature is below −39C, foreign particles are not required for the formation of ice crystals. Whenever moist air supersaturated with respect to ice is cooled below −39C tremendous numbers of ice crystals appear spontaneously. One of the easiest ways of demonstrating this effect is to use the cold-chamber method previously mentioned [29]. Tiny dry-ice fragments

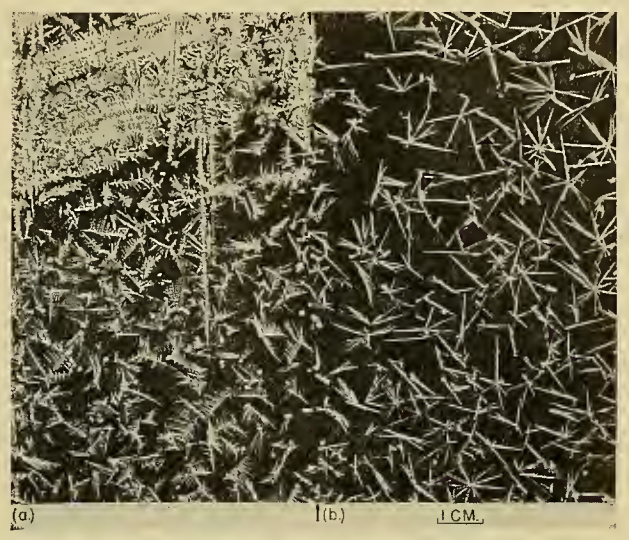


Fig. 7.—Frost crystal structure produced by surface property of a solid surface: (*a*) crystal forms on clean glass, and (*b*) crystal forms on glass coated with a very thin layer of a dichlorosilane.

dropped into cold air supersaturated with respect to ice produce many millions of ice crystals, as shown in Fig. 8. Photomicrographs of such crystals are shown in

Fig. 8.—Snow crystals forming in a cold chamber with air supersaturated with respect to ice.

Fig. 9. It can easily be shown that 1 g of dry ice may generate 10^{16} ice crystals under such conditions [26].

The critical temperature which produces this effect may be determined in a simple manner. By solidifying

2. These terms are suggested to designate the tendency of a surface to permit or to prevent, respectively, the formation of a frost layer.

a small drop of mercury with liquid air or dry ice and then passing it through saturated air at −15C, ice crystals are generated while the mercury is in the solid state. The instant it melts (melting point of *Hg* is −38.89C) ice crystals no longer are formed. Further

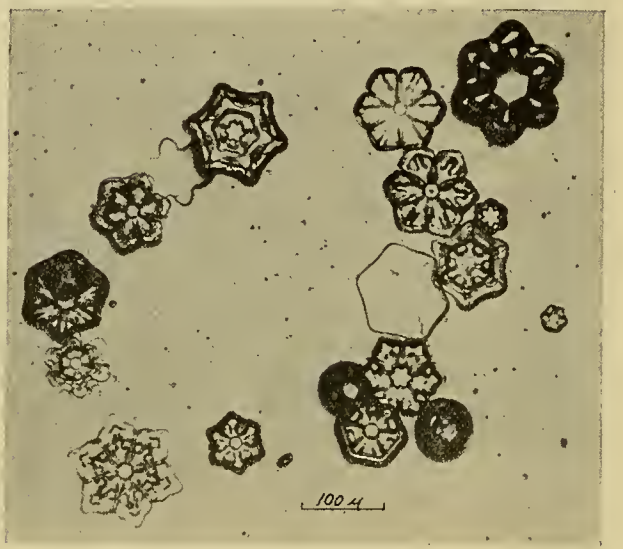


FIG. 9.—Typical hexagonal plates formed by dry-ice seeding of a supercooled cloud.

evidence that the critical transition temperature for the spontaneous nucleation of ice crystals is close to −38.9C may be demonstrated by using a sealed cold chamber with remote controls to add moisture to the air. As the air in the chamber is warmed and cooled between −37C and −41C, ice crystals suddenly appear at a temperature of −38.9C ± 0.1C. If the air remains colder than the critical temperature, snow forms and falls to the floor of the chamber continuously. This proves that foreign particles cannot be a factor in this process, since they would be quickly exhausted from the air by precipitation. A further interesting fact is that when the temperature passes from −38C to −39C many of the supercooled waterdrops present become frozen and serve as condensation centers for ice-crystal formation. Asymmetric crystals grow on such particles in such a manner as to suggest that the cirrus-type crystals shown in Fig. 10 probably grow on cloud droplets which freeze when they supercool to −39C and are seeded by several very small ice crystals spontaneously generated in their vicinity.

Another effect related to this critical transition temperature has been demonstrated by Vonnegut [38]. By suddenly expanding a cubic centimeter of air so that the rapid adiabatic expansion cools the air below −39C it is possible to form 10^{12} ice crystals within a small fraction of a second.

Since it is unusual for ordinary air to contain more than 1×10^6 cm^{-3} foreign particles of all kinds it is obvious that spontaneous nucleation occurs under these conditions.

4. *Fragmentation Nuclei.* A snow crystal formed by any of the preceding processes produces some interesting effects when in the presence of a supercooled cloud. Although the mechanism is not yet clearly understood, the effects may be produced experimentally [34]. They involve the formation of tiny ice particles in the immediate vicinity of the larger crystal, each of which subsequently grows to form new snow particles. It is believed at present that fragile dendritic growths in

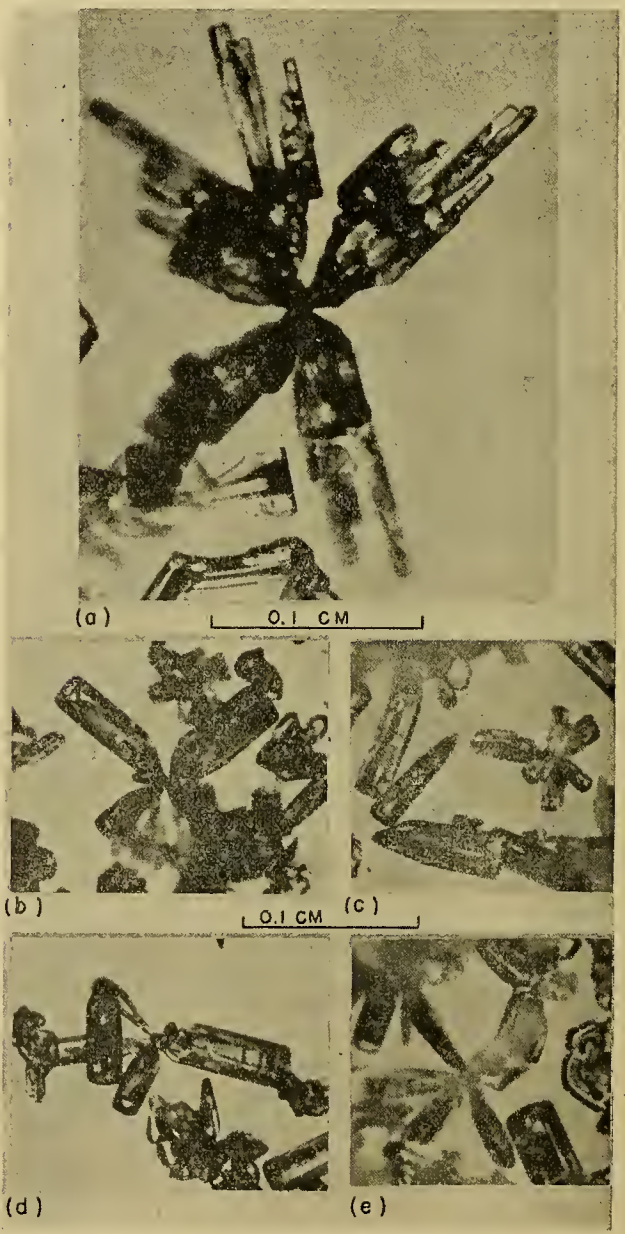


FIG. 10.—Compound columns from cirrus clouds probably growing on a frozen cloud droplet.

the form of fine needles, and frostlike structures of the type shown in Fig. 11 are required for the development of fragmentation nuclei.

Contact by liquid cloud droplets may produce thermal strains leading to the fracture of the delicate portions of such crystals. Collision with other particles could also lead to the formation of such fragments. In many snowstorms, especially those having convective activity, high winds, and supercooled clouds, broken

and irregular crystals (the latter probably formed on fragmentation nuclei) constitute the major portion of the snow reaching the ground. It is this chain reaction mechanism which must be responsible for most extensive snowstorms not seeded by high-level clouds.

5. *Electrification Ice Nuclei.* That some type of electrification may be an important producer of ice nuclei is suggested by laboratory experiments and field observation. If a field of several hundred volts per centimeter is established in a supercooled cloud, dendritic treelike growths may form if some ice crystals are present. These treelike forms seem to grow from a combination of crystals and supercooled cloud droplets. Subsequently they break and form many ice fragments, each of which becomes a potential ice crystal.

This effect has never been observed unless ice crystals were already present, and it may be associated with the Workman-Reynolds effect [42]. Although it has so far been impossible in the General Electric Research Laboratory to show any freezing effect or production of nuclei using either a-c or d-c fields up to 1 kv cm^{-1} in a supercooled cloud, a temporary coexistence of cloud droplets and ice crystals shows electrification phenomena which may be of great importance. These effects may be of primary importance in the establishment of the chain reaction which seems to be necessary to produce the many crystals required to form an extensive snowstorm. Certainly high fields are present in many snowstorms [22], particularly in those characterized by convective activity and supercooled clouds.

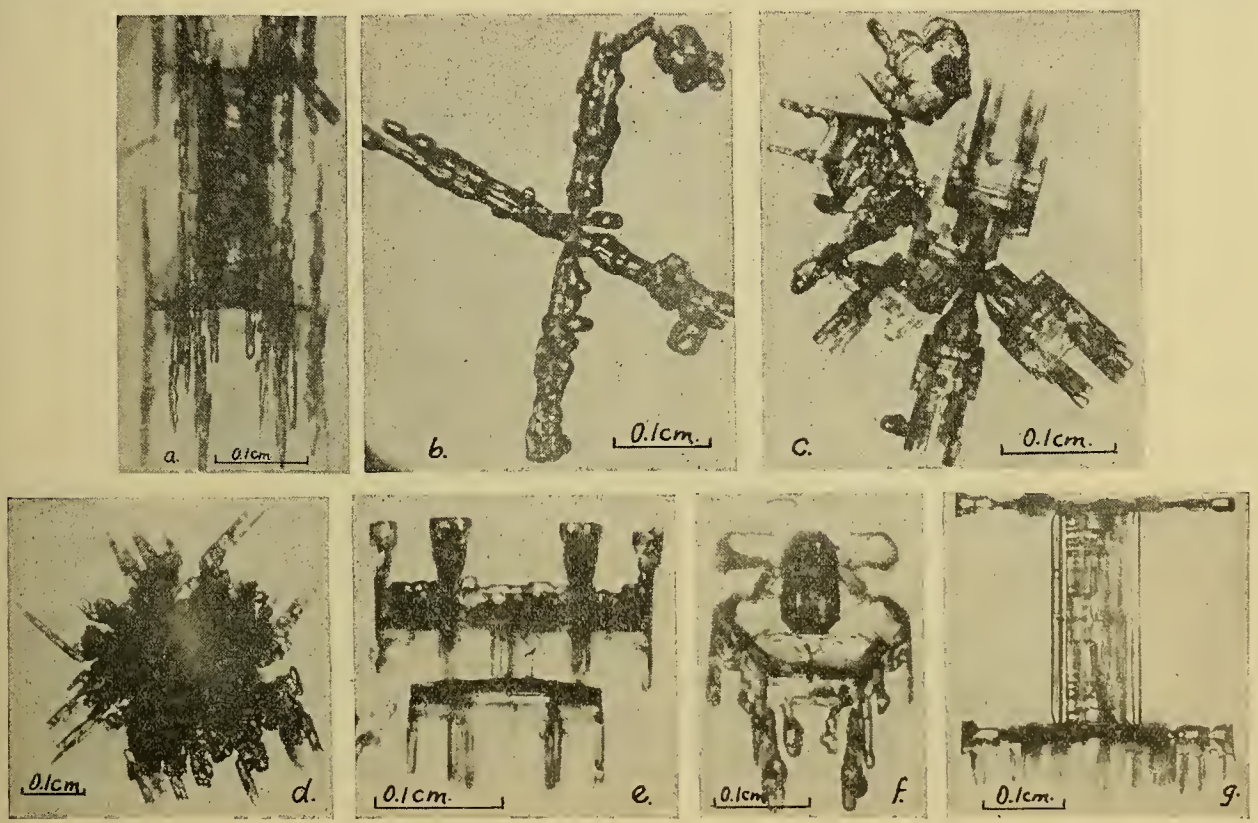


Fig. 11.—Types of snow crystals which are probably important in the formation of fragmentation nuclei.

Figure 12 illustrates the type of atmospheric electricity observed in snowstorms.

Electrical Properties of Snow

It is very likely that the chain reaction mechanisms suggested by the operation of fragmentation processes at temperatures from −12C to −20C and the interesting effects observed in electric fields are complementary mechanisms. Since snow crystals reaching the ground are often electrically charged, it seems obvious that they are indicators of mechanisms operating in the atmosphere where they were formed. The electrification may be a causative agent or an end result. The highest amounts of atmospheric electricity observed occur at temperatures not far below freezing and with broken stellars, needles, spatial dendrites, and sleet particles, all of which could be associated with fragmentation and electrification processes. Typical air-to-ground currents during a snowstorm are shown in Fig. 12.

The fact that the fragmentation of snow crystals causes a large increase in the charging potential of snow crystals shows the close interrelation which must exist between these two effects [22]. Further studies in this field should produce important results.

Factors Controlling the Life Cycle of Supercooled Clouds

The icing of aircraft, the formation of hail and graupel, and the electrical phenomena of thunderstorms are typical products of supercooled clouds. The growth,

persistence, or dissipation of such clouds is governed chiefly by evaporation and nucleation.

Evaporation may affect the life cycle of a supercooled cloud in two ways: (1) the cloud droplets may evaporate into air which is unsaturated with respect to water, or (2) they may evaporate and the water molecules condense on an ice nucleus. Either process is a very rapid one and for most purposes may be considered to be instantaneous. Nucleation under special conditions may cause a direct freezing of the supercooled cloud droplets when the concentration of ice nuclei is very high. It is unlikely that this latter process occurs very often in the atmosphere except at the spontaneous nucleation temperature of −39C. The tops of large cumulus clouds and the initial stages of some cirrus clouds show abundant evidence of this process. It is probably due to this mechanism that the optical properties of clouds at the −39C level remain unchanged. The freezing of cloud droplets does not alter the shape of the particle, and although high concentrations of small ice crystals probably form nearby they immediately disappear by evaporation since their vapor pressure is higher than that of the larger frozen cloud droplets. It is very likely that compound crystals like those in Fig. 10 form on cloud droplets frozen by this process.

When supercooled clouds are warmer than −39C and the air containing them is moist enough to prevent evaporation, several different processes control their persistence. If the cloud is of a stratiform type with few or no ice nuclei present, a freezing drizzle may develop and fall from the base of the cloud. If a few ice nuclei are present in the cloud, snow crystals form on them, grow rapidly, and fall out of the cloud as scattered snow particles. Initial growth develops from the vapor phase, but as soon as the crystals become sufficiently large to move faster than the surrounding supercooled cloud droplets, they sweep up cloud droplets which freeze as they touch the snow crystal. Graupel, hail, some forms of ice needles, and all other rimed crystals depend on supercooled cloud droplets for their formation. The presence of such crystals in the atmosphere is evidence of a low concentration of ice nuclei in the air mass and the absence of a chain reaction.

Most cumulus clouds supercool above the 0C isotherm and persist for unpredictable periods. If the moist air forming them carries with it certain types of air-borne dust which have the proper structure to serve as potential ice nuclei, they may rise only three or four thousand feet above the freezing level before shifting over to snow. At times, however, they may go nearly or all the way to the −39C level without enough snow crystals forming to initiate a chain reaction effect. It is this type of cloud which may develop into a storm producing lightning or hail, or both. At other times, the cloud may be completely dissipated into false cirrus at high altitudes without producing any precipitation.

Orographic clouds also have a strong tendency to become supercooled. Such places as the summit of Mount Washington have extensive icing storms during much of the year. Because of the very large vertical component of the wind caused by the mountain barrier, coexistence of supercooled cloud droplets and ice crystals is more likely than in most clouds in the free air, since condensation occurs faster than the evaporation-diffusion process can transport the moisture to the ice crystals. It is likely that orographic clouds in which

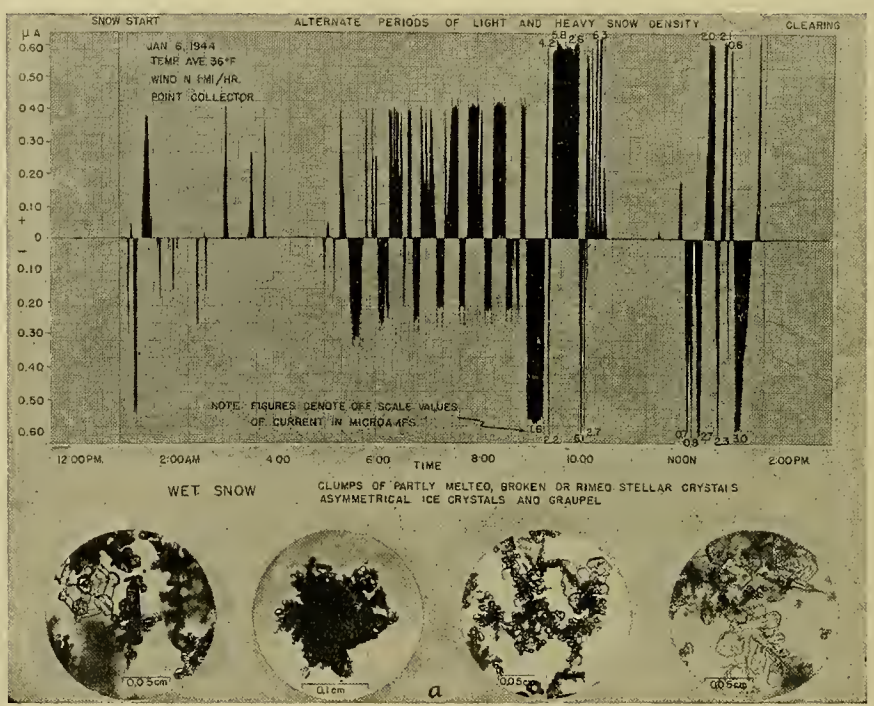


FIG. 12.—Typical air-to-ground electric currents which often occur during convective type snow storms.

the wind velocity is high have features in common with the more turbulent portions of large cumulus clouds.

The Development of Experimental Meteorology

Control of the Weather. For many years man has dreamed of the possibilities of exercising control over the weather. To a limited degree he has accomplished this by heating, air conditioning, and artificial illumination of homes, workshops, and places of culture and amusement, as well as by the use of flood-control techniques, irrigation systems, and similar technical improvements in limited areas of his countryside.

One of the first suggestions that scientific research would eventually lead to the modification of weather systems is contained in an important paper by Findeisen [5]. Others, such as Gathman [7], had previously proposed the use of dry ice to produce clouds and rain, a method subsequently tried by Veraart [36]. Their idea was to use dry ice to chill the air below its dew point to produce clouds and rain. Unfortunately, countless tons of dry ice would be required to produce even a small rainfall by this method. The insignificant results produced by this method and others, such as the formation of rain by dumping electrified sand into clouds (tried by Warren and Bancroft), led in 1926 to the conclusion (Humphreys [8]) that none of the proposed methods were of any consequence and that they had no scientific importance.

Recent Advances. In 1941 Langmuir [12] and Schaefer devised a method for producing an artificial fog which eventually was used to blanket sixty miles of the Rhine Valley as well as other large areas during the latter part of World War II. Subsequent research following the production of artificial fog involved basic studies of precipitation static on aircraft and of aircraft icing [20]. This latter study, which involved research with supercooled clouds, led the writer to discover a method of changing a supercooled cloud to ice crystals in a simple but highly effective manner [21].

When this discovery was made, it was immediately apparent that the time had arrived to attempt the modification of clouds on a large scale. Unlike all previous attempts, the methods proposed would utilize dry ice as a triggering device to release the energy stored within supercooled clouds by causing a change in phase of the unstable supercooled cloud. By utilizing the inherent instability of such clouds it was planned to introduce ice crystals in such numbers that large effects would be produced according to the Bergeron mechanism.

The first experiments designed to transform a supercooled cloud to snow crystals were carried out by Schaefer in the fall of 1946 [24] and the early months of 1947. In March 1947 cloud-seeding experiments using dry ice in supercooled clouds were undertaken on a larger scale with government support. This activity, called Project Cirrus, was sponsored by the Army Signal Corps and the Office of Naval Research, using planes and crews supplied by the Air Forces, with technical consultation provided by Langmuir and Schaefer of the General Electric Research Laboratory at Schenectady, New York.

The laboratory, field, and flight studies of this group have been described in more than a score of reports [15]. Flight studies up to July 1950 have included experimental investigations of clouds in the northeastern United States, New Mexico, Puerto Rico, and Florida. In addition, detailed observations of natural clouds have been undertaken in various parts of the country and many field and laboratory projects have been completed or are actively under way at the present time.

Cloud Modification Methods

The Use of Dry Ice to Modify Clouds. As mentioned earlier, whenever any object with a temperature less than −39C is introduced into air colder than 0C and supersaturated with respect to ice, tremendous numbers of ice crystals are generated. The simplest method for achieving this effect is to introduce small fragments of dry ice (solid carbon dioxide) into the air. This may be accomplished in the free atmosphere by using free balloons, projectiles, or airplanes, or by placing the dry ice or expanding liquid CO_2 in orographic clouds on mountains or in supercooled ground fogs.

Dry ice may be introduced into various parts of clouds with differing effects. The ice crystals it generates will immediately disappear unless the air is below freezing and supersaturated with respect to ice. Langmuir has summarized our general methods [14]. Briefly, the judicious use of dry ice may (1) dissipate clouds, (2) precipitate clouds, or (3) make existing clouds more persistent.

Clouds may be dissipated by seeding their tops from an airplane before they develop a deep supercooled layer or by seeding along a line from the side, using liquid carbon dioxide. It is much easier to dissipate clouds after they have passed their maximum vertical development than when they are actively growing, since seeding a growing cloud may trigger off a chain reaction.

Since it requires considerable experience and judgment to select a growing cumulus, experiments planned to cause precipitation may fail because of improper selection of the area to be seeded. If the larger of the towers in a cumulus system are selected for seeding while the airplane is in a position to see the general cloud system, the selected cloud will probably be in a subsiding state before seeding takes place. Under such conditions, dissipation is likely to result. The life cycle of such clouds is so short [43] that an intimate knowledge of their dynamic properties is essential if effective seeding is to be achieved.

The effects are generally spectacular when the seeding is done to release effectively the heat of sublimation stored in supercooled clouds. In unstable air, a tremendous upheaval may occur if the seeded cloud is actively growing and well supercooled. This increase in convection occurs because of the heat liberated by the shift from the water to the ice phase. In stratus clouds, this increase in temperature may amount to 0.5C, while in

large cumulus formations an increase in temperature of several degrees within the cloud is quite possible.

The Use of Silver Iodide as a Seeding Agent. An entirely different process for seeding the atmosphere with ice nuclei involves the use of submicroscopic silver iodide particles produced with a smoke generator. Since this substance is most effective at temperatures below −10C, techniques which are somewhat different from those used with dry ice are employed. Silver iodide lends itself particularly well to the use of ground generators and the modification of large cloud systems. By the use of the generators described by Vonnegut [39] it is possible to produce invisible smokes of silver iodide particles in such high concentrations as to infect many thousands of cubic miles of the atmosphere. Not only is it possible to exceed the highest concentrations found in the free atmosphere, but the silver iodide particle is a more effective ice-crystal nucleus than any which has thus far been found in nature.

The Seeding of Clouds with Water. Under certain conditions the precipitation cycle may be initiated in large cumulus clouds by introducing large water drops into actively convective portions of the cloud. This seeding initiates a chain reaction [13] involving the growth and breakup of water drops. Since this seeding mechanism is not related to snow, it will not be discussed here.

Cloud Seeding

Effects of Seeding Operations. Laboratory experiments show that a concentration of about 50 ice nuclei per cubic centimeter will exhaust all the moisture in a supercooled cloud in about 10 sec. Since supercooled clouds are sometimes invaded by ice crystals falling from cirrus clouds, it is interesting to determine their effect on a supercooled cloud. Such crystals have a falling velocity of about 1 m sec^{-1} and a concentration in the air of about 1×10^4 m^{-3}. Assuming typical atmospheric conditions with the invading crystals moving through the supercooled cloud at a rate of 1 m sec^{-1}, not more than 20 min would be required to use up the supercooled cloud droplets in a specific region of a cloud. This process may often be observed in the free atmosphere.

The Overseeding of Clouds. In most natural clouds the number of cloud droplets ranges from 100 to 500 cm^{-3}. This is about one hundred times more than the number of potential ice nuclei thus far observed under natural conditions. Supercooled droplets can be transformed to ice crystals without a decrease in the number of the individual particles when they reach the spontaneous nucleation temperature of −39C or when they are artificially seeded with dry ice or silver iodide. Concentrations of ice nuclei exceeding 10,000 cm^{-3} may be induced in natural supercooled clouds by proper seeding methods. This fact is most easily demonstrated in a supercooled ground fog [23]. Since one gram of dry ice may generate 1×10^{16} ice crystals, this minute quantity effectively distributed could fill 1×10^7 m^3 of air (a volume of fog 100 m thick, 100 m wide, and 1000 m long) with a concentration of 1000 cm^{-3}.

The optical effects produced by overseeding of clouds are very striking. Very beautiful halos, sun pillars, sun dogs, etc., of brilliant color and high intensity occur in overseeded clouds. Sometimes a peculiar bluish hue develops in an overseeded cloud because the particles are small with respect to the wave length of visible light.

As suggested by Bergeron [3] and Schaefer [31], overseeding may eventually be used to cause the transport of moisture from one region to another or may be utilized to form large reservoirs of ice nuclei. Such formations could affect clouds some distance from the seeding point or any clouds in the infected region which became large enough to form precipitation. It follows therefore that overseeding of supercooled clouds may become an effective method for preventing the formation of rain or snowstorms in specific areas.

Figure 13 illustrates the active formation of an ice-crystal cloud by dry-ice seeding in clear air. The early

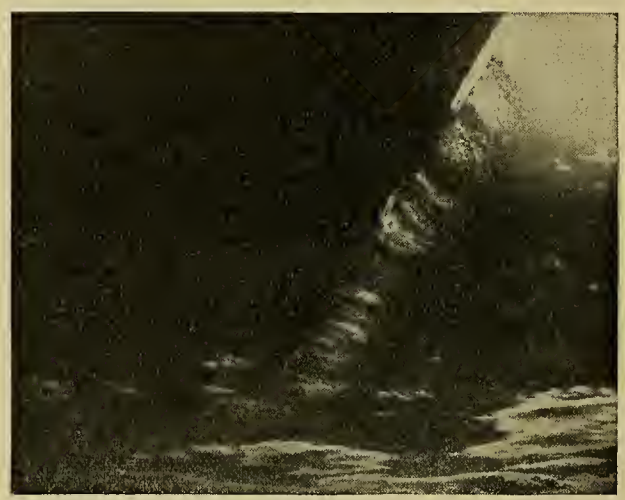


FIG. 13.—The formation of an ice-crystal cloud by dry-ice seeding of air supersaturated with respect to ice at a temperature of −18.5C.

stage of a stratus-cloud seeding, producing removal of cloud cover over a local region, is illustrated in Fig. 14. Not enough moisture was contained in these clouds to produce substantial amounts of precipitation.

The Effective Seeding of Clouds. To achieve maximum effectiveness in the form of energy release and precipitation, it is necessary to seed young, actively growing cumulus clouds in a region of the atmosphere which is potentially unstable. The clouds should be seeded when they have a vertical thickness above the freezing level of 3000–5000 ft. Care must be exercised that the dry-ice particles penetrate the entire supercooled region.

As mentioned earlier, the seeding of cumulus clouds with dry ice to effect a sudden release of the heat of sublimation should produce unusually intense convection in growing cumulus clouds. Such effects have been observed in experiments conducted in Australia [10], South Africa [1], Canada [17], and New Mexico [32]. Under optimum conditions, judicious seeding to produce this effect could lead to a local convergence by breaking through a limiting inversion layer or other restrictive atmospheric condition. At times it is conceivable that

this could develop into marked alterations of the general synoptic situation. Despite the skepticism of some meteorologists, it seems reasonable to presume that such end results are quite likely, since this would be in

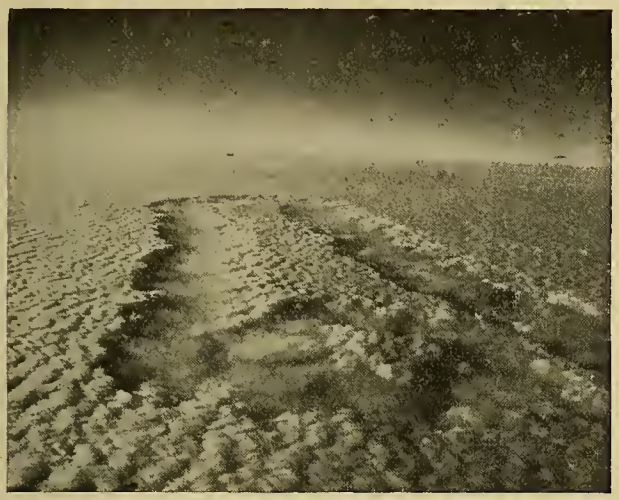


FIG. 14.—The effect produced in a supercooled stratus cloud with ½ pound per mile of dry ice. The straight legs of the pattern are 18 miles long.

accord with the physical processes which lead to the formation of certain types of widespread natural storms.

Figure 15 shows an area of cumulus clouds in New Mexico seeded with pellets of dry ice when the clouds were actively growing and had a vertical thickness above the supercooled layer of 5500 ft. The photograph illustrates the appearance of the clouds 33 min after four

FIG. 15.—A view of the cloud area in New Mexico seeded with dry ice. The towering cumulus to the right of center was seeded with one pound of dry ice 15 min earlier.

pounds of dry ice were placed in five local regions along the Monzano Mountains. Figure 16 shows the subsequent development of the rainstorm triggered off by the dry-ice seeding. This cloud system produced its initial precipitation 15 min after dry-ice seeding was completed. The development of this storm was followed in minute detail by radar, two ground stations, and two airplanes. No other precipitation could be detected by the radar within sixty miles, and the precipitation records show nothing within a hundred miles except from seeded clouds. An adjacent mountain range, the Sandia Mountains, to the northwest of the area used in these experiments, was employed as a control region. Although cumulus clouds similar to those which were seeded formed over the Sandias throughout the day, no

FIG. 16.—Rain showers falling from cumulus cloud 90 min after seeding with one pound of dry ice.

trace of precipitation occurred from them. A series of experimental studies in July 1950 produced similar results.

The Effects of Silver Iodide Seeding. Anomalous cloud effects were anticipated and have been observed in regions where experiments were conducted with silver iodide. Figure 17 shows a large precipitation area which

FIG. 17.—A rainstorm which originated in a region containing silver iodide produced by a ground generator.

began in air filled with silver iodide particles and formed rain several hours before any other clouds developed in the region on July 21, 1949 near Albuquerque, New Mexico. Further information on this subject may be obtained from Langmuir's and Vonnegut's papers [13, 38].

Possibilities and Limitations in Seeding of Clouds

In any science it is possible to establish the possibilities and limitations of certain physical phenomena.

In experimental meteorology, such factors as the stability of the atmosphere, the concentration of ice nuclei, the absolute humidity, the circulation of the air, and the heating by the sun are important.

The absence of clouds, the presence of strong temperature inversions, large variations in wind velocity with altitude, low absolute humidity, small lapse rate, weak circulation, large quantities of cirrus clouds, and other similar atmospheric situations present serious, if not absolute, limitations to cloud-seeding operations. Conversely, if large cumulus clouds occur without the formation of precipitation, if there is abundant moisture, large lapse rates, low concentrations of ice nuclei, an absence of wind shear, and few inversions, the possibilities of cloud modification are great and some unquestionable effects may follow the use of proper seeding techniques. However, even under ideal conditions, it is still possible to have poor or inconclusive results unless intelligent seeding techniques are followed.

With evidence now at hand it is possible to draw some fairly definite conclusions with regard to the possible effects of seeding methods under the following general conditions.

Thunderstorms. Thunderstorms generally occur in clouds of large vertical thickness containing large amounts of supercooled droplets and some ice crystals. By dissipating or overseeding all clouds in a thunderstorm "breeding area" [28] shortly after they pass above the freezing level, it should be possible to prevent the formation of a thunderstorm. This may be accomplished by local seeding with dry ice from an airplane or from projectiles from ground stations, but eventually it may be accomplished most effectively and over a larger area by the proper use of silver iodide generators at the ground. Initial success may be expected from orographic type thunderstorms rather than from the much more complex storms associated with frontal systems.

Hailstorms. Since it is believed that damaging hailstorms are due primarily to the presence of a deep supercooled layer and a relatively small number of ice nuclei, techniques similar to those employed in preventing thunderstorm development should be effective in preventing hail. Since such storms often possess special characteristics in regions where they are common, a careful study will suggest the best method of dealing with them. If certain features of the local topography favor their development, it may be possible to locate silver iodide generators so that their effluent will be carried into the most active part of the cloud system. It is important to provide so many ice nuclei that the competition for the available moisture is too great for individual particles to grow large. It would be more effective, however, to prevent the development of a deep supercooled layer and for this purpose a program of early seeding should be more effective.

Increased Precipitation from Cumulus Clouds. As suggested earlier, large cumulus clouds sometimes form without the development of a precipitation phase. High wind shear at upper levels in the atmosphere or a low concentration of ice nuclei are two of the factors responsible for such occurrences.

Proper seeding by artificial means may have economic value in several ways. Besides forming precipitation in the vicinity of such clouds, the process would prevent the formation of false cirrus streamers which sometimes reduce solar radiation to such an extent [30] that the formation of other convective clouds downwind is seriously affected. The formation of precipitation also may lead to increased air circulation which, if accomplished on a large enough scale, could change the weather pattern.

As suggested recently by Langmuir [14], the seeding of a given portion of the atmosphere with a sufficient concentration of ice nuclei to start a chain reaction serves primarily to increase the probability of precipitation coming from any given synoptic situation. Depending on the "triggering" nature of the seeding effect, the effectiveness of any particular seeding operation will involve many complex physical relationships in the atmosphere not yet understood in enough detail even to consider at this time.

The Prevention of Windstorms and Torrential Rains. Local windstorms of the type characterized by high gustiness and turbulence are often caused by downdrafts induced by heavy precipitation. By preventing the formation of thick cumulus clouds through judicious seeding techniques, this type of local storm damage may be prevented. Whether these techniques can be expanded to affect larger storm systems must await further experimental studies.

The Elimination of Ground Fogs. Under special conditions supercooled ground fogs may form by radiational cooling, contact of cold air with warm open water, or some similar condition. Seeding of such fogs may be effective in dispersal if care is exercised that they are not overseeded. An initial concentration not greater than 100 particles per cubic centimeter must be achieved. If the concentration is much greater, the fog will be overseeded and the visibility will become reduced from that existing in the supercooled fog.

No methods are known at present which would lend themselves to the effective dissipation of warm fogs.

The reduction of ice-crystal fogs which form at very low temperatures is also a difficult condition to overcome by seeding methods, since these fogs commonly form at temperatures below −39C. Under such conditions, any surplus of moisture exceeding the frost point comes out in the form of ice crystals. In the temperature range from 0C to −39C, the air may become supersaturated with respect to ice. Under such conditions, it may be possible to put out an optimum quantity of ice nuclei to precipitate the moisture as snow.

The Elimination of Supercooled Clouds in the Atmosphere. It is now feasible, using the mechanisms referred to in the preceding sections, to limit the formation of supercooled clouds in the atmosphere. This control over supercooled clouds is not necessarily limited to local regions; if desired it could be achieved throughout

the world. By using silver iodide in relatively insignificant quantities, a situation could be developed under which ice crystals would immediately form in *any* cloud which formed in air colder than −5C.

The anomalous effects which would follow from such an atmospheric condition are not predictable at the present time. It is obvious that important changes in storm patterns and climatic conditions could follow wherever these phenomena are governed by supercooled cloud structures.

The Use of Seeding Techniques for Studying the Atmosphere. By placing a known quantity of dry ice or silver iodide in a supercooled cloud it is possible to "mark" a cloud in such a way that subsequent changes in the physical nature of the cloud may be followed in great detail and with an accuracy hitherto impossible to achieve. By this method, the development of turbulence, the growth characteristics of snow crystals, the chain reaction mechanisms responsible for the development of widespread snowstorms, the change of insolation from cloudy sky to clear sky, the formation of new clouds, the development of precipitation, the occurrence of supersaturation, the properties of false cirrus, and many other atmospheric phenomena may be studied in great detail. Since it is possible to initiate many of these processes, it is possible for the first time to follow a particular meteorological process from beginning to end.

As experience is gained in this field, it is likely that these techniques will be most valuable in permitting us to reach a better understanding of that very complex but fascinating subject which we call *weather*. A careful scientific approach to these problems, using imagination, enthusiasm, and sound judgment, is of utmost importance if success is to be achieved in this new aspect of experimental meteorology.

REFERENCES

1. *Artificial Stimulation of Precipitation.* Council for Scientific and Industrial Research, Division of Meteorology. Pretoria, Un. of South Africa, 1948.
2. Bergeron, T., "On the Physics of Cloud and Precipitation." *P. V. Météor. Un. géod. géophys. int.*, Pt. II, pp. 156–178 (1935).
3. —— "The Problem of Artificial Control of Rainfall on the Globe. Part II, The Coastal Orographic Maxima of Precipitation in Autumn and Winter." *Tellus*, Vol. 1, No. 3, pp. 15–32 (1949).
4. Dorsey, N. E., "The Freezing of Supercooled Water." *Trans. Amer. phil. Soc.*, Vol. 38, Pt. 3, pp. 247–328 (1948).
5. Findeisen, W., "Die kolloidmeteorologischen Vorgänge bei der Niederschlagsbildung." *Meteor. Z.*, 55:121–133 (1938).
6. Fournier d'Albe, E. M., "Ice Nuclei in the Ultramicroscope and the Atmosphere." *J. Glaciol.*, 1:310–314 (1949).
7. Gathman, L., *Rain Produced at Will.* Chicago (publisher unknown), 1891.
8. Humphreys, W. J., *Rain Making and Other Weather Vagaries.* Baltimore, Williams & Wilkins, 1926.
9. Köhler, H., "An Experimental Investigation on Sea-Water Nuclei." *Nova Acta Soc. Sci., upsal.*, Ser. 14, Vol. 12, No. 6 (1941).
10. Kraus, E. B., and Squires, P., "Experiments on the Stimulation of Clouds To Produce Rain." *Nature*, 159:489–494 (1947).
11. Landsberg, H., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, (Supp.) 3:155–252 (1938).
12. Langmuir, I., "The Growth of Particles in Smokes and Clouds and the Production of Snow from Supercooled Clouds." *Proc. Amer. phil. Soc.*, 92:167–185 (1948).
13. —— "The Production of Rain by a Chain Reaction in Cumulus Clouds at Temperatures above Freezing." *J. Meteor.*, 5:175–192 (1948).
14. —— "The Control of Precipitation from Cumulus Clouds by Various Seeding Techniques." *Science*, 112:35–41 (1950).
15. —— Schaefer, V. J., and others, (A Series of More Than 20 Occasional Reports on Cloud Physics and Experimental Meteorology). Project Cirrus, Contract No. W-36-039-sc-32427, General Electric Res. Lab., Schenectady, N. Y.
16. Nakaya, U., Toda, Y., and Maruyama, S., "Further Experiments on the Artificial Production of Snow Crystals." *J. Fac. Sci. Hokkaido Univ.*, Ser. II, Vol. 2, No. 1, pp. 13–57 (1938).
17. Orr, J. L., Fraser, D., and Pettit, K. G., "Canadian Experiments on Artificially Inducing Precipitation." *Bull. Amer. meteor. Soc.*, 31:56–59 (1950).
18. Schaefer, V. J., "A Method for Making Snowflake Replicas." *Science*, 93:239–240 (1941).
19. —— "Use of Snowflake Replicas for Studying Winter Storms." *Nature*, 149:81 (1942).
20. —— *Final Report on Icing Research up to July 1, 1945.* A.T.S.C. Contract No. W-33-038-ac-9151, General Electric Res. Lab., Schenectady, N. Y., 1945.
21. —— "The Production of Ice Crystals in a Cloud of Supercooled Water Droplets." *Science*, 104:457–459 (1946).
22. —— "Properties of Particles of Snow and the Electrical Effects They Produce in Storms." *Trans. Amer. geophys. Un.*, 28:587–614 (1947).
23. —— "Methods of Dissipating Supercooled Clouds in the Natural Atmosphere." *J. Inst. Navig.*, 1:172–174 (1947).
24. —— "The Natural and Artificial Formation of Snow in the Atmosphere." *Trans. Amer. geophys. Un.*, 29:492–498 (1948).
25. —— "Types of Solid Precipitation in Snowstorms." *Weatherwise*, 1:124–125 (1948).
26. —— "The Production of Clouds Containing Supercooled Water Droplets or Ice Crystals under Laboratory Conditions." *Bull. Amer. meteor. Soc.*, 29:175–182 (1948).
27. —— "The Formation of Ice Crystals in the Laboratory and the Atmosphere." *Chem. Rev.*, 44:291–320 (1949).
28. —— *The Possibility of Modifying Lightning Storms in the Northern Rockies.* Occasional Rep. No. 11, 13 pp., General Electric Res. Lab., Schenectady, N. Y., 1949.
29. —— "The Detection of Ice Nuclei in the Free Atmosphere." *J. Meteor.*, 6:283–285 (1949).
30. —— *The Occurrence of Ice Crystal Nuclei in the Free Atmosphere.* Proc. Nat. Air Pollut. Symp., Pasadena, April 1950.
31. —— "Experimental Meteorology." *Z. angew. Math. Phys.*, 1:153–184; 217–236 (1950).
32. —— *The Effects Produced by Dry Ice Seeding of Cumulus Clouds in New Mexico.* Paper presented at the Amer. Meteor. Soc. Meeting, St. Louis, January 1950. (To be published.)
33. —— "The Formation of Frazil Ice and Anchor Ice in Cold Water." *Trans. Amer. geophys. Un.*, 31:885–893 (1950).
34. —— *The Rate of Growth of Snow Crystals.* Paper presented

to the Amer. Meteor. Soc. Meeting, Washington, May 1950. (To be published)

35. Smith-Johannsen, R., *Some Experiments on the Freezing of Water*. Occasional Report No. 3, 7 pp., General Electric Res. Lab., Schenectady, N. Y., 1948.
36. Veraart, A. W., *Meer Zonneschijn in het Nevelig Noorden; meer Regen in de Tropen*. Seyffardt's Boek en Muziekhandel, Amsterdam, 1931.
37. Vonnegut, B., "The Nucleation of Ice Formation by Silver Iodide." *J. appl. Phys.*, 18:593–595 (1947).
38. —— "Production of Ice Crystals by the Adiabatic Expansion of Gas." *J. appl. Phys.*, 19:959 (1948).
39. —— "Nucleation of Supercooled Water Clouds by Silver Iodide Smokes." *Chem. Rev.*, 44:277–289 (1949).
40. Weickmann, H., "Die Eisphase in der Atmosphäre." *Ber. dtsch. Wetterd. U.S.-Zone*, Nr. 6, 54 SS. (1949). (Wölkenrode Report and Translation No. 716, Feb. 1947, Library Translation No. 273. Ministry of Supply, Sept. 1948.)
41. Woodcock, A. H., and Gifford, Mary M., "Sampling Atmospheric Sea-Salt Nuclei over the Ocean." *J. mar. Res.*, 8:177–197 (1949).
42. Workman, E. J., and Reynolds, S. E., "A Suggested Mechanism for the Generation of Thunderstorm Electricity." *Phys. Rev.*, 74:709 (1948).
43. —— "Time of Rise and Fall of Cumulus Cloud Tops." *Bull. Amer. meteor. Soc.*, 30:359–361 (1949).

RELATION OF ARTIFICIAL CLOUD-MODIFICATION TO THE PRODUCTION OF PRECIPITATION

By RICHARD D. COONS and ROSS GUNN

Physical Research Division, U. S. Weather Bureau

Introduction

With the important discovery in 1946 that supercooled cloud elements could be artificially converted to ice crystals by introducing pellets of carbon dioxide snow [11, 12], the possibility of human control of weather seemed imminent. Claims and speculation concerning the degree of possible weather control and the probable amount of artificially induced precipitation were rife, and meteorological opinion ranged from the one extreme "of academic value only" to the other "of great economic and military importance." The more favorable claims were supported in part by a few incompletely documented single experiments conducted in 1946 and 1947. These claims excited the interest of the public and stimulated demands for immediate weather control, drought relief, and storm diversion and dissipation, even though at the time there was no definite objective evidence that these were possible.

Alfred Wegener [15] was the first to suggest that the coexistence of ice and water in supercooled clouds led to colloidal instability resulting in rapid growth of ice particles. The possibility of modifying and producing rain from supercooled clouds by adding sublimation nuclei was foreseen by Bergeron [1, 2] and Findeisen [8]. Until 1946, however, no method was known of producing the necessary ice-crystal nuclei artificially. Earlier experiments by Veraart [13] in Holland in 1930 in which he dropped solid carbon dioxide, among other things, into supercooled clouds must have produced such nuclei, although Veraart at that time did not recognize this possibility. He was credited with producing slight amounts of rain on several occasions. However, because of his sweeping claims, even his positive results were discredited. Not until Schaefer and Langmuir [12] demonstrated (1) that a single solid carbon dioxide pellet could produce prodigious numbers (of the order of 10^{16}) of sublimation nuclei in its fall through a supercooled cloud, and (2) that, in actual field tests, a number of CO_2 pellets dropped into a supercooled cloud converted it largely into ice crystals, was it possible to investigate the importance of artificial nucleation in producing rain and modifying clouds as previously suggested by Bergeron and Findeisen. More stable sources of sublimation nuclei have also been found suitable. Silver iodide in particular has been used most successfully [14]. It should be noted that Heverly [9] has demonstrated in the laboratory that spontaneous freezing of water droplets in supercooled clouds may be a more important natural precipitation instrument than the sublimation of water vapor onto suitable active nuclei. According to Heverly's results, the effect of dry ice in producing colloidal instability could possibly be due either to the resulting spontaneous freezing of the supercooled droplets or to the production of myriads of ice-crystal germs from the water vapor in the cloud. In any case, whatever the effect of the dry ice, there is no question that it is capable of almost completely converting seeded portions of supercooled clouds into ice particles. The optical effects of such conversions are startling. In Fig. 1 the reflection of the sunlight from the ice-crystal cloud produced by seeding is seen to be brilliant when compared to the light reflected from the adjacent supercooled water-droplet clouds.

FIG. 1.—Reflection of sunlight on aluminum aircraft wing, ice-crystal cloud, and supercooled water-droplet cloud. Note that the reflection from the ice-crystal cloud is nearly as bright as that from the wing.

Bergeron [3] has fully discussed the suitability of the various types of rain-producing clouds for artificial release of precipitation. Briefly, the following conditions must be met in a cloud or cloud system in order for artificial nucleation to be effective in the modification of clouds and the production of precipitation: (1) The cloud must contain water droplets at a temperature below freezing, while there are still no ice crystals present in the immediate vicinity or falling through the cloud; (2) it must have some minimum thickness (arbitrarily about 1000 m) in order for the amount of moisture that could be converted into precipitation to be appreciable; (3) there probably should be some vertical instability in order to allow the growth, if any, of the cloud to spread the ice crystals as a result of the heat liberated by the freezing of the water (this would be accentuated in the presence of large-scale horizontal convergence that would provide a continuing source of moisture into the seeded area). Less important factors determining the possibility that precipitation will reach the

ground include (4) the height of the cloud base above the terrain, and (5) the humidity of the air below the cloud. In view of all the specialized conditions and factors involved, one of the objectives of conclusive tests of cloud-seeding potentialities must certainly be the determination of the frequency of such conditions and the importance of such factors. The question and importance of particular seeding rates, the one variable under control of the experimenter, are somewhat nebulous and will be discussed later in the light of some experimental results.

The Cloud Physics Project

In view of the divergent claims made by both qualified and unqualified experimenters in the rain-making field, the requirement for a long series of well-controlled objective tests became obvious. Such a program, the Cloud Physics Project, was initiated by the U. S. Weather Bureau, the U. S. Air Force, and the National Advisory Committee for Aeronautics in 1947. Since the 170 tests of that project are the only well-documented series yet reported [4, 5, 6, 7], the following discussion of seeding results will be based upon them.

The first requirement for an evaluation program is that it be able to separate results of seeding from natural effects. This is especially important in view of the fact that the necessary conditions for successful seeding operations approach those required for natural precipitation processes. The Cloud Physics Project experiments were designed to use the maximum number of controls and observational tools available for such an atmospheric investigation. Tests were first conducted for a period of about one year in the vicinity of Wilmington, Ohio, where ideal facilities were available [4, 5]. Further tests were then conducted from Sacramento, California [6] in the Sierra Nevada Range, and from Mobile, Alabama [7] along the Gulf Coast. At Wilmington, all of the project seeding operations were controlled from a central radar site. The high-powered 10-cm V-beam radar was of utmost importance in these tests. Not only was it capable of indicating and marking permanently on film the paths of each of the aircraft involved in the investigation, but also it indicated the presence of precipitation areas, whether they were formed naturally or as a result of seeding. This radar then was an invaluable tool which indicated if the cloud chosen visually for seeding, or for that matter other clouds in the vicinity, had precipitation echo sources before seeding, while, on the other hand, it indicated if echoes resulted exclusively from seeding. This radar, of course, was also of great help in keeping the observational aircraft properly positioned with respect to the cloud area being investigated.

Whenever possible, the project's operations were carried out over a 55-station surface network south of Wilmington. The rain gages at these stations, along with recording gages of the hydrometeorological service, were used to measure precipitation amounts. Also available in the Wilmington area were two rawinsonde stations which provided, as required, information regarding the temperature, pressure, moisture, and winds of the upper atmosphere.

A number of aircraft were used in most of the seeding flights. A B-17 usually performed the seeding operation in an area such that its effects would be noted and measured over the surface rain-gage network. Its seeding conveyor was carefully calibrated and indicators of the rates and periods of seeding were photographed continuously from remote-reading instruments installed in a photo panel along with indicators of other important parameters, such as temperature, pressure, altitude, and electric field. Also installed in the seeding aircraft was a low-powered X-band radar which would indicate the presence of light-precipitation areas. Photographs of the seeded cloud were taken from this aircraft both before and after seeding. Other aircraft which participated in the flights were high-altitude photo planes which took vertical and oblique photos of the seeded area, and observation planes which flew above, within, and beneath both the seeded cloud and neighboring clouds to determine their characteristics. Paths of all of the aircraft participating in the missions were recorded on the photographs of the ground radarscopes. From these photos, it was possible to determine the exact location of any of the planes at any instant, since each of them was identified by its own radar-triggered beacon code. All oral descriptions of seeding operations and results were recorded on continuous records, making it possible to determine easily the exact time of any observation. Since the position of the particular aircraft at the time of the observation was known from the radar data, the location of the observation was also known.

In summary, observations of the results of the seeding operations were made from several points of vantage: the radar, the aircraft flying above and below the seeded cloud, and on the ground (by visual observers). The quantities measured were the height of the base and top of the seeded cloud, the relative humidity above and below, the temperatures inside and outside the cloud, lapse rates, optical characteristics, and the extent and character of resulting precipitation. In the Sacramento and Mobile operations, not all of these quantities were available. However, the basic observational plan was followed with a high-powered 10-cm radar installed in a B-17 used to obtain the important radar information.

Experiments with Stratus Clouds

The first results to be discussed will be those obtained by the Cloud Physics Project in stratus clouds. Altogether 37 seedings in such clouds were conducted in the Wilmington area in 1948 and 15 seedings into orographic clouds were made in the Sacramento area in 1949. The objectives of these tests, as set forth in 1948, were to determine "the practical limits and general utility of cloud modification processes in producing or suppressing precipitation and increasing the visibility for flying aircraft."

On January 21, 1948, over one inch of snow fell in the vicinity of Wilmington, Ohio, within eighteen hours

of the completion of a cloud-seeding mission [4]. Since this snow fell in exactly the area where the seeding had been done and was somewhat local in character, it seemed that it had resulted from the artificial nucleation processes. However, a detailed analysis of the information secured during the mission, especially the radar pictures, indicated that this was definitely not the case. Actually, the measured winds carried the seeded clouds outside the precipitating area within a couple of hours. A layer of altostratus clouds, based at 4000 ft and extending to 8700 ft, where the temperature was −13C, was seeded with dry-ice pellets at the rate of four pounds per mile. The first task in analyzing the data was to determine the exact place of the seeding. Since the seeding run was made from 1443:45 to 1445:15 EST, its exact location could be found from the pictures of the radarscope during that period. The seeding aircraft's location was easily marked by its coded beacon signal during the run. Before the seeding, it had been determined that the cloud deck was composed wholly of supercooled water droplets, but ice crystals were observed to form immediately along the seeded line as the dry ice was dropped. Observers in an aircraft flying 12,000 ft above the cloud top reported that the seeding aircraft looked like a snowplow moving across the cloud deck. A trough about 300 ft deep resulted as the conversion to ice crystals took place while, along the perimeter of the affected area, clouds built up an estimated 100 ft. This trough, which widened to 2½ miles, was easily recognizable for about 45 min, after which mixing dissipated the lines of demarcation between the treated clouds and the rest of the deck. The first radar echo resulting from the seeding was observed about 30 min after the drop (see Fig. 2). The position of this echo

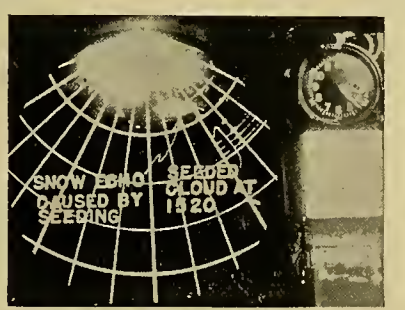


Fig. 2.—Radarscope photograph, January 21, 1948 (first seeding).

agreed exactly with the location of the seeded area as indicated by the observing aircraft and also with its position computed from the winds aloft. This seeding echo was approximately three miles wide by about eight miles long and never grew any larger. By 1540, about one hour after the seeding, it had almost completely disappeared. It can be concluded, then, that the seeding resulted in the production of a precipitation area from which a limited radar echo was observed for about 30 min. This echo, however, was not unique at the time of seeding, since other echoes of similar size and natural origin were also indicated. This seeded area was observed to move with the prevailing winds at the cloud altitude and, by the time it disappeared, it was some twenty miles to the east of the rain-gage-network area. At 1540, a new echo directly over the rain-gage area was observed to form, completely independent of the seeded echo. Snow continued to fall for some time from the unseeded cloud deck as it passed over the network, but it was not until approximately 1800 that enough had fallen to affect the rain gages. A complete series of photographs like Fig. 2 is given for this test in Weather Bureau Research Paper No. 30 [4].

In summary, to quote from this paper:

> A definite change in texture of the supercooled cloud deck was caused by the first seeding. Shortly thereafter, a radar echo was noted to be associated with the seeded area and snow was observed underneath it. However, at the same time in the immediate vicinity, snow echoes were existing or were forming. Within one hour after the seeding, the area could no longer be distinguished by the observers flying in the aircraft above the cloud. The snow echo as indicated on the radar scope had disappeared also. At no time was there a definite hole broken through the cloud as a result of the seeding, although there were large breaks in the clouds nearby.

A study of the pictures presented in that report indicates that some few ice crystals remained in the seeded area and acted as an obstruction to visibility even when the snow was reported to be falling from beneath the seeded line. Whether or not the seeding resulted in a hole in the cloud deck does not seem to be a meaningful question in this case, since the same pictures indicate very large natural breaks immediately bordering the seeded area.

Another seeding conducted in the Wilmington area, that of October 5, 1948, was also successful in producing a radar echo in the exact pattern of the seeding. On this particular day, the cloud deck extended from 8000 ft to 11,600 ft, where the temperature was −5C. Dry-ice pellets were dropped at the rate of five pounds per mile in an L pattern. Immediately after the seeding run, a texture change and the horizontal diffusion of ice crystals were observed. The area continued to spread slowly and a slight boiling effect was observed along the extremities of the pattern. At 1445, almost exactly 30 min after the dry ice was dropped, a small radar echo was observed from the seeded area. Within 5 min this echo assumed the shape of the L pattern (Fig. 3) but within 5 min thereafter completely disappeared. Observers flying underneath the seeded pattern reported that light rain fell at about the time of the maximum radar echo. They indicated that the rain was of short duration and possibly did not reach the ground.

On this day, project pilots indicated that a visual descent would have been possible in the seeded area at about the time of maximum dissipation, which was also the time of maximum radar echo. However, this dissipation was not considered particularly significant in view of the fact that much larger natural openings existed within a few miles of the seeded area. (See Fig. 4, which shows the seeded L in the upper middle

of the picture with large natural openings in the foreground.) These two examples illustrate that even the most favorable seeding missions in the Ohio tests resulted in only slight precipitation amounts and in dissipation only when the clouds were breaking up due to natural conditions. A study of the data presented in the Ohio report indicates without exception that such large-scale seeding effects, as illustrated above, were possible only when the cloud deck treated was undergoing natural dissipative action. On the "unfavorable" days, seeding resulted in the production of narrow lines of ice crystals which did not diffuse through the supercooled cloud deck.

FIG. 3.—Radarscope photograph of seeding, October 5, 1948, showing L-shaped area of light rain.

FIG. 4.—First seeded area as seen from 10,000 ft above cloud deck, photographed thirty minutes after seeding (October 5, 1948). Note natural breaks in clouds.

The Sierra Nevada tests were conducted for the purpose of evaluating the importance of dry-ice seeding in orographic type stratus clouds. It was felt that the clouds formed along the Sierra Nevada Range by the mechanical lifting of moist Pacific air would offer the most favorable conditions for seeding. The report for the Project's operations near Sacramento [6] indicates that its seeding results were no more successful than those in Ohio [4]. First of all, it was found that nonfrontal orographic clouds existed only a fraction of the time, and that the majority of cloud systems over the Sierra Nevada Range were of frontal origin and contained natural ice-crystal clouds before lower supercooled decks were formed. On only three days during a two-month operational period was it possible to find nonfrontal clouds and, on each of these days, the lack of large-scale horizontal convergence in the general area resulted in somewhat broken cloud decks. A discussion of the seeding of February 15, 1949, is included here since on this particular day fairly large-scale dissipation results were observed. An altocumulus deck extending from 9800 ft to 10,650 ft where the temperature was −7C was seeded in the shape of an L along the base of the Sierras near Auburn, California. Dry ice was used as the seeding agent and was dropped at the rate of four pounds per mile. Immediately after seeding an ice-crystal pattern conforming to the dry-ice pattern was observed and within some 20 min the ice-crystal area had spread about 1½ miles. Within one hour after the seeding, definite holes through to the ground, especially along the borders of the seeded pattern, were observed and at the apex of the L a large opening some three to four miles across was observed. However, the significance of this dissipation becomes questionable in view of the natural dissipation occurring simultaneously. The natural openings could easily be seen, and within a half-hour after the observation of the large seeded opening, the whole cloud deck began to break up rapidly and disappeared shortly thereafter. No precipitation of consequence could have fallen as a result of seeding, since its effects did not spread over a large area nor did it cause any large-scale convection. The only amount of moisture that could have fallen was that which could have been extracted from the 800-ft thick cloud.

Other experimenters in the field have indicated that they believed that the Cloud Physics Project was not successful in causing large-scale dissipation because of its use of excessive amounts of dry ice, usually two to five pounds per mile, and they recommended that rates of one pound per mile or less be used. They felt that overseeding had occurred in most of the cases studied by the Cloud Physics Project and that, as a result, the ice crystals which formed did not grow large enough to precipitate from the cloud deck in sufficient numbers actually to cause it to dissipate. As a matter of fact, they suggested the use of only one pellet of dry ice to open relatively large areas. Four such single pellet drops by the Cloud Physics Project during its California tests gave no results on a scale large enough to be recognized. There is no doubt that there exists some optimum ratio of the specific number of droplets to ice nuclei for most favorable seeding results. Bergeron [3] has discussed this in some detail in his paper and gives examples of the effects of the various ratios from underseeding to overseeding. It is readily apparent there is no one seeding rate which will give the

optimum ratio because of the variability of the water-droplet distribution in different stratus cloud systems and because of differing diffusion conditions. It is obvious, considering the tremendous number of nuclei created from the fall of one dry-ice pellet, that there will be gross overseeding in the cloud segment through which such a pellet falls while at some distance away, depending upon diffusion conditions, there will be underseeding. Somewhere, then, between where the pellet passes through the cloud segment and where its nucleation effects are just felt, there must exist the optimum ratio for the particular cloud. It is seen, then, that the problem of overseeding is not a consequence of excessive amounts of dry ice thrown out along a path, but rather is due to the lack of uniform dispersion by the diffusion of the tremendous numbers of nuclei produced by even one dry-ice pellet. To quote Langmuir [12] "The conclusion is obvious, however, that with a reasonable number of pellets dropped along a flight path into the top of a cloud, the limiting factor will not be the number of nuclei but the rate at which the nuclei can be distributed throughout the cloud." In cumulus clouds, the characteristic internal turbulence and mixing would spread the nuclei from even one pellet throughout a large portion of the cloud within a short period of time. However, in characteristically stable stratus clouds, only horizontal divergence or mixing would tend to distribute the nuclei through a large cloud volume. These factors are of course the same ones that result in natural dissipation of stratus clouds. Thus, the results of the Cloud Physics Project emphasize the importance of divergence and mixing by diffusion. In both the Ohio and California tests, spread of artificially induced nuclei and subsequent stratus dissipation were observed only when natural dissipation was occurring. The effect of heats of fusion and sublimation released as a result of seeding was never observed to be on a scale large enough to cause any important local circulation and resulting spread of artificial nucleation.

Experiments with Cumulus Clouds

It appears that the most suitable cloud for the artificial production of precipitation would be a cumulus congestus whose top, having reached above the freezing level, would contain supercooled water while still not containing ice crystals. If seeding resulted only in the conversion of the water in such a cloud into precipitation, it is possible that a maximum of 0.2 in. of rainfall might reach the ground under the most favorable circumstances. In addition, it has been suggested that the heat released in the conversion of the cloud top to ice crystals might be of sufficient value to set off additional convection within the treated cloud and thus to result eventually in the production of a heavy rain shower or thunderstorm.

The investigation of the results of seeding cumulus clouds is somewhat more difficult than the investigation of similar results in stratus clouds. In the latter, the untreated portion of the seeded deck can be used as a control or a standard, while the ever-changing configuration and often violent activity of cumuli make it impossible to use any one of them as a control. It has been suggested that the cloud to be seeded be chosen randomly from among those possessing the required characteristics and that, for each cloud seeded, a similar one, also chosen randomly, be observed in its natural course. However, the results of the Cloud Physics Project in seeding cumulus clouds indicate that the control exercised through the use of the several observational aircraft and, especially, of a rain-sensitive radar, made the scheme described above unnecessary. Altogether the project conducted 79 seedings in cumulus clouds in the Ohio area [5] and 44 in semi-tropical cumulus along the Gulf Coast [7]. A number of clouds seeded in the Ohio tests were not supercooled but were seeded with water droplets for the purpose of investigating the reality of "chain reaction processes" within warm clouds as envisaged by Langmuir in 1948 [10]. The results of seedings in such warm clouds appeared to be unfavorable from the relatively few tests conducted, and they were not continued in the Gulf Coast tests, but rather the more favorable supercooled cumulus clouds were investigated there.

As indicated in U. S. Weather Bureau Research Paper No. 31 [5], only trivial amounts of precipitation resulted from seeding the Ohio cumulus clouds and even in cases where these amounts were observed, there were nearby natural showers of the same or greater intensity. In none of the cases reported did the seeded cloud build extensively as a result of the dry-ice drop and it is thus obvious that no large amounts of precipitation could have been expected to fall. On the contrary, the most obvious effect of seeding dry ice into both supercooled and nonsupercooled cumulus clouds was the resulting rapid dissipation. There were very few cases in which the cloud built up after seeding, and in each of these the growth did not exceed a few thousand feet. On the other hand, dissipation was usual and often almost complete in its effect. The following is quoted from this report [5]:

> This dissipation appeared to be nearly independent of supercooling or of the particular agent employed. Its occurrence was consistent with the idea that convective clouds often have lapse rates steeper than the moist adiabatic as a result of mixing and entrainment between such clouds and the environment. A downward movement initiated by dry ice, large numbers of ice crystals, water or other means (aircraft flying vertically upward through the cloud have been employed successfully) might easily cause an appreciable mass of air to become colder than the surrounding clouds, and thus induce further downward motion. A similar explanation has been advanced by Byers and Braham for the formation of the thunderstorm downdraft.

Regarding the general results in the Ohio area, the following is also quoted from the same report:

> The experiments showed that the artificial modification of cumuliform clouds is of doubtful economic importance for the production of rain. Dissipation rather than new development was the general rule. There is no indication that seeding will initiate self-propagating storms, and therefore, the only precipitation that can be extracted from a cloud is that con-

tained within the cloud itself. The methods are certainly not promising for the relief of drought.

In the Weather Bureau report are included detailed examples of three different cases in which the radar, in particular, and other operational tools available to the Cloud Physics Project proved invaluable in assessing the results of seeding. One of the examples illustrates a case in which a small shower which resulted from the seeding was differentiated from other natural showers occurring almost simultaneously within five miles. In the second example, there is discussed a case in which a small radar echo and accompanying shower resulted from seeding when there were no other showers indicated by the radar within fifty miles. In the last example covered in the report, an untreated tower which was part of the seeded cloud showed an echo previous to the dry-ice drop, indicating that precipitation had already started in that portion. The tower that was seeded did not give any precipitation. However, observers underneath the cloud reported rain which, as a result of the detailed analysis procedure, was definitely attributed to the tower that showed an echo previous to the seeding. These cases show that the use of all available controls, such as radar, is most important in properly evaluating the results. Without a radar in the first example, the adjacent natural showers might either have been overlooked or, on the other hand, attributed to the seeding operation. In the last example, the rain which was already falling from the seeded cloud might easily have been attributed to the effect of the dry ice. This third example illustrates a real difficulty in the seeding of cumuliform clouds. In the Ohio area many of these clouds reached the spontaneous freezing level soon after surpassing the 0C isotherm, thus occasionally making it impossible to find clouds extending above the freezing level which did not have ice crystals in them before seeding. The following results of the tests conducted at Mobile, Alabama, indicate this to be a particularly important difficulty in the seedings conducted in that area.

In clouds over the Gulf States [7], ice crystals would usually form naturally just a few thousand feet above the freezing level, often when the temperature at the tops of the clouds was not below −6C. Such clouds would appear to be composed of only water droplets when observed visually from outside, but the occurrence of radar echoes and subsequent visual observation of ice crystals within these clouds proved external visual observation to be unsatisfactory and somewhat misleading. Thus, there was only a short period of time during cloud development in which seeding was likely to be effective; that is, between the time the cloud reached the freezing level and the time it reached the natural crystallization level, from 3000 to 6000 ft higher. As in the Ohio tests, there was no appreciable building as a result of seeding, but rather dissipation of the cloud top was the rule. Occasionally, clouds containing only ice crystals were seeded because of the unavailability of strictly supercooled ones. The report of the Cloud Physics Project indicates that even in these cases seeding hastened the process of dissipation, possibly by increasing the number of nuclei available within the cloud.

Three specific examples of the Mobile seeding operation seem of interest. On June 5, 1949 the cloud under study had its base at around 5600 ft with the top at 18,000 ft where the temperature was −6C. It was seeded twice with dry ice at the rate of five pounds per mile. On the first run, only supercooled water was observed to freeze on the aircraft. This seeding resulted in the dissipation of the pinnacle traversed during the run. On the second seeding run, 13 min later, both supercooled water and ice crystals were observed. Shortly after the dry-ice drop, a faint radar echo appeared in the seeded area, while at the same time another echo formed in a cloud four miles away. This natural echo grew to a much larger size than that in the seeded cloud. Figures 5 and 6 show the cloud before and after the second seeding and illustrate its rapid dissipation into a small stratified cloud. The second picture shows a rainbow from the light precipitation which fell from the cloud. It is interesting to note in Fig. 5 the concurrent thunderstorms in the near background.

FIG. 5.—Second seeding, June 5, 1949 (before seeding). Altitude, 18,000 ft; time, 1452; azimuth, 300°.

A very favorable situation for seeding is illustrated in the third seeding of June 7, 1949. A cloud whose base was near 5000 ft and whose top was estimated to

FIG. 6.—Second seeding, June 5, 1949 (after seeding). Note rainbow under cloud. Altitude, 18,000 ft; time, 1503.

be at 24,000 ft where the temperature was −25C was seeded with dry ice at the rate of five pounds per mile. This cloud contained only supercooled water at the 20,000-ft seeding altitude and no radar echo was observed from it prior to the seeding. Within 10 min after the seeding, a small rain echo was observed from the cloud and light virga was observed by the low observation aircraft. It was estimated that within about one-half hour after the seeding this cloud had dissipated approximately 75 per cent. Figures 7 and 8 are

FIG. 7.—Third seeding, June 7, 1949 (before seeding). Altitude, 20,000 ft; time, 1430; azimuth, 360°.

pictures of the cloud immediately before and 14 min after seeding. They show its rapid conversion into ice crystals.

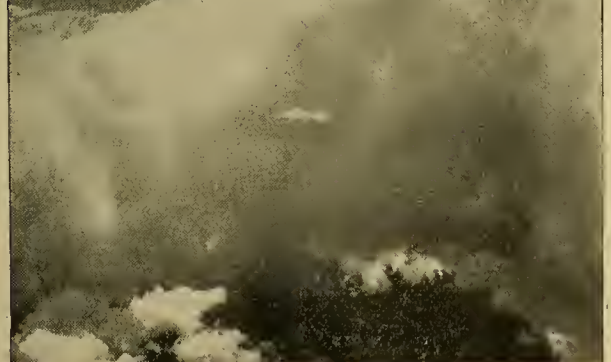


FIG. 8.—Third seeding, June 7, 1949 (after seeding). Altitude, 21,000 ft; time, 1444; azimuth, 360°.

Considering the result obtained in the Ohio tests, and especially in the Gulf tests, the project report [7] concludes in part that

> . . . it seems a logical conclusion that seeding, by usually stopping the development of cumulus soon after they reach the freezing level, possibly interferes with the growth of some few of them to full-scale thunderstorms. All of our observations seem consistent with this conclusion. Seeding then might actually be a deterrent to the production of precipitation over any particular area. It apparently inhibits the growth of cumulus clouds by initiating premature downdrafts of ice crystals which subsequently choke off the necessary moisture inflow at lower altitudes and at lesser stages of development than those decreed by nature.

The large number of independent experiments conducted by the Cloud Physics Project have been generally disappointing in terms of the practical and economic results hoped for. However, the treatment of natural clouds by large numbers of artificially dispersed sublimation nuclei has frequently induced weather perturbations whose study has added notably to our scientific understanding of precipitation processes. The objective analysis of detailed changes occurring in the tremendous spaces of the atmosphere is extremely difficult even with all the instruments and facilities available. Future progress will depend, in no small measure, upon the invention and development of better airborne instruments suitable for making rapid determinations of the detailed characteristics of each cloud to be explored.

REFERENCES

1. BERGERON, T., "Über die dreidimensional verknüpfende Wetteranalyse." *Geofys. Publ.*, Vol. 5, No. 6 (1928). (See Part I)
2. —— "On the Physics of Cloud and Precipitation." *P. V. Météor. Un. géod. géophys. int.*, Pt. II, pp. 156–178 (1935).
3. —— "The Problem of Artificial Control of Rainfall on the Globe. Part I, General Effects of Ice Nuclei in Clouds." *Tellus*, Vol. 1, No. 1, pp. 32–43 (1949).
4. COONS, R. D., GENTRY, R. C., and GUNN, R., "First Partial Report on the Artificial Production of Precipitation—Stratiform Clouds—Ohio, 1948." *Bull. Amer. meteor. Soc.*, 29:266–269 (1948). (*U. S. Wea. Bur. Res. Pap.* No. 30, same title (1948).)
5. COONS, R. D., JONES, E. L., and GUNN, R., "Second Partial Report of the Artificial Production of Precipitation —Cumuliform Clouds—Ohio, 1948." *Bull. Amer. meteor. Soc.*, 29:544–546 (1948). (*U. S. Wea. Bur. Res. Pap.* No. 31, same title (1949).)
6. —— "Third Partial Report of the Artificial Production of Precipitation—Orographic Stratiform Clouds—California, 1949." *Bull. Amer. meteor. Soc.*, 30:255–256 (1949). (*U. S. Wea. Bur. Res. Pap.* No. 33, same title (1949).)
7. —— "Fourth Partial Report on the Artificial Production of Precipitation—Cumulus Clouds—Gulf States, 1949." *Bull. Amer. meteor. Soc.*, 30:289–292 (1949). (*U. S. Wea. Bur. Res. Pap.* No. 33, same title (1949).)
8. FINDEISEN, W., "Die kolloidmeteorologischen Vorgänge bei der Niederschlagsbildung." *Meteor. Z.*, 55:121–133 (1938).
9. HEVERLY, J. R., "Supercooling and Crystallization." *Trans. Amer. geophys. Un.*, 30:205–210 (1949).
10. LANGMUIR, I., "The Production of Rain by a Chain Reaction in Cumulus Clouds at Temperatures above Freezing." *J. Meteor.*, 5:175–192 (1948).
11. SCHAEFER, V. J., "The Production of Ice Crystals in a Cloud of Supercooled Water Droplets." *Science*, 104: 457–459 (1946).
12. —— LANGMUIR, I., and others, *First Quarterly Progress Report, Meteorological Research.* General Electric Res. Lab., 15 July 1947.
13. VERAART, A. W., *Meer Zonneschijn in het Nevelig Noorden; meer Regen in de Tropen.* Seyffardt's Boek en Muziekhandel, Amsterdam, 1931.
14. VONNEGUT, B., "The Nucleation of Ice Formation by Silver Iodide." *J. appl. Phys.*, 18:593–595 (1947).
15. WEGENER, A., *Thermodynamik der Atmosphäre.* Leipzig, J. A. Barth, 1911. (See pp. 94–98)

THE UPPER ATMOSPHERE

General Aspects of Upper Atmospheric Physics *by S. K. Mitra* 245

Photochemical Processes in the Upper Atmosphere and Resultant Composition *by Sydney Chapman* 262

Ozone in the Atmosphere *by F. W. Paul Götz* 275

Radiative Temperature Changes in the Ozone Layer *by Richard A. Craig* 292

Temperatures and Pressures in the Upper Atmosphere *by Homer E. Newell, Jr.* 303

Water Vapour in the Upper Air *by G. M. B. Dobson and A. W. Brewer* 311

Diffusion in the Upper Atmosphere *by Heinz Lettau* 320

The Ionosphere *by S. L. Seaton* 334

Night-Sky Radiations from the Upper Atmosphere *by E. O. Hulburt* 341

Aurorae and Magnetic Storms *by L. Harang* 347

Meteors as Probes of the Upper Atmosphere *by Fred L. Whipple* 356

Sound Propagation in the Atmosphere *by B. Gutenberg* 366

GENERAL ASPECTS OF UPPER ATMOSPHERIC PHYSICS

By S. K. MITRA

University of Calcutta

INTRODUCTION

The present article is intended as an introduction to the articles on the different aspects of the upper atmosphere which follow. The main topics of this contribution will be a discussion of the physical origin of the upper atmosphere, methods of investigation and a general survey of the contemporary state of our knowledge of the upper atmosphere, and an account of the problems which still remain unsolved. There will, of necessity, be references to many subjects which are dealt with in the articles that follow. These references will be included at the risk of a certain amount of repetition, as the aim of the present article is to give a general idea of the physical aspects of the upper atmosphere. The reader who desires detailed knowledge in any particular subject should seek it in the specialised articles that follow. It should also be mentioned that the treatment of some of the topics in the present article will follow closely that given in the author's recent book [72].

The region of the earth's atmosphere denoted by the term *upper atmosphere* is not yet well defined. To the meteorologist, upper atmosphere may mean the regions investigated by the conventional sounding balloons; to the geophysicist studying the aurora or the ionosphere, upper atmosphere may mean the high regions near and above 100 km. In the present article, the words *upper atmosphere* will generally refer to the regions above the troposphere. However, for purposes of reference the whole atmosphere may sometimes be divided into three regions: the lower atmosphere (troposphere); the middle atmosphere (stratosphere); and the upper atmosphere, extending above 100 km. The lower stratosphere, as reached by sounding balloons, will be generally left out of consideration in the present discussion.

In studying the upper atmospheric region it is helpful to bear in mind its high tenuity. At 15 km, near the base of the stratosphere, the atmospheric density is about one-eighth of that at sea level. Above this the density falls off rapidly, and at 100 km it is only a millionth of that at sea level. The pressure is thus about 10^{-3} mm, which is of the same order as in the so-called "vacuum" of ordinary electric bulbs. At 300 km the pressure is of the order 10^{-6} mm. This is the pressure attained by only high quality modern vacuum pumps. If we consider the mean free path of a molecule, we note that its value near sea level is 10^{-5} cm, at 15 km it is 10^{-4} cm, at 100 km it is 10^{2} cm, and at 300 km it is 10^{6} cm.

It may seem surprising that regions of such extreme tenuity in the upper atmosphere could be the seat of any phenomenon of geophysical importance or of interest for any aspect of our daily life. Nevertheless such is the case. But for the presence of the ionospheric regions at heights of 200 km and above, long distance radio communication at night would be impossible. Auroral displays which illuminate the long winter nights in the polar regions occur with greatest frequency near a height of 100 km, and auroral streamers sometimes extend up to heights of 1000 km and beyond. The common phenomenon of shooting stars appears most frequently in the region 50 to 150 km. During World Wars I and II the sound of cannon fire in France could be heard in England because the wave of explosion was bent downward by refraction at heights of about 35 km.

Many upper atmospheric phenomena are also of great scientific interest because they occur under conditions and on a scale which cannot be reproduced in laboratories. As a matter of fact the upper atmospheric regions may be regarded as constituting a vast physical laboratory where Nature carries out experiments on a gigantic scale on such phenomena as bombardment of air masses by charged and uncharged particles, electric discharge, magnetic double refraction, ionization by collision, photochemical reaction, and recombination of ions and electrons. In laboratory experiments the ionization track of a charged particle in a Wilson cloud chamber may be only a few centimetres long; in the upper atmosphere it may be hundreds of kilometres long. In a laboratory, rarefied gas for study of discharge phenomena has of necessity to be confined within a glass vessel. The walls of the vessel are responsible for the quick disappearance of electrons and ions and the consequent extinction of the discharge glow. In the rarefied upper atmosphere there are no glass walls. Unhindered by the wall effect, the ions and the electrons and the luminescence persist for a long time.

One further point of interest in connection with upper atmospheric phenomena may be mentioned. Unexpected associations (some still unexplained) have been found to exist between physical conditions in the upper atmosphere, tens or hundreds of kilometres above the surface of the earth, and weather conditions in the tropospheric regions.

Finally, accurate knowledge of the physical properties of the high regions of the atmosphere is of utmost importance to those engaged in the development of the many new types of high-flying aircraft.

THE UPPER ATMOSPHERE

We may now consider the origin of the observed structure of the upper atmosphere. If the earth's atmosphere were at rest, undisturbed by any external agency, conduction of heat from one part to another—slow as it is—would, after a sufficient length of time, produce uniform temperature throughout the mass. Further, if the atmosphere consisted of more gases than

one, the pressure and density of each gas would be distributed according to the well-known exponential law of Dalton. Thus if, for a gas, m is the molecular mass, P the partial pressure at a height h, and P_0 that at the ground, then

$$P = P_0 \exp\left(-\frac{mg}{kT}h\right).$$

Such an atmosphere is said to be in isothermal equilibrium.

The terrestrial atmosphere is, however, subject to turbulence due to heating and other causes and is characterized by convective motions. There is therefore a tendency for adiabatic equilibrium to be set up, since conduction in gases is very slow. In the ideal case of adiabatic equilibrium an element of gas transferred from one place to another does not lose or gain any heat by conduction and takes up the requisite temperature and pressure in its new position. The density (ρ) distribution with height in this case is given by

$$\rho = \left[\rho_0^{\gamma-1} - \frac{\gamma-1}{A\gamma}gh\right]^{\frac{1}{\gamma-1}},$$

where A is equal to P_0/ρ_0^{γ} and γ is the ratio of the specific heats of air at constant pressure and at constant volume.

If the troposphere were in ideal adiabatic equilibrium, then the lapse rate—the fall of temperature with height—would have been 9.8C km^{-1}. Also, theoretically, an atmosphere in ideal adiabatic equilibrium has a natural limit. From the expression for ρ it is easily seen that ρ is equal to zero at $h = \gamma P_0/[\rho_0 g(\gamma-1)]$. For the terrestrial atmosphere in ideal adiabatic equilibrium, this limit would be 27.5 km. However, owing to various factors, the troposphere is not in perfect adiabatic equilibrium. The actual lapse rate is only 5C km^{-1}. Also, the height of the troposphere varies between 8 km over polar regions to about 18 km over equatorial regions, instead of averaging 27.5 km. The limit of the adiabatic atmosphere, under actual conditions, cannot be a surface separating a region of perfect vacuum above from a region containing gas molecules below. Because of thermal agitation, molecules from the atmosphere below will constantly be evaporating, as it were, across the separating surface in much the same manner as molecules evaporate from the body of a liquid to the space above it. The region above the natural limit will therefore contain molecules. The atmosphere in adiabatic equilibrium may thus be considered to be capped by a region which we call the *outer atmosphere*. The outer atmosphere is necessarily in isothermal equilibrium.

According to the simple consideration given above, the outer atmosphere need be only a few metres thick. However, account has to be taken of the fact that radiation from the relatively hot gases below is continually reaching and heating this isothermal layer and that in order for the atmospheric gases to be in equilibrium the radiation and the absorption of heat by each element must balance. These considerations lead to the conclusion that the isothermal layer above the adiabatic layer, instead of being a few metres thick, extends to great heights.

Assuming that the atmosphere consists of two layers, the lower in adiabatic and the upper in isothermal equilibrium, and working out the condition for radiative equilibrium, it has been shown that for an atmosphere of uniform constitution, the adiabatic state cannot extend to a height greater than that given by $P = P_0/2$, where P_0 is the pressure at the surface of the earth. For an atmosphere of nonuniform constitution, the adiabatic layer may extend to greater heights. Under certain assumptions concerning the amounts of radiation absorbed by water vapour at different heights, it has been shown that the height of the adiabatic layer cannot exceed that given by $P = P_0/4$ which, for the case of the terrestrial atmosphere, is 10.5 km. This, as indicated above, is roughly the height of the troposphere.

Limit of the Outer Atmosphere. The outer atmosphere, being more or less in isothermal equilibrium, has no natural limit, though the point of minimum density resulting from the balance of the centrifugal and the gravitational forces is sometimes spoken of as the limit of the outer atmosphere. But long before this height is attained (distance 6.6 earth-radii in the equatorial plane) a limit of the outer atmosphere is reached due to the rarity of collisions and the escape of molecules.

We are thus led to enquire into the limit of the outer atmosphere or, more popularly, to ask, Where does the atmosphere end? The question is of considerable importance, since it is closely related to the question of the escape of the atmospheric gases from the gravitational field of the earth.

A limit of the outer atmosphere may be understood from the following considerations. In any region above the surface of the earth, the atmosphere in the ordinary sense of the term can exist only if the molecules in the region are prevented from escaping by collisions with the molecules above. Now, as one goes up, the atmosphere becomes increasingly rarefied and the frequency of collisions between the gas particles becomes smaller and smaller. Ultimately a height is reached where the collisions are so few and far between that a molecule from the denser atmosphere below has little chance of returning to earth by collision with molecules above. The height at which this state of affairs prevails may be said to be the limit of the atmosphere. There is, of course, a transition region of considerable thickness from which a particle, projected upwards, *may* escape without collisions. The mean height of this transition region is variously estimated to lie between 500 and 1000 km [98].

Escape of Atmospheric Gases—The Problem of Helium. This leads to the question of atmospheric gases overcoming the gravitational field and escaping from the earth. The process of escape may best be understood as follows (after Milne [66] and Jones [51]). Let us suppose that an observer ascends upwards from a layer where molecular density is small but appreciable. If the molecules were opaque, the hemispherical sky

above the observer would appear to him absolutely opaque; a line drawn from the observer towards any direction would pass through many molecules one behind another. If the observer continues his ascent, the molecules overhead will gradually thin away and he will reach a level where the molecules overhead will just fill the sky, that is, a line drawn vertically upwards will pass through only one molecule, whereas a line drawn in any other direction will pass through more than one molecule. Mounting still higher, the observer will find his sky overhead gradually clearing up and he will "see" a cone with its axis vertical within which his sky will be clear. It is obvious that this cone will open out with height and the observer will finally "see" the whole sky clear. This cone is called by Milne the *cone of escape* of the molecules. A molecule moving within the solid angle of this cone will have some chance of escaping without a collision. The escape velocity v is given by $v^2 > 2ga^2/r$, where v is the velocity of the molecule, g is the acceleration due to gravity at the level from which the molecule escapes, a is the earth's radius, and r is the distance of the level of escape from the centre of the earth.

Since the mean velocity of the molecules depends upon temperature it may be expected that a considerable quantity of gas, particularly of the light variety, will escape from the upper atmosphere if a high temperature is assumed [49]. The problem of helium in this connection is particularly interesting. Measurements show that the atmosphere near the surface of the earth contains 5×10^{-4} per cent helium by volume. If, however, account is taken of the helium discharge from the earth's crust during the geological ages, the amount of the helium in the atmosphere should be very much more than this. The reason for the scarcity of helium is believed to be that it is continually diffusing upwards and escaping. In order that the escape may be possible a temperature of the order of 1000K has to be assumed near the limit of the atmosphere. We shall see later that considerations of a number of other atmospheric phenomena lead to a similar conclusion.

The Fringe Region or the Exosphere. Beyond the limit of the atmosphere as discussed above, the molecules move freely with the velocity acquired at their last collision in the lower region and, being subject only to the force of gravity, describe elliptic, parabolic, or hyperbolic paths according to the magnitude of their velocities. Escaped molecules whose velocity is not enough to overcome the gravitational pull, will fall back to the atmosphere after describing elliptic paths. These high-flying particles constitute the *fringe region* or the *exosphere* of the atmosphere. The exosphere obviously commences at the level where the semivertical angle of the cone of escape approaches 90°. In the exosphere the particles move without collision in enormous orbits. The heights to which these particles will rise depend on the magnitude and direction of velocity acquired at the last collision and also on g at the point of collision. A few tracks of such particles are shown in Fig. 1. The merging of the atmosphere with interstellar space (average density of matter one particle per cc) is estimated to take place in the region about 2500 km above the surface of the earth [73].

If the atmospheric particles are ionized, the phenomenon of escape becomes complicated. This is because the

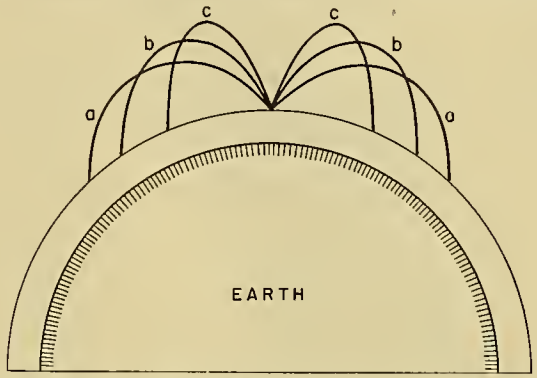


FIG. 1.—Trajectories of a particle projected from 1000-km level and moving freely without collision. Velocity 3.5 km sec^{-1}; initial angle with the vertical—(a) 60°, (b) 45°, (c) 30°.

motion of an ionized particle is profoundly influenced by the terrestrial magnetic field. Thus, although for a neutral particle the critical velocity of escape is independent of latitude, it is not so for an ionized particle [73].

According to Vegard [103], the topmost layer of the atmosphere is an electric double layer formed by electrons, ions, and neutral molecules, the sun being supposed to emit radiation corresponding to soft X rays. The particles in the double layer are disposed of by the earth's magnetic field in a manner similar to that of matter in the solar corona, which is also known to be highly ionized. The terrestrial atmosphere is thus supposed to be topped by a corona—a cloud of particles—similar to the solar corona [103].

According to Hulburt, the particles in the exosphere may be the scattering matter of the zodiacal pyramid which is seen in the evening (or early morning) to rise above the horizon as a faintly luminous cone after the disappearance (or before the appearance) of twilight [48].

In connection with the above two hypotheses of Vegard and of Hulburt it should be mentioned, however, that according to some authors, the particles spend such short time in the transition layer or in the exosphere that they have not much chance of being ionized [98].

SOLAR CONTROL OF THE UPPER ATMOSPHERE

The upper atmospheric regions are under strong solar control due, firstly, to absorption of the whole of the ultraviolet radiation below 2900 A and, secondly, to bombardment by charged particles emitted by the sun. The ultraviolet absorption produces dissociation, allotropic modification, and ionization of the upper atmospheric gases. The bombardment by charged particles—which is concentrated round the magnetic axis poles—produces auroral displays associated with optical excitation and ionization of the upper atmospheric gases. The effects of the different parts of the ultraviolet solar spectrum are presented in Table I.

It is to be noted that if the sun is assumed to be radiating like a black body, then the number of emitted quanta decreases rapidly with the decrease of wave length. As such, the available number of quanta may not be sufficient to produce the observed effects in the region of extreme ultraviolet. It is therefore believed that the sun must be sending out continuously (with occasional outbursts) line radiations in the extreme ultraviolet, for example, principal series of H, He and He^+. Besides, there may also be ultraviolet radiation

Illumination of the Upper Atmosphere by Solar Rays. In connection with the solar control of the upper atmosphere it is important to remember that the hours at which the sun's rays first strike in the morning or disappear in the evening from the high atmospheric regions are quite different from those at the surface of the earth. For example, the high atmosphere above 100 km as far from polar regions as lat. 55°, is illuminated at midnight by solar rays during high summer. The method of calculating the hours of sunrise and sunset at

Table I. Effects of Solar Ultraviolet Radiation on Upper Atmospheric Gases

Spectral region	Reaction	Remarks
3000–2100 A (Hartley absorption bands of O_3)	$O_3 + h\nu \rightarrow O_2 + O^*$ (excited)	Very strong absorption by ozone (50–60 km).
1925–1760 A (Runge-Schumann absorption bands)	$O_2 + h\nu \rightarrow O_2$ (excited) $O_2^* + O_2 \rightarrow O + O_3$ $O + O_2 + M \rightarrow O_3 + M$ (Note: The last reaction is a three-body collision process, in which M is the so-called *third body* which takes away the excess of momentum and energy.)	Comparatively weak absorption. Production of ozone.
1751–1200 A (Runge-Schumann continuum)	$O_2 + h\nu \rightarrow O + O^*$ (excited)	Very strong absorption. Dissociation of O_2 above 80 km.
1012–910 A	$O_2 + h\nu \rightarrow O_2^+$ (normal) $+ e$	Weak absorption. First ionization potential of O_2. Production of D-region (?) (50–80 km).
910–795 A	$O + h\nu \rightarrow O^+ + e$	Very strong absorption. Ionization of O. Production of F_1- and F_2-regions (?) (200 km upwards).
795–755 A	$N_2 + h\nu \rightarrow N_2^+$ (normal) $+ e$	Comparatively weak absorption. First ionization potential of N_2. Production of E_2-region (?) (140–160 km).
744–661 A	$O_2 + h\nu \rightarrow O_2^+$ (excited) $+ e$	Strong absorption. Second ionization potential of O_2. Production of E_1-region in the transition region $O_2 \rightarrow O + O$ (90–120 km).
661–585 A	$N_2 + h\nu \rightarrow N_2^+$ (excited) $+ e$	Very strong absorption. Second ionization potential of N_2.

and soft X rays coming out of the corona which is regarded as having a very high temperature of the order of some million degrees.

The phenomenon of radio fade-out strikingly illustrates the solar control of the upper atmosphere. During an intense solar flare, when bright spots of hydrogen light appear on the sun, the magnetic disturbances and the ionization in the absorbing D-region of the ionosphere both show an abnormal and simultaneous increase. This causes total stoppage of radio traffic over the sunlit portion of the earth [64].

different atmospheric levels and the results obtained for a few typical cases are given below.

The shadow cast by the earth in space is of cylindrical shape. We have from Fig. 2,

$$H = a\left(\frac{1}{\cos\theta} - 1\right),$$

where H is the height at which the cylinder cuts the zenith, a is the radius of the earth, and θ is the depression of the sun below the horizon.

The hour angle h of the sun at its rising at any height H may be obtained as follows. We have

$$\cos Z = \sin \phi \sin \delta + \cos \phi \cos \delta \cos h,$$

where Z = zenith distance of the sun, δ = declination of the sun, and ϕ = latitude of the place of observation. From Fig. 2, $Z = 90^\circ + \theta$.

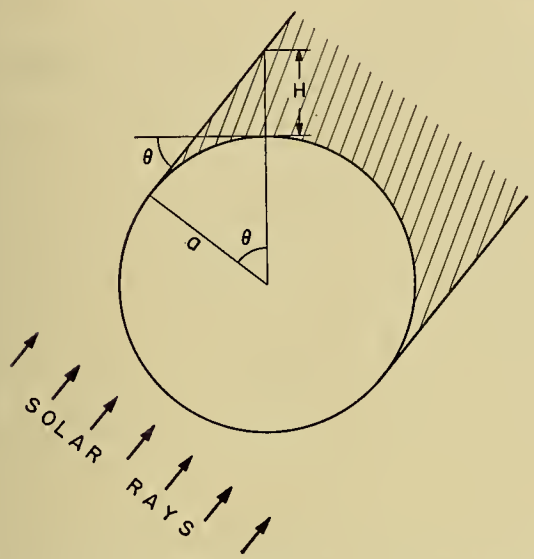


FIG. 2.—The height H at which the cylindrical shadow cast by the earth cuts the zenith is given by $H = a\left(\frac{1}{\cos\theta} - 1\right)$.

The effect of refraction—which accelerates the time of rising—has not been taken into account in the equation above. The true zenith distance of the sun's centre when it rises at height H above the ground is given by

$$Z = 90^\circ + \theta + 50'.$$

Here $34'$ has been allowed for horizontal refraction and $16'$ for the semidiameter of the sun.

Now, at actual noon the apparent time is 12^h and the mean time is $12^h + \epsilon$, where ϵ is the equation of time. Hence the mean time t of sunrise at height H above the ground is given by

$$t = 12^h + \epsilon - h.$$

Similarly, the hour of sunset at H is given by

$$t = 12^h + \epsilon + h.$$

The values of ϵ and δ may be obtained from the *Nautical Almanac*. Hence the hours of sunrise and sunset at different atmospheric levels may be calculated with the help of these equations.

Curves in Figs. 3*a* and 3*b* delineate (after Ghosh [40]) the variation of the hour of sunrise with height for the whole year at intervals of about a fortnight for the latitude of Calcutta (22° 32′ 48″ N). Figure 4 depicts (after Bartels [10]) the height above which the atmosphere is illuminated by solar rays at midnight during the course of the year in the latitudes 40° to 90°.

Useful curves depicting the hours of sunrise at different latitudes on the surface of the earth and also at a given set of heights (*e.g.*, 50 km, 180 km, 250 km, and 500 km) are also given by Bartels [10].

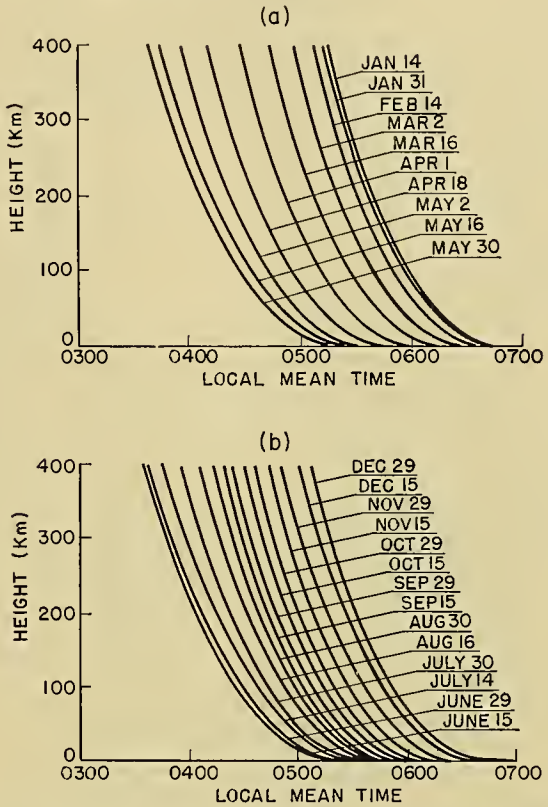


FIG. 3.—Illustrating how the hour of sunrise at different heights in the upper atmosphere changes in the course of the year at the latitude of Calcutta. (*After Ghosh.*)

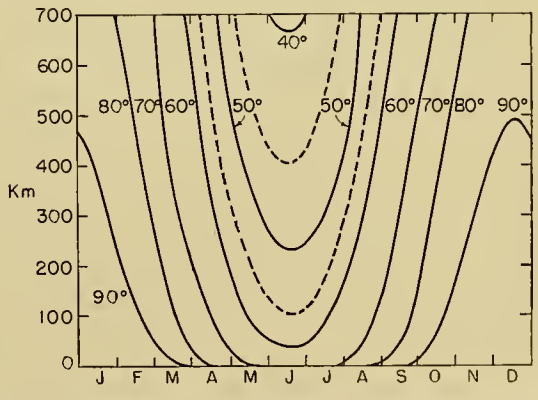


FIG. 4.—Illustrating the heights above which the atmosphere is illuminated at midnight in the course of the year at different latitudes. (*After Bartels.*)

A PICTORIAL REPRESENTATION OF THE PHYSICAL FEATURES OF THE UPPER ATMOSPHERE

In Fig. 5 an attempt has been made to depict the physical features of the upper atmosphere and some of the phenomena occurring there.

Starting from the tropopause we note that the region

from 20 km to 40 km is rich in atmospheric ozone. The total quantity of ozone is small—only about 0.25 cm thick if reduced to standard temperature and pressure—but it suffices to cut off all solar radiation in the near ultraviolet below 2800 A. In the region round 80 km noctilucent clouds occur. Meteors appear and disappear most frequently in the region from 50 to 150 km.

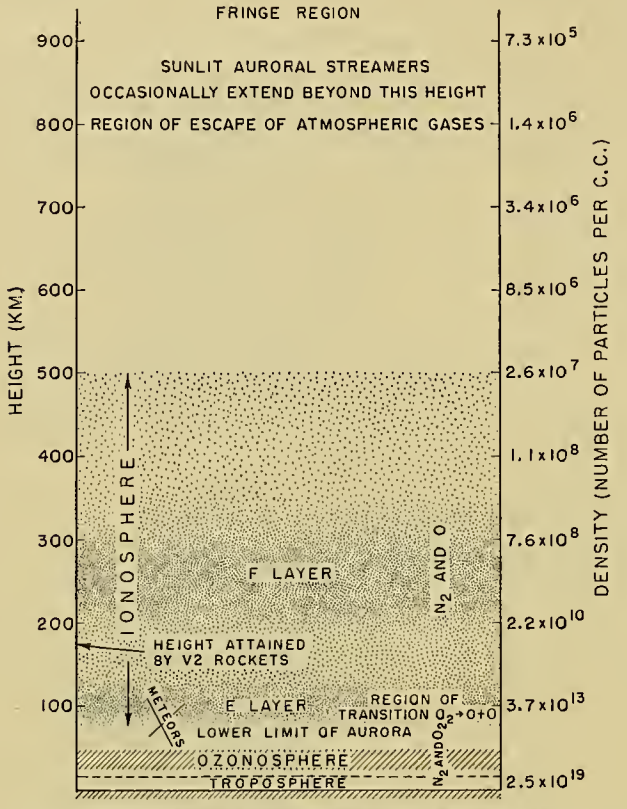


Fig. 5.—Illustrating some of the physical features of the upper atmospheric regions. The ionospheric regions are shown by shading with dots, the depth of shading roughly indicating the relative intensity of ionization.

The regions above 70 km are ionized by the solar extreme-ultraviolet rays and are collectively called the *ionosphere*. The region from 80 to 130 km is the region of transition from an atmosphere consisting of N_2 and O_2 to one of N_2 and O, due to the dissociative action of solar ultraviolet rays. The region from 80 to 120 km is the region of the most frequent occurrence of aurorae. But auroral rays are sometimes known to extend up to a height of 1000 km and beyond.

Atmospheric gas particles which are not ionized by solar ultraviolet rays escape from the region estimated to lie between 500 and 1000 km. Such particles as escape from the atmosphere but are unable to overcome the gravitational pull travel in closed orbits and form the exosphere or the fringe region of the atmosphere.

The physical features depicted in Fig. 5 are, of course, only illustrative and should not be taken too literally. This remark is particularly applicable to the density distribution above 100 km. Here, because of the uncertainty of the temperature distribution and of the mean molecular mass, the density distribution is uncertain. For example, if it is assumed that the region above 100 km is in diffusive equilibrium, the higher regions would consist almost entirely of oxygen atoms. But the spectrum of auroral rays shows that, up to the highest limits, the intensity of the negative bands due to ionized nitrogen molecules vies with that of the lines of atomic oxygen.

METHODS OF EXPLORING THE UPPER ATMOSPHERIC REGIONS

The methods of exploring the upper atmospheric regions may be classified under two heads—direct methods and indirect methods.

Direct Methods. Under this caption we include those methods in which the physical agency employed for exploration is under the control of the investigator. A straightforward direct method is to send up craft carrying recording apparatus which may register temperature, pressure, humidity, wind, and any other measurable physical quantities. A variation of this method is the so-called *radiosonde method* in which the craft carries a small radio transmitter which automatically sends out signals of the recorded data. The great advantage of radiosondes is that instantaneous knowledge of the data is gained and it is not necessary to wait until the craft comes down and the registering apparatus is collected.

Until recently the craft used for this purpose were generally sounding balloons filled with hydrogen gas. (Aircraft are also used for regions up to about 14 km.) The height reached by such balloons seldom exceeds 30 km. But a new departure was made in 1946 by utilising the V-2 rockets devised during the war. These reach enormous heights (up to 180 km) and even the very first trial flights have yielded many important results.

A few words may be said here of the technique of this latest method of sounding the upper atmosphere. (General accounts of it are to be found in articles by Krause [52] and by Newell [80]; an excellent summary is also given by Sheppard [96].) The war head which contains the recording instruments bursts towards the end of the flight and the instruments are landed by means of a parachute. The height and the velocity of the rocket are tracked by radar throughout its flight. For the flight at White Sands, New Mexico, on March 7, 1947, the highest velocity, 1600 m sec^{-1}, was attained at the height of 127 km, 80 seconds after launching the rocket. The highest altitude attained was nearly 180 km. Measurement of the velocity is important because it is necessary for evaluation of the temperature. Much of the recorded data was also telemetered to the ground during the flight. By various instruments and devices pressure down to 10^{-5} mm was measured. The temperature was derived from the pressure-height curve and/or from the ratio of the ram pressure (the pressure at the stagnation point on the nose of the rocket) to the static pressure. The probable errors of measurement are still

rather large, averaging ± 20C. The recorded data will be discussed in the next section.

Another agency for direct study is pressure waves produced by explosion. These waves proceeding upwards may be bent down if they meet a region of high temperature where the phase velocity is greater. Records of the downcoming wave yield information regarding the density and temperature distribution.

An analogous method is to use radio waves instead of sound waves. These waves penetrate beyond the highest limits attained by V-2 rockets and when they come down, "reflected" by the ionized regions of the upper atmosphere, they carry with them indelible messages regarding the physical conditions of such regions, which are no less useful than those recorded by instruments carried by sounding balloons or rockets.

Indirect Methods. These consist in studying critically such of the geophysical phenomena occurring near the surface of the earth or in the upper atmosphere as are known to have bearings on the physical state of the upper atmosphere. These phenomena include aurorae, lights from the night sky, meteoric flashes, noctilucent clouds, oscillations of barometric pressure, terrestrial magnetic variations, and spectra of direct or scattered sunlight.

The various direct and indirect methods that have been employed in the study of the upper atmospheric regions and the main results obtained therefrom are shown in Tables II and III. Short accounts of the methods together with a brief survey of the contemporary state of ourknowledge of the physical state of the upper atmosphere will be given in the next section.

TABLE II. DIRECT METHODS OF STUDYING THE UPPER ATMOSPHERE

Method	Region explored	Atmospheric conditions inferred from the study
Sounding balloon	Up to 30 km	Troposphere is in quasi-adiabatic equilibrium; in the stratosphere the temperature distribution is nearly constant with height; information about composition and wind systems.
Smoke shell	Up to 30 km	Existence of seasonal wind systems.
V-2 rocket	Up to 120 km	Generally confirms results on temperature and density distribution as inferred from the indirect methods of study; the chemical composition of the atmosphere at 70-km level is practically the same as that in the troposphere; solar spectrum is extended beyond the ozone absorption limit; Mg^+ doublet (2802 A) obtained in emission.
Sound explosion	35–60 km	Rise of temperature in the middle atmosphere; existence of wind systems.
Radio-wave exploration	70–500 km	Atmospheric constituents from 70 km upwards are ionized; measurements of scale height H, intensity of magnetic field, recombination coefficient. High temperature (approx. 1000K) in the region 250 km. "Bursts" of ionization are produced along meteor trails.

TABLE III. INDIRECT METHODS OF STUDYING THE UPPER ATMOSPHERE

Phenomena studied	Region explored	Atmospheric conditions inferred from the study
Meteoric phenomena	40–150 km	Rise of temperature in the middle atmosphere and correspondingly greater density; drop in temperature at 80-km level; existence of seasonal winds.
Ultraviolet spectrum of direct or scattered sunlight	20–60 km	Contains ozone with maximum concentration at about 25-km level; thickness of the ozone layer reduced to S.T.P. is only about 0.25 mm.
Noctilucent clouds	70–90 km	High-velocity wind and low temperature (approx. 200K).
Barometric oscillations	50–400 km	Tidal motions in the upper atmosphere; temperature rise in the middle atmosphere with a cold top.
Terrestrial magnetic variations	70–100 km	High electrical conductivity and world-wide electric current systems.
Night-sky luminescence	60–500 km	Sodium atoms in the middle atmosphere (60–80 km); atomic oxygen in the higher regions above 100 km; nitrogen molecules are ionized by solar rays.
Aurorae	80–1000 km	Entry of high-speed charged particles; atomic oxygen and atomic nitrogen present.

A SURVEY OF THE CONTEMPORARY STATE OF OUR KNOWLEDGE OF THE UPPER ATMOSPHERE

Composition. Near the surface of the earth the atmosphere consists of nearly 99 per cent by volume of the gases nitrogen and oxygen. Next in importance are argon 0.93, carbon dioxide 0.03, neon 1.8×10^{-3}, and helium 0.5×10^{-3} per cents by volume [26].

Chemical analysis by laboratory methods of samples of air collected in stratosphere flights and by pilot balloons has shown that this composition is maintained up to a height of about 29 km [84]. Contemporary experiments made with samples of air collected by V-2 rockets show that practically speaking the same composition is maintained up to 70 km [23]. This is very satisfactory, because from various considerations the same conclusion had been arrived at long before the observations with V-2 rockets. (Ozone is present in the middle atmosphere in appreciable quantities. This will be discussed later.)

The atmosphere above 80 km begins to change in composition because of the dissociative action of solar ultraviolet rays on molecular oxygen ($\lambda < 1751$ A). The region 80–130 km is the region of transition from

an atmosphere consisting of N_2 and O_2 below, to one of N_2 and O above [25, 57, 88, 90, 108]. Since low pressure is conducive to diffusive separation, it might be expected that atomic oxygen would predominate in the highest regions of the atmosphere [75]. There is, however, no evidence of it. Spectra of auroral streamers extending up to 1000 km and beyond show lines due to atomic oxygen and bands due to N_2^+ with almost equal intensity. Presence of atomic nitrogen has also been reported in low-latitude aurorae [32, 41], but whether atomic nitrogen is as generally distributed as atomic oxygen is not yet known (see section on aurorae below).

Temperature Distribution. The temperature distribution in the troposphere and the lower stratospheric regions is now well known from direct observations with the help of sounding balloons and smoke shells [50]. There is a falling temperature (lapse rate about 5C km^{-1}) in the troposphere, the depth of which varies between 8 and 18 km. Above the troposphere the temperature is nearly constant. For the higher regions up to 120 km, we now have results of direct observation thanks to successful V-2 rocket flights. These observations are now confined only to isolated places (in the United States) but it is hoped that in the near future reliable data based on rocket observations will also be available for the different latitudes at different hours of the day and night and in different seasons. A complete picture of the world distribution of temperature, at least up to Region E of the ionosphere, will thus be available. In Fig. 6 the temperature distribution obtained from the rocket flight at White Sands, New Mexico, on March 7, 1947 is depicted [15]. For comparison, the distributions obtained from results of indirect observations (abnormal sound propagation, meteoric flashes, and noctilucent clouds), are also given.

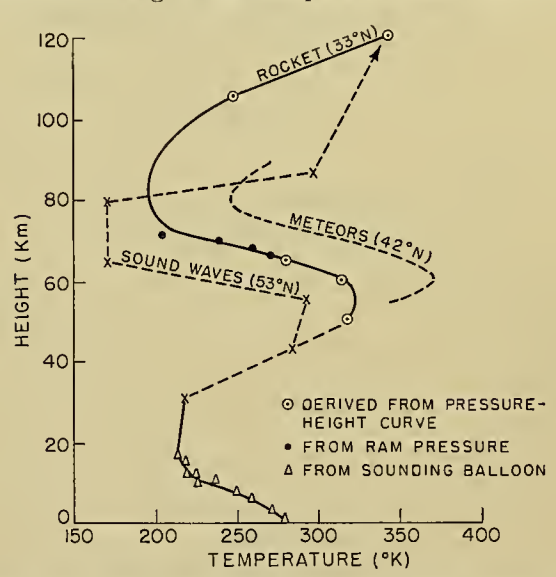


FIG. 6.—Temperature distribution with height. The rocket data are from the flight on March 7, 1947 at White Sands, New Mexico. The sounding balloon data were obtained within an hour of the rocket flight. For comparison, temperature distributions as deduced from abnormal sound propagation experiments and from meteoric data are also given after Sheppard [96].

It will be seen that the trend of the rocket-observation curve is the same as the general trend of the results from indirect observations.

Some remarks about the origin of this peculiar temperature distribution in the middle and in the upper atmosphere and the indirect methods of investigation by which it had been surmised long before may be made here. The rise of temperature in the middle atmospheric region (30–50 km) is due to absorption by ozone [43, 87]. The existence of any marked temperature variation in the stratosphere was, however, not suspected until the famous work of Lindemann and Dobson on meteors [54]. These authors developed a theory of the appearance and disappearance of meteors and applied it to determine the density distribution in the middle atmosphere. Their findings could be reconciled with observed meteoric data only if it was assumed that an isothermal region of high temperature (about 100C) existed above 55 km.

A later interpretation of meteoric data by F. L. Whipple [106] based on a different theory of meteors (first suggested by Sparrow [97] and subsequently developed by Öpik [83]) also confirms these results. The meteoric data, according to Whipple, can best be reconciled if it is assumed that there is a flat temperature maximum of about 375K near the 60-km level (as in the case of the Lindemann-Dobson theory), a rapid drop to 250K near 80 km (as in the Taylor-Pekeris theory, see below), and a constant or slowly rising temperature up to about 110 km. A seasonal variation is also indicated. The upper atmosphere, under average midsummer temperatures, is raised 5.3 ± 1 km above its height under average midwinter temperatures.

Existence of high temperature in the middle atmosphere, as indicated by the study of meteors, is also confirmed by the study of abnormal propagation of sound waves. The sound of an explosion or the firing of a cannon can be heard not only in the neighbourhood of the source of sound but also at a greater distance, in an area separated from the audible zone by a silent zone. The phenomenon is caused by refraction of the sound waves by the region of high temperature in the middle atmosphere. Since the trajectory followed by the sound ray in the region of high temperature depends upon the temperature gradient of the region, it is possible, from a systematic study of the abnormally propagated waves (the source being an artificial explosion), to deduce the gradient as well as the temperature. Such studies have been made in Germany, in England, in the United States, and recently in India [62].

In the region from 60 to 80 km there is no strong absorption of solar radiation. (Below 60 km there is absorption by ozone; above 80 km there are absorptions leading to dissociation of molecular oxygen and ionization of the atmospheric gases.) Hence, a falling temperature is to be expected in this region. The rocket curve shows this clearly, the temperature dropping to about 180K near the 80-km level. A drop was also inferred from indirect evidence long before rocket experiments. For example, the so-called noctilucent clouds which are observed in the region around 80 km are

believed to be composed of ice crystals. According to some authorities the presence of the ice crystals indicates a temperature of 160K in this region [60]. Again, radio measurements show that the scale height (kT/mg) for this region has a value which corresponds to a temperature of about 200K [19]. Meteor observations by Whipple, however, indicate a drop to only 250K near 80 km (see above). Work on atmospheric tides [104] also confirms the general trend of the drop in temperature in this region.

Above 80 km a rise in temperature would be expected because of strong absorption by O_2 molecules (1751 A). This is fully supported by the rocket observations. A rise from 180K to 335K is found between the levels 80–120 km.

Beyond 120 km there are several indirect evidences indicating that the rising gradient is maintained and a temperature near 1000K to 1200K is attained at the height of Region F_2. These evidences are (1) increased scale height in Region F_2 as compared to that in Region E [3], (2) escape of helium from the outer atmosphere, which demands a temperature of above 1000K [49, 98], and (3) large width of the green line of atomic oxygen in the spectrum of night-sky emission [8]. (It should be mentioned, however, that in regard to the last-named evidence more recent measurements of the green oxygen line [102] show that the previous estimation of temperature was too large.)

It appears that there is considerable variation, both spatial and temporal, in the value of this high temperature within the ionospheric Regions E, F_1, and F_2, 80 to 400 km [95]. In the F_2-layer the temperature is found to vary (both spatially and temporally) from 100K to 1000K. In lower levels the range is smaller. In the stratum from 200 to 300 km, in the noon meridian, two centres of high temperature are found, one from 30°N to 50°N and the other at 35°S. The former is less peaked than the latter. It should be mentioned that these temperatures have been derived by Seaton [95] from the values of the recombination coefficient computed from ionospheric data for January 1947, collected at eighteen stations situated between 71°N and 43°S. Existence of such thermal patterns implies the existence of quite strong wind systems in the ionospheric regions (see below).

The existence of a high temperature in the uppermost regions of the atmosphere can also be inferred on strong theoretical grounds. The primary effect of the absorption of solar radiation might be dissociation, ionization, or excitation of the constituent gases of the atmosphere; but, like all other forms of energy, the absorbed energy must ultimately degrade into thermal energy of molecular agitation, causing a rise of temperature. (The applicability of the concept of temperature for the very high regions of the atmosphere has been discussed by the author [72, pp. 507–510].)

Study of Meteors. It has already been mentioned that the existence of a region of high temperature in the middle atmosphere was first inferred from a study of the heights of appearance and disappearance of meteors. Winds in the upper atmosphere have also been studied from measurements of the drift and distortion of long-lived meteor trails [47, 82, 99]. Study of meteors, in fact, provides a valuable means of investigating the physical characteristics of upper atmospheric regions. It is, therefore, very satisfactory that a new technique—the radar technique—has been adapted for this study, and developed in the research laboratories of different countries [46, 55, 58, 63]. Not only is the ionization produced by meteoric impacts being systematically studied, but methods have also been developed for measuring the velocity [34] and other characteristics of meteors (*e.g.*, height, range, and radiant point). Accurate determination of the meteor velocity is very important in the application of the theory of meteors in deducing various upper atmospheric data.

Ozonosphere—Origin of High Temperature in the Middle Atmosphere. The high temperature of the middle atmosphere mentioned above is a consequence of absorption of solar radiation by atmospheric ozone. Because of this absorption the solar spectrum ends abruptly at about 2900 A. The absorption bands responsible for this are known as Hartley absorption bands.

The ozone content of the atmosphere may be determined by the measurement of the variation of the intensity of solar radiation near the ultraviolet end of the spectrum as the zenith distance of the sun changes [20]. A more accurate method is spectrophotometric study of sunlight scattered from the zenith sky. The intensities of two spectral regions near the absorption edge, one of which is rather strongly and the other weakly absorbed, are compared when the sun is setting [41]. This makes possible, with the help of the so-called Umkehr effect, an approximate estimate of the distribution of ozone with height. Contemporary improvements in the technique of ozone measurement, utilising photoelectric multiplier tubes should be mentioned. At the 1948 Oslo meeting of the Union of Geodesy and Geophysics it was claimed that with the improved technique, the scattered light from the moon and the stars sufficed for making measurements of the ozone content at night.

It has been found that the proportion of ozone to air by volume is a maximum at a height of about 35 km. The atmospheric ozone near the region of 50 km acts as an enormous reservoir of heat. This height is much above the centre of gravity of ozone because the solar ultraviolet radiation responsible for heating is almost completely absorbed in the top layer [87].

The ozone content shows a diurnal variation, being greater at night or at least never less than that during the day. There is also a seasonal variation; the maximum occurs in spring and the minimum in autumn in both hemispheres. This annual variation is greatest in the high latitudes and least near the equator.

It is to be noted that solar radiation is responsible for both production and destruction of ozone. It is produced by absorption by molecular oxygen in the range of the so-called Runge-Schumann absorption bands (1760–1925 A) when excited O_2 molecules are

produced. It is destroyed by absorption in the near ultraviolet in the Hartley bands (2100–3000 A).

Observations show that the variation of the atmospheric ozone content has little if any association with the 11-year solar cycle. It appears, however, that there are 27- and 15.5-day periods of variation with small amplitudes.

Ionization in the Upper Atmosphere—Radio Exploration. Atmospheric constituents from 70 km upwards are more or less ionized by the action of the extreme solar ultraviolet rays. Under certain ideal conditions each atmospheric constituent ionized may be supposed to produce a distinct layer or region of ionization. Such a region of ionization is called a *Chapman region* [24]. There are several ionized regions and these are known collectively as the ionosphere. The two main regions of maximum ionization, Region E and Region F, are at average heights of 100 km and 275 km respectively. The average maximum ionization densities of these two regions, in the epoch midway between the maximum and minimum of solar activity, are of the orders 10^5 and 10^6 electrons cm^{-3}, respectively. During daytime, Region F splits up into two regions, F_1 and F_2. A subsidiary Region E_2 (above E, which is sometimes called E_1) also appears during the daytime. The average heights of the maximum ionization of F_1 and E_2 are 200 km and 140 km, respectively. The ionization below 80 km (Region D), present during the daytime, causes absorption of radio waves of medium length.

As may be expected, the ionization densities of the various regions vary with the hour of the day, the season of the year, and also with the solar cycle. It should be noticed, however, that while Regions E and F_1 follow, in the main, the simple $\sqrt{\cos \chi}$ law (*i.e.*, intensity of ionization is proportional to $\sqrt{\cos \chi}$, where χ is the zenith angle of the sun), Region F_2 refuses to do so. Region F_2 is, in fact, notorious for its erratic behaviour.

Besides the regions mentioned above, which are more or less permanent features of the ionosphere, mention should also be made of what is known as *sporadic E*. Part, at least, of sporadic E is ascribed to ionization produced by meteoric impacts [55].

The knowledge of upper atmospheric ionization summarized above has been gained mainly through exploration by radio waves. The radio wave is now the most powerful experimental tool for studying upper atmospheric ionization. The most important method of radio exploration is as follows: Packets or "pulses" of radio waves are sent upwards. These are generally scattered if they meet a region of ionized atmosphere. If the ionized region is extensive—of dimensions much larger than the wave length—then, depending upon the frequency, the wave packet may be totally or partially reflected. Or it may penetrate through, or be absorbed by, the ionized region.

The technique of the method has now been greatly perfected. Records of ionosphere characteristics are now kept in many stations of the world as regularly as weather and magnetic data. Predictions of ionospheric conditions, well in advance, for determining the maximum usable frequencies for communication over given distances, are also made regularly by many ionosphere stations.

The other physical properties of the upper atmospheric regions which have been determined from the ionospheric studies are the following:

1. The absorption phenomena of radio waves enable us to estimate the collisional frequencies and the molecular densities in the different regions. Thus, it is found that for Region E, the pressure is of the order 10^{-3} mm and for Region F 10^{-5} mm.

2. Analysis of the records of diurnal variations of the heights and the maximum ionization densities of the Regions E and F_2 has confirmed the existence of tidal motions in the high atmosphere. Observations on absorption in the lowermost ionospheric region—the D-region—has revealed the existence of lunar tidal oscillation effects in this region also [5].

3. The study of the magnetic splitting of the radio waves in the ionosphere affords a means of estimating the intensity of the earth's magnetic field at great heights above the surface of the earth.

It has now been established that the different ionospheric regions are produced by the ionization of the different atmospheric constituents at different levels. According to one hypothesis [16], Region F_2 is produced by ionization of atomic oxygen, Region F_1 by ionization of N_2, Region E by ionization of O_2 at second ionization potential, and Region D by the ionization of the same molecule at its first ionization potential [74]. According to another hypothesis, Region F_1 is produced by ionization of atomic oxygen in the normal manner and Region F_2 by splitting up of this region into two regions [78]. According to still another hypothesis, some of the ionized regions may be produced by radiation from the solar corona [107]. In regard to Region E it may be noted that for a long time it had been difficult to explain theoretically the location of its height of maximum ionization, because the height at which it was expected was much above the observed height. The difficulty has been reconciled by the hypothesis that the E-layer is located in the region (90–110 km) where the density of O_2 molecules diminishes rapidly with height on account of the dissociative action of solar ultraviolet rays [67]. Further, it is believed that in this region the O_2 molecules are pre-ionized, rather than directly ionized, by solar radiation of appropriate wave length [81].

Luminescence of Upper Atmospheric Regions. On dark moonless nights, the portions of the sky devoid of stars are found to emit a faint light. Part of this light is due to starlight scattered by the atmosphere and part due to other sources. But a certain proportion, about 40 per cent, has been found to be due to the self-luminescence of the upper atmospheric gases. Spectroscopic study of the night-sky radiation shows the presence in the upper atmosphere, besides O_2 and N_2, of O and Na atoms and of H_2O and OH.

The most prominent lines and bands in the night-sky spectrum in the visible region are lines due to O atoms and bands due to N_2 molecules. The luminescent region emitting these lines and bands may possibly be identified with Region F of the ionosphere (200–400 km). It

has been found that the night-sky brightness due to these lines and bands fluctuates erratically when the ionospheric conditions in Region F do so [69]. The reaction which excites these spectra is most probably due to the mutual neutralization of N_2^+ and O^- ions [67], thus,

$$N_2^+ + O^- \rightarrow N_2 \text{ (excited)} + O \text{ (excited)}.$$

In the infrared region there are strong emissions due to *OH* bands [65]. The region from which these are emitted is still uncertain.

There is a luminescent layer (60–80 km) emitting sodium *D* lines [12] and also possibly one emitting O_2 bands at the height of Region E.

The intensity of the night-sky radiations varies with the season of the year and with the solar cycle and shows how upper atmospheric conditions are controlled by solar radiation and by the corpuscular emission from the sun.

Winds in the Upper Atmosphere—Atmospheric Tides. There is evidence that up to the height of Region F the atmosphere is subject to winds, caused partly by temperature gradient and partly by tidal effects. Tentative curves depicting the variation of wind with height, for summer and winter in the middle latitudes, are drawn after Sheppard in Fig. 7 [96]. The data for the region

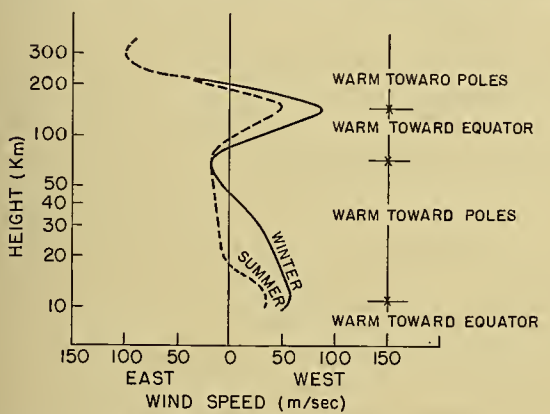


FIG. 7.—Probable variation of wind (east and west components) with height in middle latitudes for summer and winter. The height is given in logarithmic scale as this enables one to infer directly the approximate mass flow. The meridional temperature gradient for quasi-geostrophic winds is shown on the right. (*After Sheppard* [96].)

around 30 km are from sounding-balloon [44] and smoke-shell observations [50]. For higher regions, the evidence is from observations on noctilucent clouds and from meteor trails. For the latter (in the region of 120 km) an optical technique has been developed for measurement of wind from the drift and distortion of long-lived meteor trails by Störmer in Norway [99], Olivier in the United States [82], and Hoffmeister in Germany and Southwest Africa [47]. Radio observations on the movements of what are called *ion clouds* also confirm the existence of winds in this and in still higher regions [14, 36, 77, 79].

Regarding the tidal effect it is to be mentioned that there exist in the atmosphere, just as in the ocean, tidal motions due to the pulls of the sun and the moon. These tidal effects have been observed up to the highest regions of the ionosphere—the F-region (250–400 km). A direct effect of the tidal oscillations is seen in the rhythmic variation of the barometric pressure. An indirect effect is the rhythmic variation of the terrestrial magnetic elements on a magnetically quiet day. The amplitudes of these variations are small—only about a few thousandths of the total value—but they are singularly persistent.

The tide-raising force of the moon is about 2.5 times as strong as that of the sun. But, contrary to expectation, it is found that the solar tidal effect is very much more prominent than the lunar one. The origin of this anomaly is traced to a resonance effect. It was first pointed out by Taylor and Pekeris that, because of the peculiar temperature distribution in the middle atmosphere (a temperature rise in the neighbourhood of 50 km followed by a cold top), the atmosphere has a mode of oscillation with a 12-hour period [86, 100]. Hence the solar tidal effects are greatly enhanced by resonance.

The ionized regions of the upper atmosphere are also necessarily subject to the tidal oscillations. The oscillations of the lowest of these regions are responsible for the quiet-day magnetic variations as follows: As the conducting ionized region, in its horizontal tidal motions, cuts the magnetic lines of the earth, emf's are developed in these regions. These emf's produce world-wide electric current systems in the upper atmosphere. The magnetic field of these current systems produces the quiet-day variations of the terrestrial magnetic elements. It is believed that the most probable location of these current systems is the lower regions of the ionosphere (D- and E-regions) where the mean free paths of electrons and ions are small [64].

Aurorae—Entry of Fast Charged Particles into the Upper Atmospheric Regions. The luminescence of the aurora is due to the excitation of air molecules by collision with the fast charged particles emanated from the sun and incident on the upper atmosphere. Besides the lines and bands observed in the night-sky light (with different intensity distribution) a special feature of the auroral spectrum is the great intensity of the so-called negative bands of nitrogen due to N_2^+ ions. It also appears that nitrogen atoms are present in the aurora. Identification of forbidden lines of atomic nitrogen in the auroral spectrum has been reported by more than one worker. Bernard [13] claims to have identified the forbidden ultraviolet doublet of N, $^2P \rightarrow {}^4S$ (3466 A) while Dufay, Gauzit, and Tcheng Mao-lin [33] announce that they have observed the forbidden green doublet, $^2D \rightarrow {}^4S$ (5199 A) in low-latitude aurorae. It should be mentioned that according to quantum-mechanical calculation by Pasternack [85], the lifetime of the excited state for 5199 A is 10 hours. According to Götz [42], the strength of this radiation in the case of low-latitude aurorae (which generally lie at greater heights) may continue undiminished for hours after the aurora ends. From the intensity measurement made on a particular aurora, the concentration of excited N atoms, at an altitude somewhat above 200 km, was estimated to be of the order 10^6 cm^{-3}.

As is to be expected, auroral activity and magnetic disturbances are closely associated with solar phenomena. For instance, both auroral and magnetic activity show remarkable correlation with sunspot numbers [28]. Also both have seasonal variation—maxima in March and October and minima in June and January [27].

Weather and the Upper Atmosphere. It may seem surprising that weather conditions in the tropospheric regions should have any association with the physical conditions of the atmosphere tens or even hundreds of kilometres above the surface of the earth. Nevertheless, evidence of such associations (some of them yet unexplained) has been obtained.

It has been found that the ozone content in the middle atmosphere is markedly related to weather conditions in the troposphere [31, 32]. Almost without exception the ozone value is found to be high with cyclonic systems (low pressure) and low with anticyclonic systems (high pressure). Periods of fine and settled weather have been observed to be associated with relatively low ozone value [101].

Atmospheric densities at heights of 75 km (as deduced from observations on meteors) show a marked seasonal variation, and this variation is correlated with the mean ground temperature [105], a maximum density being indicated at the time of maximum ground temperature.

Association between Region E ionization and thunderstorms has been reported by some observers [6, 17, 92, 93]. Evidence has also been obtained of association of weather conditions in the tropospheric regions with variations of ionization densities in the ionospheric regions [60]. For instance, it has been observed that the variations of the minimum heights of Region F and of the Region E tend to follow the variations of the barometric height. Again, day-to-day variations of noon ionization of Region E seem to correspond to the day-to-day variations of barometric height. Some remarkable correlations of the variation of Region F_2 ionization and cyclonic conditions near the ground have also been reported from Australia [9]. Other similar correlations, scarcely less remarkable, have been reported from Shanghai, China [38]. It has been found that there is close relationship between the occurrences of the E, F_1, and F_2 echoes (when the ionosphere apparatus is set to the fixed frequency, 6 mc sec^{-1}, the mean E-critical frequency of the place) and the movements of the three main air masses—polar, maritime, and equatorial—which cause weather all over the world. It is claimed that these observations can be very successfully used for weather prediction.

Extension of the Atmosphere to Great Heights. Streamers of aurorae of the ray type extend to great heights (800 to 1100 km) into the sunlit portion of the upper atmosphere. Spectroscopic study of the auroral light shows that even up to such heights the atmosphere consists of atomic oxygen and molecular nitrogen.

SOME UNSOLVED PROBLEMS

The brief account of the contemporary state of our knowledge of the upper atmosphere given above shows that in spite of the many difficulties, a considerable amount of information has already been gained. However, there are still gaps in our knowledge. This is due not only to the inaccessibility of the regions concerned, but also to the lack of many fundamental laboratory data of the atmospheric constituents. But perhaps the greatest obstacle has been our imperfect knowledge regarding the nature of extreme ultraviolet radiation from the sun. Let us hope that this gap will soon be filled up by observations made with the help of V-2 rockets.

Let us attempt a survey of the upper atmospheric problems which need further investigation or are still awaiting solution. It is very satisfactory that direct observations with V-2 rockets have confirmed the surmise made long ago that the main atmospheric constituents up to about 70 km occur in the same proportion as near the ground [23]. It is very probable that not far above this level the dissociation of O_2 molecules begins. But the relative distributions of O_2 and O with height are not known with certainty. Several possible distributions have been obtained by different authors under different assumptions [25, 57, 88, 90, 108]. But the distribution which best represents the actual state of affairs has still to be found. Accurate determination of this distribution is very necessary because, according to more than one author, the electron-density distribution in Region E depends upon the distribution of molecular oxygen in this transition layer [16, 67].

The question at what height above 70 km diffusive separation becomes important is of considerable interest because on it will depend the relative abundance of N_2 and O in the highest regions of the atmosphere [75]. With diffusive separation the highest regions should consist almost entirely of O atoms. But, as mentioned earlier, the spectrum of auroral streamers up to 1000 km contains atomic oxygen lines and molecular nitrogen bands with comparable intensities. It is still not clear how nitrogen molecules which are nearly twice as massive as oxygen atoms reach such great heights.

In the ozonosphere the modes of production and destruction of ozone and the ozone equilibrium resulting therefrom are in need of further elucidation. Interesting relations have been found between the ozone content of the atmosphere and weather conditions. But the ozone observations are still confined to a few stations. It is extremely desirable that a world-wide network of stations be established for simultaneous and systematic studies of atmospheric ozone, of temperature distribution in the middle atmosphere (by abnormal sound propagation method), and of meteors (by radar technique [34, 46]). Such studies will be helpful in understanding the correlations that have been observed between weather conditions in the troposphere and temperature variations in the middle atmosphere and may also be of help in long-range weather forecasting.

The temperature distribution in the region 70–90 km needs more accurate determination. Rocket observations confirm the results of studies by various indirect methods indicating that there is a rapid drop in temperature in this region. But the magnitude of the

drop is still not known with any reasonable degree of certainty.

Thanks to the work of Taylor [100] and of Pekeris [86], the long-standing problem of semidiurnal barometric oscillations and, along with it, the difficulty of the dynamo theory of quiet-day (solar) magnetic variations appears to have been solved. But the lunar semidiurnal barometric oscillations and the corresponding magnetic variations still have their puzzles. The barometric oscillations have unaccounted-for variations both in regard to amplitude and phase.

Our knowledge of upper atmospheric ionization has been extended remarkably by radio exploration utilising the powerful pulse technique. Notable additions to our knowledge have also been made by the adaptation of radar technique for this purpose. The ionosphere, however, is still full of mystery. Very plausible hypotheses have been put forward regarding the production of the different ionospheric regions, but it is safe to say that the last words on the subject have not yet been spoken.

Part, at least, of what is known as sporadic E is now known certainly to be due to meteoric ionization. The sudden bursts of ionization that had been noticed by many ionospheric workers have been traced to ionization produced by individual meteor trails [55]. It may be recalled, in this connection, that 450 kg of meteoric material burn up every day in the atmosphere near the level of Region E. Still, there is clear evidence that meteors cannot be assumed to be responsible for all sporadic E [91]. Also the increase of meteoric echo frequency, when the wave length is increased from 6 to 8 m (the large background rate), is still not quite explained [55, 89, 105].

An interesting observation on radio echoes from meteoric trails requires further clarification. Echoes of long duration show, besides a regular decay, a periodic fluctuation in intensity. It has been suggested that such a fluctuation may be caused by variations of wind with height [45]. These winds deform the straight column left by the meteor. Further observations in different latitudes are needed to test this suggestion and also to find out how far the magnetic field of the earth influences the cross section of the ionized column. But perhaps the central problem in the study of radio echoes from meteors is to find the mechanism by which sufficient electron density, for periods of more than a minute, could be maintained against the forces of diffusion [18, 45, 56].

The reason for the bifurcation of the F-layer into F_1 and F_2 during daytime is not yet understood. Associations between F_2 ionization and tropospheric conditions have been observed. But what is the origin of such association?

It is now suspected that many of the anomalous behaviours of Region F_2 may be traced to the effect of the terrestrial magnetic field, because ions and electrons in Region F, unlike those in Region E, have long mean free paths. For example, a geomagnetic control of Region F, in the form of a belt of low ionization round the geomagnetic equator, has been observed [4]. A plausible explanation of this has also been given [71] but the phenomenon needs much fuller investigation [53].

An attempt has also been made to explain some of the F_2-region anomalies by action of atmospheric tidal movements [59]. For example, the variations of h' (minimum height) and h_{max} (height of the region of maximum ionization) have been explained as due to simple rising and falling of isobaric surfaces, caused by tides. Variation of N_{max} (maximum electron density) cannot, however, be so simply explained. For this a theory has been developed in which it is shown that horizontal tidal motion of ionized masses gives rise to electrodynamic forces which produce a vertical component everywhere except near the magnetic equator. Many of the observed anomalous behaviours of Region F_2 have been explained by this theory. However, a complete theory of the F_2-region is still lacking.

An ionospheric region is by no means smooth in its ionization. Patches or clouds of more intense ionization, against a general background of uniform ionization, have been detected by workers in different countries [91]. As a matter of fact, the existence of winds in the high regions of the ionosphere has been established by systematic study of the movements of these clouds. The clouds are generally believed to be produced by meteor ionization. But their exact nature and life history are still little known. More observational data in different latitudes are needed.

The coefficient of recombination of electrons and ions in the high ionized regions, as deduced from observations, is found to be several orders higher than the theoretical value. This discrepancy appeared to have been removed by the hypothesis of effective recombination coefficient [7, 61]. Further, the identification of the nocturnal Region F with the luminescent layer of the upper atmosphere emitting the O lines and N_2 bands has helped to unify the effective recombination hypothesis with the emission process of these lines and bands [39, 69]. Contemporary work, however, questions the soundness of the effective recombination hypothesis. It has been pointed out by Martyn [59] that if account is taken of the electrodynamic forces developed as a result of the tidal motions, a term appears in the expression of the recombination equation which becomes important in Region F. The computed result, taking into account the contribution of this term, agrees with the observed value of the recombination coefficient. A closer comparative study of the effective recombination hypothesis and Martyn's hypothesis is needed to estimate the relative importance of the two in bringing agreement between observed and theoretically computed values.

Evidence of association between weather conditions near the ground and ionization density in Region F_2 has been obtained. No theoretical explanation of such association has yet been given.

It is highly desirable that more systematic observations on ionospheric data, particularly in regard to ionospheric absorption and ionospheric tides, be made in different parts of the world. With regard to the tidal

effect, it is very necessary to know how the phase of the tide varies with latitude. It appears that there are more observational stations in the Northern than in the Southern Hemisphere. Consequent lack of data is a great handicap in obtaining a complete world picture of the ionosphere. More ionosphere stations, particularly in the southern latitudes, are therefore very necessary. It has been suggested that a sea expedition be sent out to secure observations at different latitudes and longitudes which are scientifically important but for which data are lacking.

In the night-sky spectrum there are still lines and bands the origin of which is uncertain. Special interest lies in the identification of 3471 A. If this be due to atomic nitrogen, then photo-dissociation of N_2 like that of O_2 will have to be assumed, though laboratory experiments do not show any such dissociation effect.

It is satisfactory to note that the strong radiations 6580 A in the red and 10,444 A in the infrared, the origins of which had long been matters of controversy, have now been identified as bands due to the radical OH [65]. Bands of O_2 are reported to have been identified in the night-sky spectrum. If this is confirmed, then there must be another luminescent layer—perhaps at the level of the E-layer. The presence of sodium D-lines has been established and the location of the luminescent sodium layer has been studied. But the presence of sodium in the high atmosphere has raised many questions which are still unanswered. The continuous spectrum which forms the background of the lines and bands of the night-sky light appears to have received inadequate attention and awaits further study regarding its intensity variations and origin [30].

The intensity variation of night-sky light has a regular part (diurnal and seasonal) and an irregular part. Some investigators have sought to associate the irregular part with the magnetic disturbances. Some correlation has been found, but the whole subject is still in a speculative state and requires more intensive study.

A very plausible hypothesis regarding the excitation of the atomic oxygen lines and of the observed N_2 bands (first positive and Vegard-Kaplan bands) has been given [68], but the subject is still controversial [11].

Several attempts have been made to determine the heights at which night-sky luminescence originates. The values obtained vary from 60 km to a few hundred kilometres [21, 22, 35]. It is very likely that, like the several ionospheric regions, there are several layers of maximum luminescence. It is highly desirable that the many ionospheric stations which are being established in different parts of the world have attached to them observatories for study of night-sky luminescence. There are reasons to believe that some at least of the luminescent layers may be identified with some of the ionospheric layers [39, 69]. Close comparison of the variations of the night-sky intensity with those of ionized regions will help in the determination of the existence of such association.

Of the other lights from the night sky, the mystery of zodiacal light is still far from solved. There is little doubt that the light is due to scattering by some sort of dust cloud in extraterrestrial space. But the location of the cloud is still a matter of controversy; whether it belongs to the earth or to the sun is not known definitely. Many theories as to the origin of this cloud have been put forward, but none of them can be considered fully satisfactory [37, 48, 103].

The problems of the aurorae and of the incidence of the magnetic disturbances, so closely related to each other, are only partially solved. There is little doubt that the primary cause of these phenomena is emission of fast charged particles from the sun. It may be noted that direct evidence has been obtained of the existence of Ca^+ ions and protons between the earth and the sun during magnetic storms [1, 25, 65a]. That these may be at least one of the kinds of charged particles emitted from the sun had long been suspected, but the mechanism of the emission of such particles is still only guesswork. Further, the fundamental dilemma still remains: The charged particles, to produce the observed effects, must arrive in bundles, whereas they are bound to be dispersed on their way by electrostatic repulsion. Attempts at solving this problem by imagining that the particles constitute a neutral beam have met with only partial success [2, 29]. (See, however, [59a].)

The lines and bands of the auroral spectrum may be classified under two heads: first, spectra excited by direct bombardment of charged particles—first negative bands of N_2^+; and second, spectra excited as a result of reactions amongst neutral and charged particles produced by the bombardment (second positive, first positive, and the Vegard-Kaplan bands of N_2, as well as the forbidden lines of O [70]). Of these, the excitation processes of the first positive and the Vegard-Kaplan bands may be the same as those of the night-sky light spectrum. But there remains the problem that whilst the Vegard-Kaplan bands are strong in the night sky they are comparatively faint in the auroral spectrum. It has been suggested that this might be due to the fact that the auroral spectrum originates at much lower heights than the night-sky luminescent layer and as such the metastable $N_2(A)$ molecules from which these bands originate are de-excited by collision. However, objection to this has been raised on the ground that the strength of the Vegard-Kaplan bands does not increase with height in the spectra of auroral streamers. The excitation of second positive bands has been suggested as due to radiative recombination of N_2^+ ions and electrons. But here again it may be objected that the calculated intensity of such radiation is very small compared to the observed intensity.

In regard to the insufficiency of laboratory data which is still standing in the way of upper atmospheric study, mention might be made of the electronic spectra and of the absorption coefficients of N_2, O_2, and O in the extreme ultraviolet region. More complete knowledge is very necessary because the ionization densities of the ionospheric regions are controlled by the photo-ionization of these particles by absorption in the extreme ultraviolet. A more detailed study by the quantum-

mechanical method of the collisional cross sections of these molecules and atoms is also necessary. It may be recalled that ionospheric studies yield collisional frequencies of electrons with atmospheric particles, and one is tempted to infer from this the molecular densities in the respective regions. However, this is not justifiable since the cross sections of the atoms and molecules for low-velocity colliding electrons (corresponding to a temperature, for instance, of 1000K—a few tenths of an electron volt) may be widely different from the gas kinetic cross section [76, 94]. More theoretical and experimental data on the collisional processes between meteor atoms and air molecules are also needed. This will be helpful in improving the theoretical calculations on meteors.

It is a pleasure to record my thanks to Dr. S. N. Ghosh, Imperial Chemical Industries Research Fellow of the National Institute of Sciences of India, now working in my laboratory, for the help he has given me in the preparation of this article by collecting data, checking references, and suggesting improvements.

REFERENCES

1. Adams, W. S., "Survey of the Year's Work at Mount Wilson." *Publ. astr. Soc. Pacif.*, 56: 213–219 (1944).
2. Alfvén, H., "A Theory of Magnetic Storms and of the Aurorae." *K. svenska VetenskAkad. Handl.*, (3) Vol. 18, No. 3, 39 pp.; No. 9, 39 pp. (1939).
3. Appleton, E. V., "Regularities and Irregularities in the Ionosphere—I." (The Bakerian Lecture) *Proc. roy. Soc.*, (A) 162: 451–479 (1937).
4. —— "Two Anomalies in the Ionosphere." *Nature, Lond.*, 157: 691 (1946).
5. —— and Beynon, W. J. G., "Lunar Oscillations in the D-layer of the Ionosphere." *Nature, Lond.*, 164: 308 (1949).
6. Appleton, E. V., and Naismith, R., "Weekly Measurements of Upper Atmospheric Ionization." *Proc. phys. Soc. Lond.*, 45: 389–398 (1933).
7. Appleton, E. V., and Sayers, J., "Recombination in the Ionosphere." *Int. Sci. Radio Un., Proc. Gen. Assemb.* 5: 272–277 (1938).
8. Babcock, H. D., "A Study of the Green Auroral Line by the Interference Method." *Astrophys. J.*, 57: 209–221 (1923).
9. Bannon, J., and others, "The Association of Meteorological Changes with Variations of Ionization in the F_2 Region of the Ionosphere." *Proc. roy. Soc.*, (A) 174: 298–309 (1940).
10. Bartels, J., "Überblick über die Physik der hohen Atmosphäre." *Elekt. Nachr.-Tech.*, Bd. X (Sonderheft), 40 SS. (1933).
11. Bates, D. R., "Theoretical Considerations Regarding the Night Sky Emission" in *Emission Spectra of Night Sky and Aurora*. Gassiot Comm. Rep., Phys. Soc., London, 1948. (See pp. 21–33)
12. Bernard, R., "Observation d'un nouveau phénomène de fluorescence dans la haute atmosphère. Présence et variations d'intensité de la radiation λ5893 Å dans la lumière du ciel au crépuscule." *C. R. Acad. Sci., Paris*, 206: 448–450 (1938).
13. ——"The Emission of the Night Sky" in *Emission Spectra of Night Sky and Aurora*. Gassiot Comm. Rep. Phys. Soc., London, 1948. (See pp. 91–92)
14. Beynon, W. J. G., "Evidence of Horizontal Motion in Region F_2 Ionization." *Nature, Lond.*, 162: 887 (1948).
15. Best, N., Havens, R., and LaGow, H., "Pressure and Temperature of the Atmosphere to 120 km." *Phys. Rev.*, 71: 915–916 (1947).
16. Bhar, J. N., "Stratification of the Ionosphere and the Origin of the E_1 Layer." *Indian J. Phys.*, 12: 363–386 (1938).
17. —— and Syam, P., "Effect of Thunderstorms and Magnetic Storms on the Ionization of the Kennelly-Heaviside Layer." *Phil. Mag.*, 23: 513–528 (1937).
18. Blackett, P. M. S., and Lovell, A. C. B., "Radio Echoes and Cosmic Ray Showers." *Proc. roy. Soc.*, (A) 177: 183–186 (1941).
19. Budden, K. G., Ratcliffe, J. A., and Wilkes, M. V., "Further Investigation of Very Long Waves Reflected from the Ionosphere." *Proc. roy. Soc.*, (A) 171: 188–214 (1939).
20. Buisson, H., et Fabry, C., "Mesures de longueurs d'onde dans l'extrémité ultra-violette du spectre solaire." *J. phys. Radium*, 2: 297–302 (1921).
21. Cabannes, J., et Dufay, J., "Analyse spectrale de la lumière du ciel nocturne au Pic du Midi." *C. R. Acad. Sci., Paris*, 198: 306–309 (1934).
22. —— and Gauzit, J., "Sodium in the Upper Atmosphere." *Astrophys. J.*, 88: 164–172 (1938).
23. Chackett, K. F., Paneth, F. A., and Wilson, E. J., "Chemical Composition of the Stratosphere at 70 km. Height." *Nature, Lond.*, 164: 128–129 (1949); "Chemical Analysis of Stratosphere Samples from 50 to 70 km. Height." *J. atmos. terr. Phys.*, 1: 49–55 (1950).
24. Chapman, S., "The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth." *Proc. phys. Soc. Lond.*, 43: 26–45 (1931); 43: 483–501 (1931).
25. —— "Some Phenomena of the Upper Atmosphere." *Proc. roy. Soc.*, (A) 132: 353–374 (1931).
26. —— "The Upper Atmosphere," in *Reports on Progress in Physics.*, 3: 42–65. Phys. Soc., London, 1936.
27. —— *The Earth's Magnetism*. London, Methuen, 1936. (See p. 104)
28. —— and Bartels, J., *Geomagnetism*, Vol. I, *Geomagnetic and Related Phenomena*. Oxford, Clarendon Press, 1940. (See p. 370)
29. Chapman, S., and Ferraro, V. C. A., "A New Theory of Magnetic Storms." *Terr. Magn. atmos. Elect.*, 36: 77–97; 171–186 (1931); 37: 147–156 (1932).
30. Dobronravine, P. P., et Khvostikov, I. A., "Spectre de luminescence nocturne du ciel dans la région ultra-violette." *C. R. (Doklady) Acad. Sci. URSS*, 23: 233–237 (1939).
31. Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., "Meteorology of the Lower Stratosphere." *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
32. Dobson, G. M. B., Harrison, D. N., and Lawrence, J., "Measurements of the Amount of Ozone in the Earth's Atmosphere and Its Relation to other Geophysical Conditions. Part III." *Proc. roy. Soc.*, (A) 122: 456–486 (1929).
33. Dufay, J., Gauzit, J., et Tcheng Mao-lin, "Spectre de l'aurore du 1er mars 1941." *Cah. Phys.*, No. 1, pp. 59–64 (1941).
34. Ellyett, C. D., and Davies, J. G., "Velocity of Meteors Measured by Diffraction of Radio Waves from Trails during Formation." *Nature, Lond.*, 161: 596–597 (1948).
35. Elvey, C. T., and Farnsworth, A. H., "Spectrophotometric Observations of the Light of the Night Sky." *Astrophys. J.*, 96: 451–467 (1942).

36. Ferrell, P., "Winds in the Ionosphere Indicated by Radio Reflecting 'Clouds' of High Ionic Density." *Bull. Amer. meteor. Soc.*, 25: 371 (1944).
37. Fesenkov, V. G., "On the Mass of the Moon's Atmosphere." *Astr. J. Sov. Un.*, Vol. 20, Pt. 2, pp. 1–8 (1943).
38. Gherzi, E., "Ionosphere and Weather." *Nature, Lond.*, 165: 38 (1950).
39. Ghosh, S. N., "Measurements of the Intensity of the Night Sky at Calcutta." *Indian J. Phys.*, 20: 205–213 (1946).
40. Ghosh, S. P., M.Sc. (Pure Physics) Thesis, Univ. of Calcutta, 1938.
41. Götz, F. W. P., Meetham, A. R., and Dobson, G. M. B., "The Vertical Distribution of Ozone in the Atmosphere." *Proc. roy. Soc.*, (A) 145: 416–446 (1934).
42. Götz, F. W. P., "Zum Nordlichtspektrum des 17. April 1947." *Experientia*, 3: 185 (1947).
43. Gowan, E. H., "Ozonosphere Temperatures under Radiation Equilibrium." *Proc. roy. Soc.*, (A) 190: 219–226 (1947); "Night Cooling of the Ozonosphere." *Ibid.*, 190: 227–231 (1947).
44. Gutenberg, B., "New Data on the Lower Stratosphere." *Bull. Amer. meteor. Soc.*, 30: 62–64 (1949).
45. Herlofson, N., "The Theory of Meteor Ionization," in *Reports on Progress in Physics*, 11: 444–454. Phys. Soc., London, 1946–47.
46. Hey, J. S., Parsons, S. J., and Stewart, G. S., "Radar Observations of the Giacobinid Meteor Shower, 1946." *Mon. Not. R. astr. Soc.*, 107: 176–183 (1947).
47. Hoffmeister, C., "Die Strömungen der Atmosphäre in 120 km. Höhe." *Z. Meteor.*, 1: 33–41 (1946).
48. Hulburt, E. O., "The Upper Atmosphere" in *Terrestrial Magnetism and Electricity*, J. A. Fleming, ed. (Vol. VIII of *Physics of the Earth*). New York, McGraw, 1939. (See pp. 567–568)
49. Jeans, J H., *The Dynamical Theory of Gases*. Cambridge, University Press, 2nd ed., 1916. (See pp. 351–363)
50. Johnson, N. K., "Wind Measurements at 30 km." *Nature, Lond.*, 157: 24 (1946).
51. Jones, J. E., "Kinetic Theory of Non-uniform Rarefied Gas and the Escape of Atmospheric Molecules." *Trans. Camb. phil. Soc.*, 22: 535–556 (1923).
52. Krause, E. H., "High Altitude Research with V-2 Rockets." *Proc. Amer. phil. Soc.*, 91: 430–446 (1947).
53. Liang, P. H., "F_2 Ionization and Geomagnetic Latitudes." *Nature, Lond.*, 160: 642–643 (1947).
54. Lindemann, F. A., and Dobson, G. M. B., "A Theory of Meteors and the Density and Temperature of the Outer Atmosphere to Which It Leads." *Proc. roy. Soc.*, (A) 102: 411–437 (1923).
55. Lovell, A. C. B., "Meteoric Ionization and Ionospheric Abnormalities," in *Reports on Progress in Physics*, 11: 415–442. Phys. Soc., London, 1946–47.
56. —— "Meteor Ionisation in the Upper Atmosphere." *Sci. Progr.*, 38: 22–42 (1950).
57. Majumdar, R. C., "The Theory of Molecular Dissociation and the Fundamental Mechanisms in the Upper Atmosphere." *Indian J. Phys.*, 12: 75–86 (1938).
58. Manning, L. A., "The Theory of the Radio Detection of Meteors." *J. appl. Phys.*, 19: 689–699 (1948).
59. Martyn, D. F., "Atmospheric Tides in the Ionosphere." *Proc. roy. Soc.*, (A) 189: 241–260 (1947); 190: 273–288 (1947); 194: 429–463 (1948).
59*a*.—— "The Theory of Magnetic Storms and Auroras." *Nature, Lond.*, 167: 92–94 (1951).
60. —— and Pulley, O. O., "The Temperatures and Constituents of the Upper Atmosphere." *Proc. roy. Soc.*, (A) 154: 455–486 (1936).
61. Massey, H. S. W., "Dissociation, Recombination and Attachment Processes in the Upper Atmosphere—I." *Proc. roy. Soc.*, (A) 163: 542–553 (1937).
62. Mathur, L. S., "Reflection of Sound Waves from the Stratosphere over India in Different Seasons of the Year." *Indian J. Meteor. Geophys.*, 1: 24–34 (1950).
63. McKinley, D. W. R., and Millman, P. M., "A Phenomenological Theory of Radar Echoes from Meteors." *Proc. Inst. Radio Engrs., N. Y.*, 37: 364–375 (1949).
64. McNish, A. G., "Terrestrial-Magnetic and Ionospheric Effects Associated with Bright Chromospheric Eruptions." *Terr. Magn. atmos. Elect.*, 42: 109–122 (1937).
65. Meinel, A. B., "Hydride Emission Bands in the Spectrum of the Night Sky." *Astrophys. J.*, 111: 207 (1950); "Identification of the 6560 A Emission in the Spectrum of the Night Sky." *Ibid.*, 111: 433–434 (1950).
65*a*.—— "Evidence for the Entry into the Upper Atmosphere of High-Speed Protons during Auroral Activity." *Science*, 112: 590 (1950).
66. Milne, E. A., "The Escape of Molecules from Gaseous Stars." *Trans. Camb. phil. Soc.*, 22: 483–517 (1923).
67. Mitra, S. K., "Origin of the E Layer of the Ionosphere." *Nature, Lond.*, 142: 914–915 (1938).
68. —— "Light of the Night Sky." *Sci. Cult.*, 9: 46–48 (1943–44).
69. —— "Night Sky Emission and Region F Ionization." *Nature, Lond.*, 155: 786 (1945).
70. —— "The Auroral Spectrum." *Nature, Lond.*, 157: 692 (1946).
71. —— "Geomagnetic Control of Region F_2 of the Ionosphere." *Nature, Lond.*, 158: 668–669 (1946).
72. —— *The Upper Atmosphere*. Calcutta, Royal Asiatic Society of Bengal, 1948. (See pp. 507–510)
73. —— and Banerjee, A. K., "The Fringe of the Atmosphere and the Ultra-violet Light Theory of Aurora and Magnetic Disturbances." *Indian J. Phys.*, 13: 107–144 (1939).
74. Mitra, S. K., Bhar, J. N., and Ghosh, S. P., "The Lower Ionosphere." *Indian J. Phys.*, 12: 455–465 (1938).
75. Mitra, S. K., and Rakshit, H., "Distribution of the Constituent Gases and Their Pressures in the Upper Atmosphere." *Indian J. Phys.*, 12: 47–61 (1938).
76. Mitra, S. K., Ray, B. B., and Ghosh, S. P., "Cross-Section of Atomic Oxygen for Elastic Collision with Electrons and Region F Absorption." *Nature, Lond.*, 145: 1017 (1940).
77. Mitra, S. N., "A Radio Method of Measuring Winds in the Ionosphere." *Proc. Inst. elect. Engrs.*, 96 (Part III): 441–446 (1949).
78. Mohler, F. L., "Recombination and Electron Attachment in the *F* Layers of the Ionosphere." *J. Res. nat. Bur. Stand.*, 25: 507–518 (1940).
79. Munro, G. H., "Short-Period Changes in the F Region of the Ionosphere." *Nature, Lond.*, 162: 886–887 (1948).
80. Newell, H. E., Jr., "Exploration of the Upper Atmosphere by Means of Rockets." *Sci. Mon.*, 64: 453–463 (1947).
81. Nicolet, M., "Le problème des régions ionosphériques." *J. geophys. Res.*, 54: 373–381 (1949).
82. Olivier, C. P., "Long Enduring Meteor Trains." *Proc. Amer. phil. Soc.*, 85: 93–135 (1942).
83. Öpik, E., "Atomic Collisions and Radiation of Meteors." *Acta Univ. dorpat. (tartu.)*, 26: 2–41 (1933). Harvard Repr. No. 100 (1933).
84. Paneth, F. A., "The Upper Atmosphere." *Quart. J. R. meteor. Soc.*, 65: 304–328 (1939).
85. Pasternack, S., "Transition Probabilities of Forbidden Lines." *Astrophys. J.*, 92: 129–155 (1940).
86. Pekeris, C. L., "Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 158: 650–671 (1937).
87. Penndorf, R., "The Temperature of the Upper Atmosphere." *Bull. Amer. meteor. Soc.*, 27: 331–342 (1946). (Translated from *Meteor. Z.*, 58: 1–10 (1941) by C. C. Chapman.)
88. —— "The Vertical Distribution of Atomic Oxygen in the Upper Atmosphere." *J. geophys. Res.*, 54: 7–38 (1949).

89. Pierce, J. A., "Abnormal Ionization in the E Region of the Ionosphere." *Proc. Inst. Radio Engrs., N. Y.*, 26: 892–908 (1938).
90. Rakshit, H., "Distribution of Molecular and Atomic Oxygen in the Upper Atmosphere." *Indian J. Phys.*, 21: 57–68 (1947).
91. Ratcliffe, J. A., "The Ionosphere and the Propagation of Radio Waves." (Summary of the summer meeting, July 14–16, 1949, of the London Physical Society.) *Nature, Lond.*, 164: 511–513 (1949).
92. —— and White, E. L. C., "An Automatic Recording Method for Wireless Investigations of the Ionosphere." *Proc. phys. Soc. Lond.*, 45: 399–410 (1933)
93. —— "Some Automatic Records of Wireless Waves Reflected from the Ionosphere." *Proc. phys. Soc. Lond.*, 46:107–115 (1934).
94. Ray, B. B., "Absorptions of Radio Waves in the Ionosphere." *Sci. Cult.*, 3: 679–682 (1937–38).
95. Seaton, S. L., "State of the Upper Atmosphere." *J. Meteor.*, 5: 204–219 (1948).
96. Sheppard, P. A., "Meteorology: The Exploration of the Upper Atmosphere." *Sci. Progr.*, 37: 488–503 (1949).
97. Sparrow, C. M., "Physical Theory of Meteors." *Astrophys. J.*, 63: 90–110 (1926).
98. Spitzer, L., "The Terrestrial Atmosphere above 300 km" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed. Chicago, University of Chicago Press, 1949. (See pp. 213–249)
99. Störmer, C., "The Meteor Train of March 24, 1935." *Astrophys. norveg.*, 3: 117–138 (1939).
100. Taylor, G. I., "The Oscillations of the Atmosphere." *Proc. roy. Soc.*, (A) 156: 318–326 (1936).
101. Tönsberg, E., and Chalonge, D., "Ozone Measurements of the Auroral Observatory, Tromsö (70° N. L.)." Conference on Atmospheric Ozone (Oxford) September 9–11, 1936. *Quart. J. R. meteor. Soc.*, Supp., 62: 55–58 (1936).
102. Vassy, A., et Vassy, E., "La température de la haute atmosphère." *J. Phys. Radium*, 3: 8–16 (1942).
103. Vegard, L., "Die Deutung der Nordlichterscheinungen und die Struktur der Ionosphäre." *Ergebn. exakt. Naturw.*, 17: 229–281 (1938).
104. Weekes, K., and Wilkes, M. V., "Atmospheric Oscillations and the Resonance Theory." *Proc. roy. Soc.*, (A) 192: 80–99 (1947).
105. Whipple, F. L., "Meteor Astronomy." *Observatory*, 68: 226–232 (1948).
106. —— "Meteors and the Earth's Upper Atmosphere." *Rev. mod. Phys.*, 15: 246–264 (1943).
107. Wooley, R. G. D., *Report on Certain Aspects of Solar Knowledge Relevant to Ionosphere.* Proc. of the First Meeting of the Mixed Commission on the Ionosphere, 1948. International Council of Scientific Unions, Brussels, 1949. (See pp. 85–90)
108. Wulf, O. R., and Deming, L. S., "The Distribution of Atmospheric Ozone in Equilibrium with Solar Radiation and the Rate of Maintenance of the Distribution." *Terr. Magn. atmos. Elect.*, 42: 195–202 (1937).

PHOTOCHEMICAL PROCESSES IN THE UPPER ATMOSPHERE AND RESULTANT COMPOSITION

By SYDNEY CHAPMAN

Queen's College, Oxford, England

INTRODUCTION

1. Chemistry and the Troposphere. The meteorology of the lower atmosphere, the region of weather, is essentially a *physical* science, in which chemistry plays practically no part; chemists may become good meteorologists, but meteorologists need not know much chemistry. Their work is concerned with fluid dynamics, with heat interchanges by radiation and conduction and convection, with mixing and changes of state of water vapor—evaporation, condensation, sublimation, and precipitation.

Chemists are indeed interested in the composition of air,[1] but their work has generally been supposed to end with the chemical analysis. This shows that air is a *mixture* of several gases, merely coexisting and not reacting chemically with each other to any significant degree in the lower atmosphere.[2] This is true for the main permanent constituents, nitrogen N_2 (78 per cent of dry air, by volume) and oxygen O_2 (21 per cent); for the chief variable constituent, water vapor H_2O; and also for carbon dioxide CO_2 and the rare gas constituents, helium, neon, argon, krypton, xenon, and radon.

In the lower atmosphere the only changes in the composition of dry air are minor and local, such as the *withdrawal* of a small amount of oxygen (and ozone) by plants or by oxidation of iron and organic matter; or the *emission* of carbon dioxide by plants, of organic and other gases by decaying matter and by factories and chimneys, and of helium and radon from radioactive matter below ground. Among the few constituents of the lower atmosphere that undergo appreciable chemical change are ozone and radon; the latter disintegrates radioactively, so that its concentration decreases upwards. This is a process of *nuclear* chemistry, outside the realm of ordinary chemistry, which is concerned with reactions affecting only the outer structures of atoms and molecules. The incidence of cosmic rays must also produce nuclear chemical transformations, relatively very rare, but worthy of study. However, the main fact of tropospheric chemistry is the uniformity of the composition of dry air all over the globe, near the ground, as shown by Paneth.[1]

Thus ozone is almost the only chemically active constituent of the tropospheric air, though it may be that there are other very rare but chemically active gases present, whose changes have not yet been considered. Ozone is always present in the air near the ground,[3] though it is reduced in concentration over industrial areas and to windward of them by oxidation processes. For a long time, however, little attention was paid to the chemical problems raised by the continued presence in the air of a somewhat unstable gas like ozone.

2. Atmospheric Chemistry. A more active interest in the chemical processes of the atmosphere was aroused when it was found that the ozone density is less in the air near the ground than in the air well up in the stratosphere—the height of maximum density being 20 to 25 km, though early estimates were 40 to 50 km. Such a distribution differs from that of all the other known constituents of the lower atmosphere, which decrease upwards in density, maintaining the same relative concentrations at least up to 20-km height, except in the cases of radon, whose relative concentration decreases upwards, and water vapor, whose distribution is irregular and variable.

This peculiar height distribution of ozone may be explained by attributing the formation of ozone to the dissociation of oxygen molecules into atoms O by sunlight; the O atoms then combine with O_2 molecules to form ozone ($O + O_2 \rightarrow O_3$), which is itself dissociated by sunlight ($O_3 \rightarrow O_2 + O$). The consequent reactions between the three forms of oxygen (O, O_2, O_3), which achieve a slowly changing equilibrium mixture, are very complicated, and the relative importance of these reactions varies with the height. Above about 80 km the ozone concentration sinks to insignificance and the atomic oxygen concentration increases to importance, and at levels probably of 100 to 120 km O becomes predominant over molecular oxygen. Its presence is indicated by the emission spectra of the atmosphere, namely those of the night sky and of the aurora. Similar spectral evidence reveals the presence of atomic sodium Na and also of atomic nitrogen N in the upper atmosphere, and the terrestrial part of the absorption spectrum of sunlight indicates that some oxides of nitrogen also exist in the atmosphere. The presence of such chemically active gases as atomic oxygen and sodium raises very interesting chemical problems, on which there is now a growing literature.

The chemical reactions and their energy mainly originate from the absorption of *light* from the sun, so that meteorological chemistry is in the main a branch of *photo*chemistry. It includes also some *impact*-chemistry, as it may be called, concerned with reactions initiated by the entry of fast-moving particles into the atmosphere from outside—a process somewhat analogous to the chemistry of reactions inside discharge tubes, in which fast-moving ions and electrons produce chemical changes in the gas. In the atmosphere the impacts are mainly those of the solar particles that cause aurorae and magnetic storms.

1. Consult "The Composition of Atmospheric Air" by E. Glueckauf, pp. 3–10 in this Compendium.

2. Consult "Some Problems of Atmospheric Chemistry" by H. Cauer, pp. 1126–1136 in this Compendium.

3. Consult "Ozone in the Atmosphere" by F. W. P. Götz, pp. 275–291 in this Compendium.

The chemistry of the atmosphere differs from laboratory chemistry because the atmosphere is without any material boundary except the ground, which exerts an appreciable chemical influence only on the air at the lowest levels. Above these levels, the wall-reactions, so important in laboratory chemistry of gases, play no part in meteorological chemistry.

The chemistry of the upper atmosphere is a borderline subject between the main fields of chemistry and meteorology, and naturally involves many aspects and conceptions that may be unfamiliar to meteorologists. The chief aim of this article is to assist them to read past and future papers on the subject by outlining its principles and some of its technicalities. The present primitive state of the subject is also briefly described, and some of its main problems are indicated. In so doing it is necessary to pay some attention to atmospheric spectroscopy and ionization.

SOME GENERAL PRINCIPLES OF GAS CHEMISTRY

3. Atomic and Molecular Particles. Atomic particles consist of a positively charged nucleus surrounded by (negative) electrons; molecular particles comprise more than one such nucleus, with electrons around and between them. If the number of electrons is equal to the number of positive electronic charges (each of magnitude ϵ) in the nucleus or nuclei, the particle is electrically neutral, and is called a (neutral) atom or molecule respectively; otherwise the particle is called an ion, atomic or molecular. In general, ions are positive, the number of electrons being less than that necessary for electrical neutrality; but certain gases, notably (in the atmosphere) atomic and molecular oxygen (but not nitrogen or the rare gases) form negative ions, taking on an extra electron beyond their normal number in the neutral state. Negative ions are indicated thus: O^-, O_2^-. Positive ions may be similarly indicated, for example, O^+, N_2^+ or, if doubly ionized, N_2^{++}. Otherwise I, II, III, are added to the chemical symbol to signify respectively the neutral state, the first positively ionized state, the second, and so on: for example, N_2 I, N_2 II, N_2 III instead of N_2, N_2^+, N_2^{++}.

The charge on an atomic nucleus is $Z\epsilon$, where Z is an integer, called the *atomic number;* it determines the chemical nature of the atomic particle. For hydrogen H, carbon C, nitrogen N, oxygen O, and sodium Na, Z is 1, 6, 7, 8, and 11 respectively.

The unit of atomic mass m_0 is one-sixteenth of the mass of the chief type of oxygen atom, the isotope O^{16}; $m_0 = 1823\ m_e$, where m_e denotes the mass of an electron [32]; hence the electrons make a very minor contribution to the masses of atoms. The mass of an atom or molecule is expressed respectively as Am_0 or Mm_0; A or M is called the atomic or molecular *weight.* Atoms may have the same Z (and therefore the same chemical nature) but different atomic weights A; the different forms are called *isotopes.* They usually differ greatly in their relative abundance; for example, in the case of oxygen the isotope for which $A = 16$ far predominates over the isotopes for which A (always nearly an integer number) is approximately 15 or 17, so that the average atomic weight of oxygen (including all isotopes) differs very little from 16. The isotopic constitution of the chemical elements in the atmosphere forms an interesting subject of study, as yet little developed. It will not be further mentioned here.

Chemists call M grams of a substance whose molecular weight is M a *gram-molecule* or mole or gram molecular weight (and, similarly, A grams of an element whose atomic weight is A, a *gram-atom*). The number N of molecules (or atoms) in a gram-molecule (or gram-atom) is M grams divided by the mass of one molecule, namely Mm_0, that is, $N = 1/m_0$ if m_0 is measured in grams; likewise for the number of atoms in a gram-atom. This number N is called *Loschmidt's* (or, less appropriately, *Avogadro's*) *number:*

$$m_0 = 1.660 \times 10^{-24} \text{ gram}; \quad N = 6.023 \times 10^{23}.$$

4. Energy Levels. Atomic and molecular particles, neutral or ionized—hereafter for brevity often referred to simply as particles—may exist in more than one state, and each state has a definite amount of internal energy (as distinguished from the kinetic energy of translatory motion). These states form a discrete (not continuous) series.

In an atomic particle the internal energy is determined solely by the *electronic configuration* (including the electron spins). In a molecular particle it may also include energy of internal vibration and of rotation. These latter energies mainly depend on the disposition and motion of the nuclei, because of their preponderant share of the molecular mass. Each type of energy has a discrete series of values, associated with *quantum numbers,* electronic, vibrational, and rotational. There is consequently a series of electronic *energy levels,* and in the case of molecular particles these levels (for states without vibration and rotation) are supplemented by many neighbouring levels of higher energy, for states in which there is also vibration or rotation or both.

The state of lowest energy is called the *ground state;* the others are called *excited states.* The energies of the latter are reckoned as differences from the energy of the ground state, and are called *excitation energies* or *excitation potentials.* The minimum energy needed to remove an electron altogether from a particle in its ground state is called the *ionization energy* or *ionization potential;* in atmospheric chemistry one mostly considers only the first ionization potential, for the removal of only one electron from the neutral particle. The energy required to detach the excess electron from a negative ion such as O^- or O_2^- is called the *attachment energy* or *electron affinity.* The energy necessary to divide a molecular particle (neutral or ionized) into two atomic (or molecular) particles is called the *dissociation energy* or *dissociation potential.*

In a molecule the distances between its atomic nuclei are in general different in different electronic configurations. A molecule may be raised to a state in which its energy, including vibrational energy, is greater than that required for dissociation, yet the internuclear distances may not be such that the molecule will divide, without some rearrangement of the nuclei and electrons. If the molecule is such that this redistribution can take

place, it is said to be, before the change occurs, in a state of *pre-dissociation* [12]; a similar definition applies to *pre-ionization*.

Symbols are assigned to the various states of particles; the details of the electronic configuration may be specified; for example, by $(1s)^2(2s)^2(2p)^4$ for the lowest levels of the O atom [16, p. 45], but these symbols will not be explained here. For this configuration there are three main energy levels associated with the *term* symbols. ^3P (ground term) and $^1D_2$, $^1S_0$, the two latter being *metastable* terms (§ 12). The ground term itself has three closely spaced levels, $^3P_2$, $^3P_1$, and $^3P_0$, the first being the lowest. Often the suffixes in these term symbols are omitted.

For atomic sodium the ground state has the symbol ^2S, and the first excited state has two closely spaced levels $^2P^0_{\frac{3}{2}}$ and $^2P^0_{1\frac{1}{2}}$. The symbols for the states of molecules are, as might be expected, more complicated than those for atoms. For example, the ground state and the next succeeding energy states for neutral molecular nitrogen have the symbols $X^1\Sigma_g$, $A^3\Sigma_u$, and $B^3\Pi_g$ [12].

5. Energy Units. The cgs unit of energy is the erg, and a larger unit is the joule (1 joule = 10^7 ergs). Another unit much used by chemists is the calorie (or gram-calorie—the terms have identical meanings), which is the energy required to raise the temperature of 1 cc of water at 15C by 1C (1 calorie = 4.185 joules). Chemists often quote energies in kilocalories (or large calories); 1 kilocalorie = 1000 calories.

Physicists often use another unit called the *electron volt* (ev). An energy of V ev is equal to that acquired by an electron in traversing a fall of potential of V volts; this is ϵV joules, if ϵ (the electronic charge) is measured in coulombs:

$$\epsilon = 4.802 \times 10^{-10} \text{ electrostatic cgs units}$$
$$\epsilon = 1.602 \times 10^{-20} \text{ electromagnetic units}$$
$$\epsilon = 1.602 \times 10^{-19} \text{ coulombs.}$$

Hence, as a volt is 10^8 electromagnetic units,

$$1 \text{ electron volt} = 1.602 \times 10^{-12} \text{ ergs.}$$

Hence also for a change of energy of V ev per particle, the change expressed in gram-calories per mole (N particles) is $1.602 \times 10^{-12}\, NV/4.185 \times 10^7$, so that 1 ev per particle is equivalent to 23,053 calories per mole.

Yet another form of energy-reckoning by physicists, the *wave number*, is explained in the next section in connection with the quantum relation.

6. Light Absorption and Emission; the Quantum Relation. An atomic or molecular particle may gain energy of amount W ergs by absorption of light of frequency ν (the frequency is the number of light-vibrations per second), or it may emit the energy W as light of this frequency: in either case W and ν are connected by the *quantum relation*

$$W = h\nu,$$

where h denotes *Planck's constant:*

$$h = 6.624 \times 10^{-27} \text{ erg sec,}$$

according to the latest estimate by R. T. Birge [5].

This relation can also be expressed in other forms, in terms of the wave length λ of the light, or in terms of its wave number n, which is the number of waves per centimetre: so that if λ is expressed in centimetres (denoted by λ_{cm}),

$$n = 1/\lambda_{cm} ;$$

ν and λ_{cm} are connected by the relation

$$\lambda_{cm}\nu = c,$$

where c denotes the speed of light, which *in vacuo* is

$$c = 2.998 \times 10^{10} \text{ cm sec}^{-1} .$$

Hence the relation $W = h\nu$ can be expressed as $\lambda_{cm}W = hc$ or $W = hcn$; energies and energy differences can therefore be expressed in terms of wave numbers, where

$$1 \text{ wave number} = hc \text{ ergs} = 1.986 \times 10^{-16} \text{ erg.}$$

Hence an energy expressed as V ev has a wave number n given by

$$n = 8068V, \text{ or } V = 1.2395 \times 10^{-4}n.$$

Wave lengths are generally expressed in angstrom units (A) or in microns (μ):

$$1 \text{ A} = 10^{-8} \text{ cm}, \ 1\ \mu = 10^{-4} \text{ cm} = 10^4 \text{ A}.$$

Thus λ expressed in A is $10^8\ \lambda_{cm}$. The relation $\lambda W = hc$, when W is expressed in electron volts V and λ in angstrom units, takes the form

$$\lambda V = 12395 \qquad (\lambda \text{ in A}).$$

Similarly if W is expressed in calories per mole,

$$\lambda W = 2.858 \times 10^5 \quad (\lambda \text{ in A}).$$

7. Spectra Associated with Atomic and Molecular Transitions. When an atomic or molecular particle undergoes a transition from a state of higher energy to one of lower energy, one way in which it can dispose of the energy difference W thus released is by radiation in light of frequency $\nu = W/h$. The spectrum of the light emitted by a gas whose particles are undergoing one or more such transitions is called an *emission* spectrum.

If, however, the particles are undergoing transitions from lower to higher energy levels, one way in which the necessary energy W may be acquired is by absorption of light, which must be of frequency W/h. The spectrum of the beam of light will be darkened at this frequency, because of the absorption and the consequent reduction of the transmitted intensity. This darkening, at as many frequencies as are concerned in the transitions taking place in the gas, gives what is called an *absorption* spectrum.

The spectrum of sunlight shows absorption of two kinds: one (the Fraunhofer spectrum) due to absorption by gases of the sun's own atmosphere, and another due to absorption in the terrestrial atmosphere (§18).

From our knowledge of spectra for various substances, either studied experimentally in the laboratory, or in simple cases obtained from theoretical calculation of the energy levels of particles, it is often (though as yet not always) possible to infer the nature of the

emitting or absorbing particles responsible for given emission or absorption spectra, and also, from the intensity of the spectrum, to infer how many transitions of any such kind are taking place, per unit cross section of the beam of light per second. (See §§18, 19.)

8. Line, Band, and Continuous Spectra. Atomic particles have definite configurations and energy levels, and definite energy differences and associated frequencies, corresponding to definite *lines* in their emission or absorption spectrum. Their spectra are therefore called *line* spectra.

Molecular particles have far more energy levels and energy differences and associated frequencies than atomic particles have, because of their additional energy of vibration and/or rotation. Any "line" in their spectrum, associated with a change purely of electron configuration, may be accompanied by many lines of rather different frequencies, corresponding to the same electronic change accompanied by any one of many possible changes of vibratory or rotational energy. In spectra of moderate dispersion some of these lines may be so crowded together as to appear like a continuous band; hence the name *band spectra* for molecular spectra.

A definite minimum energy (see § 4) is required to ionize or dissociate a particle. Corresponding minimum frequencies are associated with these processes when they are induced by light absorption; but light of any greater frequency may also induce the process, the excess energy going into the kinetic energy of separation of the resulting two particles. Hence the absorption spectrum can be *continuous* on the high frequency (or short-wave) side of the frequency corresponding to the ionization or dissociation potential. Similarly for the emission spectrum resulting from the *recombination* (with radiation) of ions and electrons or combining particles; for example, in the combinations

$$O^- + O^+ \rightarrow O_2 \text{ , or } O + O_2 \rightarrow O_3,$$

the energy released may exceed the ionization or dissociation energy, because of the kinetic energy of approach of the combining particles. Such continuous spectra are an indication of the occurrence of ionization or dissociation, or of their converse, recombination. Certain bands in molecular spectra indicate pre-dissociation or pre-ionization (§4) leading to dissociation or ionization with certain probabilities.

9. Atomic and Molecular Spectra: Different Ranges of Wave Length. The range of the visible spectrum extends from about 4000 A (violet) to 7600 A (red); for $\lambda < 4000$ A the spectrum is called *ultraviolet*, and for $\lambda > 7600$ A, *infrared*. The energy differences between the lower electronic levels of atoms and molecules are often as much as 1 to 10 ev, corresponding to light of wave lengths from 12395 A to 1239 A, extending from the infrared to the ultraviolet. The energy differences between different vibrational states of a molecule in the same electronic configuration are of order 0.1 ev to 1 ev, corresponding to wave lengths from 12395 A to 123950 A (or from 1.2 μ to 12 μ) in the "near" infrared. Purely rotational transitions involve energy differences about one-tenth as great, corresponding to wave lengths from 12 μ to 120 μ in the "far" infrared. Light of a single wave length is called *monochromatic*.

Numerous bands are named after their discoverers or interpreters, working either in the laboratory or with natural light in emission or absorption. Examples are the Hartley and Huggins bands of ozone, extending respectively from about 2000 to 3200 A and from 3200 to 3500 A, and the Chappuis band in the visible region (maximum absorption, in air, at about 6100 A); ozone also has infrared bands. Oxygen (O_2) has Schumann-Runge and Herzberg bands with ranges from about 1750 to 1930 A and 3100 to 3800 A, respectively, and others in the near infrared. Nitrogen (N_2) has the "first positive" system (about 6000 to 6500 A) in the red, as well as the Vegard-Kaplan system (3100 to 4500 A), and ionized nitrogen (N_2^+) has the "first negative" system (3800 to 4700 A). Hydroxyl (OH) has strong infrared (Meinel) bands ($\lambda > 7200$ A). The ranges of wave length here specified are somewhat rough and depend upon conditions which may be different in the laboratory from those in "natural" emission or absorption in the atmosphere.

10. Absorption Coefficients. When light of frequency ν passes through a gas which contains particles that can absorb it, it is weakened proportionately to its own intensity and to the number n of the absorbing particles per cubic centimetre. This is expressed by

$$dI = -k_\nu n I dl,$$

where dI is the reduction of the intensity I in traversing a path length dl; hence k_ν has the same dimensions as $1/ndl$, which is length squared or area. The factor k_ν is called the *atomic (or molecular) absorption coefficient*, or alternatively the *absorption cross section*. In the case of some processes in which light is absorbed k_ν can be calculated theoretically; for example [30], for the ionization of neutral atomic oxygen by radiation whose quanta have energy between 13.55 and 16.86 ev, k_ν is of the order 3×10^{-18} cm^2 . In other cases k_ν must be determined experimentally by observing the decrease of intensity in passing through a known amount of the gas.

If $n = n_0$, where n_0 is the number of molecules per cubic centimetre of gas at normal temperature and pressure, and $\alpha = k_\nu n_0$, the absorption in a gas of uniform density with this value of n is given by $dI = -\alpha I dl$, where α is constant. This leads to the relation

$$I = I_0 e^{-\alpha l} = I_0 10^{-0.4343\alpha l}.$$

Here α is called the *absorption coefficient;* 0.4343α is sometimes called the *decimal absorption coefficient* and denoted by α_{10}, the suffix bearing reference to the replacement of e by 10 in the formula above. For ozone at about $\lambda = 2500$ A, α_{10} is of order 100, and it has the same order of magnitude for molecular oxygen at about 1500 A.

The value of k_ν or α depends greatly on the wave length. For example, for ozone, α_{10} sinks to 10^{-3} for λ somewhat greater than 3000 A.

11. Monochromatic Absorption in an Exponential Atmosphere. The term *exponential atmosphere* is used to

refer to an atmosphere in which the density ρ varies with height h above the ground according to the relation

$$\rho = \rho_0 e^{-h/H}, \; n = n_0 e^{-h/H},$$

where H is a constant (called the scale height), and ρ_0 denotes the density at the ground ($h = 0$). The alternative form $n = n_0 e^{-h/H}$ refers to the number of particles per cubic centimetre, n at height h, n_0 at the ground. Such formulas do not apply to the actual atmosphere unless H is itself regarded as a function of h, but they are useful approximations over a range of height in which H does not vary greatly. The value of H is $kT/\bar{m}g$, where k is Boltzmann's constant (1.380×10^{-16} ergs per degree C); T denotes the absolute temperature, g the acceleration of gravity (981 cm sec^{-2}), and $\bar{m}$ the mean mass of the particles composing the atmosphere. Also $H = RT/\bar{M}g$, where $\bar{M}$ ($= N\bar{m}$) is the mean molecular *weight*, and R ($= kN$) is called the gas constant; $R = 8.314 \times 10^7$ ergs per degree C per mole.

It is of special interest to consider the absorption of light in such an atmosphere, because of the simplicity of the relations involved, and their approximation to the actual conditions in our atmosphere.

It will be supposed that outside the atmosphere the intensity of the light considered is I_∞, and that at height h it is I. The absorption will be supposed to take place with a definite absorption coefficient k_ν, so that in effect the light must be monochromatic, of a definite frequency ν (or rather, within a narrow band of frequency ν to $\nu + d\nu$). Instead of I and I_∞ one might therefore alternatively consider the flux of photons of this frequency, Q per square centimetre per second at height h, and Q_∞ at $h = \infty$ (outside the atmosphere): $Q \propto I$ or $Q/Q_\infty = I/I_\infty$.

The absorption may be due to a particular atmospheric constituent, and if so, n and ρ will refer not to the whole air but to this particular constituent only. If the atmosphere is so static that the different constituents are each distributed in accordance with their molecular weight, each will have its own value of H; for example, for O_2 at 273° absolute (0C), H is about 7.2 km, and for O it is twice as great; the only assumption here made, however, is that H can be regarded as independent of h.

Let χ denote the zenith angle of the sun, which is also the angle made by the beam of light with the vertical (the curvature of the level layers of the atmosphere being neglected). The equation of absorption is $dI = -k_\nu n I \sec\chi \, dh$, of which the integral is

$$I = I_\infty \exp(-k_\nu n_0 H e^{-h/H} \sec\chi).$$

The rate of absorption q per cubic centimetre at height h is $-dI \cos\chi/dh$, which has the value $k_\nu n I$. This has its maximum at the level

$$h_{\max} = H \ln (k_\nu n_0 H \sec\chi) = h_0 + H \ln \sec\chi,$$

where ln denotes the logarithm to the Napierian base e ($= 2.718$), and h_0 is the value of $h_{\max}$ for $\chi = 0$ (vertical sunlight); at this level n has the value $n_{\max}$ given by

$$n_{\max} = (1/k_\nu H) \cos\chi,$$

and q has the value $q_{\max}$ given by

$$q_{\max} = (I_\infty \cos\chi)/H \exp 1,$$

where $\exp 1 = 2.718$.

Let $z = (h - h_{\max})/H$, and $q' = q/q_{\max}$; then the relations above lead to

$$q' = e^{1-z-e^{-z}}.$$

A graph of this function is shown in Fig. 1. It represents the proportionate distribution of absorption per unit volume as a function of height z, reckoned upwards or downwards from the level of maximum, in terms of H as the unit of height. The form of the curve is independent of the particular value of H and of the initial light-intensity I_∞, as well as of the absorption coefficient k_ν. Though the actual *level* of maximum absorption depends on k_ν, n_0, H, and χ, the *value* of the maximum absorption depends on I_∞, H, and χ, but not on k_ν or n_0.

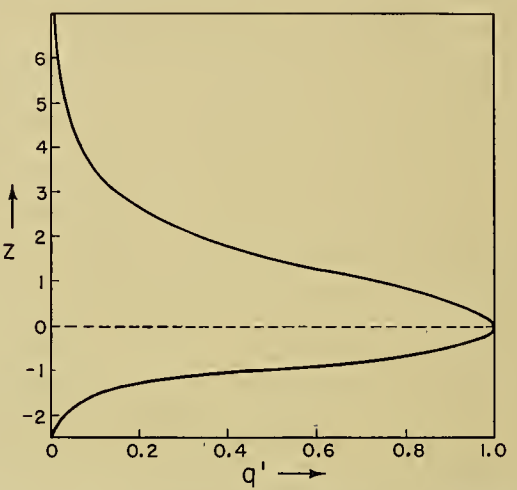


FIG. 1.—The proportionate distribution of absorption of monochromatic radiation per unit volume of gas in an exponentially distributed atmosphere, given as a ratio (q') of the absorption at any level, to that at the level of maximum absorption; the level is indicated by z, reckoned from the height of maximum absorption, in units of the scale-height of the atmosphere.

Below the level $h_{\max}$ the beam intensity and the absorption fall away very rapidly, the beam being quickly attenuated by the increasingly dense air. The decrease of q' upwards is slower; at great heights q' varies approximately as e^{1-z}. The main part of the absorption lies in a layer of thickness about $3H$ (from $z = -1$ to $z = 2$). Throughout the day, $q_{\max}$ varies as $\cos\chi$, and $h_{\max}$ is ln sec χ above h_0. When $\chi = 60°$, sec $\chi = 2$, ln sec $\chi = 0.7$, and so $h_{\max} = h_0 + 0.7H$.

The height h_0 ($= H \ln k_\nu n_0 H$) may be negative for sufficiently small values of k_ν and n_0, that is, if the absorption is weak; q then increases downwards to the ground, and much of the radiation penetrates to ground level. For oxygen (O_2), for which n_0 is of order 10^{19}, and H of order 10^6 cm (7.2 km), the minimum value of k_ν for which h_0 will not be below ground, namely

$1/n_0H$, is 10^{-25} cm^2. For maximum absorption at about the height of the ozone layer ($h = 3H$, for example), $k_\nu n_0 H = e^3 = 20.1$. These formulas for h_0 and $n_{\max}$ or $q_{\max}$ are very useful in the study of atmospheric dissociation and ionization.

If the absorption of the light quanta produces dissociation or ionization, this will partly exhaust the absorbing gas, which will not be distributed exponentially at great heights, where the separation products will become predominant. (See § 22.)

If the absorbing gas is not distributed exponentially (and the ozone layer is such a case) the above formulas will not apply to it, but even so they may give some help towards estimating the nature of the distribution of absorption. If the light absorbed is not monochromatic, but includes radiation of widely differing atomic or molecular absorption coefficients, the resultant distribution of the rate of absorption will be a superposition of layers of the type discussed above, with differing values of $h_{\max}$ and $q_{\max}$ for each small interval of frequency. In the case of the regions of absorption by molecular oxygen or ozone, for which the values of k_ν vary at least 10^5 fold, much of the radiation will be filtered out at levels far above those to which the less absorbable frequencies penetrate.

The light considered here is unidirectional, but a parallel beam of sunlight will be *scattered* as well as absorbed, and the scattered light, relatively more intense for the higher frequencies ν, will itself be scattered and absorbed. The rate of absorption at any level will thus depend on contributions from light of all directions —a complication to which little attention has yet been given in the field of atmospheric photochemistry.

If we knew k_ν and I_∞ for each frequency, and the distribution of each absorbing constituent, the distribution and amount of the absorption would be calculable. In many cases, however, some of these data are unknown, especially I_∞ at frequencies for which the radiation is completely absorbed at high levels. But this gap in our knowledge is now becoming partly closed by information concerning the spectra of sunlight obtained at different heights in the atmosphere by rocket-borne instruments.

12. Reversibility. De-excitation with Radiation. Atomic and molecular processes are reversible. The converse of excitation by radiation is de-excitation (a fall from a higher to a lower energy level) with emission of radiation. The fall need not, however, be made in one stage only; there may be two or more falls through intermediate states. The emission occurs spontaneously, and not at one definite interval after attaining the higher energy level. For each type of transition there is a constant T, called the *half-life* (or more loosely the *lifetime*), such that the probability of the transition occurring during any given short interval from time t to time $t + dt$, after attaining the higher level, is $e^{-t/T}dt$. The transition probability is briefly expressed as $1/T$.

There are certain *selection rules* concerning the permissible changes of the various quantum numbers, from the initial to the final state of the particle; these determine whether or not any particular transition is *allowed* or *forbidden*. In the case of allowed transitions, T is generally very small (*e.g.*, 10^{-8} sec).

Among the various states of a particle there may be some from which there is no allowed transition to a lower level. Such states are called *metastable* states. Even in such cases, however, there may be possible transitions (not of the normal type to which the selection rules apply) with finite though small probability, and such states consequently have a finite lifetime, much longer than that of ordinary excited states.

Neutral atomic oxygen is such a case, of great interest for upper-atmospheric chemistry. Its terms $^1D_2$ and $^1S_0$ of its first electronic configuration (see § 4) are both metastable. Their energy levels (or excitation energies) are approximately 1.96 and 4.17 ev above the ground term ^3P. There is a finite probability 2.0 sec^{-1} for the transition $^1S_0$ to $^1D_2$, involving a change of energy by 2.21 ev and giving rise to the light of the famous "green auroral line" 5577 A. The transition probabilities for fall from $^1D_2$ to the ground levels $^3P_2$, $^3P_1$, and $^3P_0$ are 0, 2.5×10^{-3}, and 7.5×10^{-3} sec^{-1}, giving rise (in the last two cases) to the two "red auroral" lines of atomic oxygen, 6364 A and 6300 A, of which the latter, because of its threefold greater probability, is three times the more intense. Thus the lifetimes of the $^1S_0$ and $^1D_2$ states are of the order ½ sec and 100 sec respectively.

The first excited states of neutral atomic sodium are not metastable; transitions occur with probability 0.62×10^8 sec^{-1} from $^2P^0_{\frac{3}{2}}$ and $^2P^0_{1\frac{1}{2}}$ to the ground state ^2S, giving rise to the sodium yellow line 5893 A (really a close doublet or pair of lines, with wave lengths 5890 A and 5896 A, of which the former is twice as intense as the latter, because the number of atoms in the $^2P_{1\frac{1}{2}}$ state is twice that in the $^2P_{\frac{1}{2}}$ state). The lifetime of the ^2P state is 1.6×10^{-8} sec.

13. Excitation and De-excitation by Collision. Atomic and molecular particles can be excited by impact as well as by absorption of radiation, and the process is reversible, that is, such particles can be de-excited by collision. In the first case the kinetic (and possibly also other) energy of the impinging particle (which may be atomic, molecular, or an electron) is reduced by the energy required for excitation; in the second, it is increased by this amount. Such collisions are called collisions of the *second kind*, or *inelastic* or *superelastic*, in contrast to the elastic collisions considered in the kinetic theory of gases. In these, the speed of separation of the particles after collision equals that of their approach, so that the translatory kinetic energy of their motion relative to their mass-centre is unaltered. A collision between two uncharged unexcited particles is elastic if this kinetic energy is too small to raise either to its first excited state. For such elastic collisions the particles, though without definite boundaries, have a certain joint collisional *cross section* (which in the case of rigid elastic spheres would be πR^2, where R denotes the sum of the radii of the two particles). Similarly there is a collisional cross section for any given type of excitation, depending on the nature of the particles and on their relative speed of approach; it is one mode of ex-

pressing the probability of such excitation occurring in such a collision.

The *mean free path* l_{12} of a particle of type 1, between collisions with particles of type 2, is of the order $1/n_2S_{12}$, where S_{12} is the cross section for the type of collision (elastic or otherwise), and n_2 is the number of particles of type 2 per cubic centimetre of the gas. If V_1 is the mean speed of the particles of type 1, the corresponding *collision-interval* t_{12} is l_{12}/V_1, and the mean collision-frequency ν_{12} is $1/t_{12}$. The number of the collisions per cubic centimetre per second is $n_1\nu_{12}$, which equals $V_1n_1n_2S_{12}$ or $\alpha_{12}n_1n_2$, where $\alpha_{12} = V_1S_{12}$; α_{12} depends on T, the absolute temperature, through V_1 (both directly and in S_{12}), which is proportional to $(T/m_1)^{1/2}$.

Whether, in a gas containing excited particles, de-excitation occurs mainly by collision or by radiation depends on whether the lifetime T of the excited state is greater or less than the collision interval t_{12}. The greater the height in the atmosphere, and therefore the smaller the values of n_1 and n_2, the more likely are the excited particles to emit their excess energy spontaneously, with radiation. For example, the collision frequency of electrons in the D-layer of the ionosphere has lately been estimated to be 1.5×10^6 at 91.7 km height and 2.8×10^6 at 86.3 km (giving a scale height H_1, at this level, of 8.5 km). The collision frequency of atomic particles in the same region is likely to be smaller by perhaps two powers of ten, so that de-excitation by allowed transitions (such as those that give the yellow lines of sodium) would be little reduced by collisions, whereas atoms of oxygen in their metastable states ^1D and ^1S would not radiate appreciably at this level.

14. Kinetic Energy of Particles of a Gas at Absolute Temperature *T*. In a gas in thermal equilibrium at the absolute temperature T, the distribution of velocities for each type of particle (mass m, number n per cubic centimetre) is given by Maxwell's formula,

$$dn = n\left(\frac{m}{2\pi kT}\right)^{3/2} e^{-m(U^2+V^2+W^2)/2kT}\, dUdVdW;$$

dn is the number of these particles, per cubic centimetre, whose velocity components U, V, and W (relative to the mean motion, if any, of the gas) lie within the ranges

$$U \text{ to } U + dU, \qquad V \text{ to } V + dV, \qquad W \text{ to } W + dW.$$

Hence in terms of the speed v of a particle (where $v^2 = U^2 + V^2 + W^2$), and its translatory kinetic energy E $(= \frac{1}{2}mv^2)$,

$$dn = 4\pi n\left(\frac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT}\, v^2dv$$

$$= n\frac{2}{\sqrt{\pi}}\left(\frac{1}{kT}\right)^{3/2} e^{-E/kT}\, E^{1/2}dE,$$

giving the number of particles dn per cubic centimetre with speed between v and $v + dv$, or energy between E and $E + dE$. The last formula does not contain m, whence it appears that the mean E for particles of any mass is the same, and equal to $(\frac{3}{2})kT$ ergs, where k denotes Boltzmann's constant; the energy per mole is NE_T or $(\frac{3}{2})RT$ calories, if the gas constant R is expressed in calories per degree per mole (its value then being 1.986); see § 11.

Table I gives the energy E_T in ergs and electron volts, and also, in the last column, its value per mole (NE_T)

Table I. Mean Translatory Kinetic Energy At Various Temperatures

T deg K	$10^{14}\,E_T$ ergs per particle	E_T ev per particle	NE_T cal per mole
200	4.1	0.026	596
300	6.2	0.039	894
400	8.3	0.052	1192
500	10.4	0.065	1490
750	15.5	0.097	2235
1000	20.7	0.129	2980
1500	31.1	0.194	4470

expressed in calories, for T from 200K to 1500K (the range which is of interest for the upper atmosphere). For particles of mass m, the mean square $\overline{v^2}$ of the speed $(2E_T/m)$, and the mean speed $\bar{v}$, are given by

$$\overline{v^2} = 3kT/m, \qquad \bar{v} = (8\overline{v^2}/3\pi)^{1/2} = 0.921\sqrt{\overline{v^2}}.$$

The fraction of particles (of whatever mass) whose energy is fE_T or more is

$$\frac{2}{\sqrt{\pi}}\left(x_0\, e^{-x_0^2} + \int_{x_0}^{\infty} e^{-x^2}\, dx\right),$$

where $x_0^2 = (\frac{3}{2})f$. This fraction is tabulated. For $f = 5$ the fraction is 0.00182.

Even at the highest temperature given in Table I, E_T is too small to produce appreciable excitation (and still less dissociation or ionization) by impact; and the fraction of the particles that have an energy of 1 ev is likewise very small, even at 1500K. It is not an entirely negligible fraction, however, and the few particles of high energy may have some small influence in determining the state of the upper atmosphere. In particular, they will excite rotation,[4] detectable in absorption spectra, which if obtained from high rockets may give some information concerning the temperature of the upper atmosphere.

15. Dissociation and Ionization in the Upper Atmosphere. A particle can be divided into two (thus being dissociated or ionized) either by radiation or by impact. Each quantum of radiation absorbed, of appropriate frequency ν and energy $h\nu$, divides one particle, so that the rate of production of the separate component particles per cubic centimetre per second is equal to the rate of absorption q of quanta, and is a function of the height, as illustrated for monochromatic absorption in an exponential atmosphere in § 11. The process of division by absorption of radiation is called *photolysis*.

The energy necessary for ionization from the neutral state is rather high for most of the gases of the atmos-

4. A molecular particle without rotational energy is not set in gentle rotation by the impact of another particle, however well directed, unless the transferable kinetic energy (allowing for conservation of momentum as well as of energy) at least equals the minimum rotational excitation energy. (See §4)

phere (sodium is an exception), as shown in Table II; the electron detachment energies for O^- and O_2^- are also given. Table III gives various dissociation potentials of interest in connection with the upper atmosphere.

Table II. Ionization Potentials V ev

Gas	N	O	O^-	Na	N_2	O_2	O_2^-	N_2O
V ev	14.5	13.6	2.2	5.1	15.5	12.2	1	12.7

Table III. Dissociation Potentials V ev

Gas	N_2	O_2	O_3	NaO	H_2	OH
V ev	7.4, 9.8?	5.09	1.1	3?	4.5	4.3

The fraction of particles with thermal kinetic energies of the amounts shown in the two tables above is very small in the upper atmosphere (see § 14), and division by impact is produced mainly by fast-moving particles coming in from outside—meteors, cosmic rays, and (of most importance) the streams or clouds of gas emitted by the sun, which produce aurorae and magnetic storms—and by atmospheric particles to which these external particles communicate sufficient of their energy. They dissociate N_2 and O_2, ionize them and the atoms N, O, and excite N_2, N, O_2, O, and their ions. The influence of such transient (though frequent) impact processes on the average composition of the air in auroral latitudes has received little attention as yet. But there can be little doubt that over most of the earth the chemistry of the upper atmosphere is mainly determined by the effect of sunlight.

Among the chief of these effects is the ionization of the various layers of the ionosphere—D, E, F_1, and F_2, in ascending order of height (respectively about 90, 120, 220, and 300 km). Their detailed explanation still presents difficulties [29]. It is uncertain whether the main ionized particles in the E- and F-layers are O_2 and O respectively, or O and N. The high degree of F-layer N_2-dissociation implied in the latter case seems to conflict with the evidence of the sunlit auroral spectrum indicating that there is undissociated nitrogen 200 km and more *above* the F-layer. There seems to be little N_2^+ in the F-layer. The D-layer has been ascribed to atomic sodium. The formation of these layers filters out all or most of the sunlight of highest frequency, whose quanta exceed 12 ev.

The light whose quanta have energy between about 5 and 10 ev dissociates molecular oxygen, being absorbed at levels between about 15 or 20 km and 120 km height. It is still uncertain to what extent molecular nitrogen is dissociated by part of this radiation, or radiation of somewhat greater frequency; the nitrogen molecule seems to be more readily ionized than dissociated by radiation.

Other constituents of the upper atmosphere that are partly dissociated by absorption of sunlight are ozone ($O_3 \rightarrow O_2 + O$), water vapor ($H_2O \rightarrow OH + H$), and sodium oxide ($NaO \rightarrow Na + O$).

16. Combination of Particles; Conservation of Energy and Momentum; Two- and Three-Body Collisions. The reverse of the division of one particle into two (dissociation or ionization) is the combination of two into one. This is called *recombination* in the case of ions and electrons, and *attachment* in the case of a neutral particle and an electron (for example, $O + e \rightarrow O^-$). The act of combination essentially involves juxtaposition, that is, a collision. The mean interval after division, before the two particles collide and may recombine, depends upon the density (see § 13), and therefore may have any value. Thus it differs essentially from the mean interval between an excitation and a subsequent de-excitation with radiation, the *lifetime*, which depends solely on the nature of the particle itself and on its spontaneous processes.

The results of separation can consequently persist for hours or even days in air of sufficiently low density; during this period the energy of dissociation or ionization is stored up in the gas as a kind of potential energy. In the upper atmosphere some of the energy absorbed from the sunlight during the day hours, causing dissociation and ionization, remains in this potential form after sunset, and during the night it is partly transformed slowly into radiation, which is observed as a faint luminosity of the night sky—the *airglow*.

In general the two particles resulting from the division of any one particle do not recombine with each other. The parent particle gives birth to its progeny among a host of jostling neighbours, and the "new-born infants" at once lose each other in the crowd. Each may combine in due course with some other particle of the same kind as its lost brother, or it may enter into a new partnership of a different kind.

The division of a particle can take place in either of two ways: by absorption of light or by impact. The process that is the reverse of division by absorption of light is combination with emission of light, in a *two-body collision*. The process that is the reverse of the division of a parent particle AB into components A and B, by the impact of a particle X—a transformation symbolized by

$$AB + X \rightarrow A + B + X$$

—is the combination

$$A + B + X \rightarrow AB + X$$

produced by the *three-body* or *triple* collision of A and B and X, to form the two bodies AB and X. The two-body radiative combination is symbolized by

$$A + B \rightarrow AB + h\nu.$$

In all types of collision there is conservation of both energy and momentum. In a radiative combination, by two-body collision, the initial velocities of the two combining particles determine the initial total momentum, which after the collision is divided between the single combined particle and the emitted photon; as

the momentum of the photon is insignificant, the initial momentum determines the velocity of the combined particle, and therefore also its kinetic energy. The energy released in the photon is the energy of the division of AB into A and B, that is, the dissociation or ionization potential, less the change (a reduction) of the kinetic energy of translation resulting from the combination; the amount of this change is not governed by the intrinsic nature of A, B, and AB, but depends on whatever particular velocities A and B may happen to have just before the collision. The chance that in the process of combination it will be possible to emit a photon of just the right energy thus determined is in general small; if the circumstances permit its emission, the combination occurs, otherwise the particles A and B part again without combining, the collision being an elastic one (see § 13). The probability of a radiative combination is expressed by means of its effective collisional cross section.

In a combination by three-body collision, the existence of two particles after the collision permits the ready fulfilment of the two conditions of conservation of momentum and of energy, and the chance that combination will result from the collision is consequently higher. But the chance of the *occurrence* of a triple collision is much less, in a rare gas, than that for a two-body collision; when particles are sparsely scattered, the simultaneous concourse of three particles in one small vicinity is naturally much rarer than that of two particles. The number of two-body collisions of A and B in a gas, per cubic centimetre per second, is $\alpha n_A n_B$ (see § 13); the corresponding number of three-body collisions is $\alpha' n_A n_B n_X$; here the n's denote the numbers of particles per cubic centimetre, of the three kinds. With increasing height all the n's will in general decrease, but the triple product will decrease faster than the double one. Hence the radiative combinations must increase upwards relative to the three-body combinations; at the confines of the atmosphere the combinations are predominantly radiative.

17. Reactions in General; Transference; Activation Energy; Resonance. A *combination* of A and B to form AB (whether by two-body or three-body collision) is a simple special case of chemical transformation or reaction; another general type of reaction involves a *transfer* of part of one of the colliding particles to the other particle, breaking only one molecular bond, as in

$$O + O_3 \rightarrow O_2 + O_2\,, \qquad O + OH \rightarrow O_2 + H\,, \quad (1)$$

or a division of both particles and subsequent interchanged unions, as in

$$OH + OH \rightarrow H_2 + O_2\,. \quad (2)$$

Both of these result from a two-body collision, and as there are two bodies also after the collision, energy and momentum can readily be conserved. It would be expected, therefore, that the *efficiency* of the processes would be greater than that for *radiative* combination, provided of course that they are *exothermic* (*i.e.*, give out heat) and not *endothermic* (*i.e.*, absorb heat). While this is indeed generally true, it is found that transfer does not occur unless the relative kinetic energy in the collision is above a certain limit E_A called the *activation energy*. For endothermic reactions it must at least equal the difference between the internal energies of the initial and final particles. For exothermic reactions, although (by definition) they give out heat, the activation energy is *not*, in general, zero; its value, expressed in calories per N collisions [31], is generally as much as a few kilocalories for reactions of type (1), and some tens of kilocalories for reactions of type (2). It can easily be seen that this means that at moderate temperatures the former are far more important than the latter. Thus at room temperature the fraction of collisions of energy 5 kilocalories or above is about 10^{-4}, whereas the fraction of energy 50 kilocalories or above is only 10^{-39}.

Besides *chemical* transfer reactions there are reactions in which what is transferred is energy of excitation, as in

$$N_2' + O_2 \rightarrow N_2 + O_2'\,, \quad (3)$$

(where the primes denote excitation), or charge, as in

$$O^- + O^+ \rightarrow O' + O''\,, \qquad O^+ + XY \rightarrow O + XY^+\,. \quad (4)$$

The efficiency of transfer of *excitation* is not high unless the change of kinetic energy of relative motion is small; when this is so there is said to be *resonance*, and the effective cross section for the collision may exceed the gas-kinetic cross section. In contrast, transfer of *charge* can occur most readily when the reaction is exothermic by a few tens of calories.

Yet another type of two-body reaction is *dissociative recombination*, as in

$$O_2^+ + e \rightarrow O' + O''\,, \qquad XY^+ + e \rightarrow X' + Y'\,, \quad (5)$$

in which the ionization energy released is taken up in some form by the fragments of the original molecular ion. Quantal arguments have recently been advanced which indicate that the mechanism may be extremely rapid. Bates and Massey [29] have indeed suggested that (5) may be responsible for the observed high (and pressure-independent) coefficient of recombination α for electrons in the E-layer of the ionosphere. The pressure-independence of α in this layer implies that the recombination occurs by two-body collisions; if the rate of combination or reaction between particles of types (a) and (b) is expressed by $\alpha n_a n_b$, and the process took place predominantly by three-body collisions, α would be proportional to the total number of particles per unit volume, and therefore to the pressure (and inversely proportional to the temperature). If both two- and three-body collisions were of importance, α would have a constant part and a part proportional to the pressure. It may be noted that the theory of the rate of recombination observed in discharge tubes, where the conditions are subject to control, still involves unexplained difficulties.

CONSTITUENTS AND REACTIONS IN THE UPPER ATMOSPHERE

18. Absorption-Spectral Evidence Regarding Atmospheric Composition. The absorption spectrum of sun-

light has a part which is of terrestrial origin and includes bands due to N_2, O_2, O_3, CO_2, and H_2O. Except for O_3, the major part of all these gases lies in the lowest and densest part of the atmosphere, but the absorption in the ultraviolet region by N_2, O_2, and O_3 occurs at higher levels.

The amount of ozone present varies with the latitude, season and weather, but is of the order 3 atmo-millimetres. (It is convenient to express the amount of the rarer gases of the atmosphere by the thickness of the layers they would form if all of each such gas were separated out from each vertical column of atmosphere, and collected at ground level at normal temperature (0C) and pressure (760 mm of mercury); the amount may then be named as being in atmo-centimetres or atmo-millimetres.)

There is also good evidence for the presence of about 1 atmo-cm of nitrous oxide (N_2O); it seems likely that it is mainly in the lower atmosphere. This gas is very transparent for radiation of wave lengths greater than 2000 A, and must therefore be very stable photochemically. Study of its ultraviolet spectrum indicates an ionization potential of 12.66 ev, but its dissociation potential and products are still uncertain. Other oxides of nitrogen may also be present; it appears to be possible to set upper limits of 0.1 atmo-mm for the amounts of NO_2 and NO_3; there is some evidence for the presence of less than 1 atmo-mm of N_2O_5. If formaldehyde (CH_2O) is present (presumably in the troposphere and lower stratosphere), its amount must be less than 1 atmo-mm [26].

The sun's absorption spectrum gives information concerning atmospheric composition by day; similar information about the atmosphere at night could be obtained from the spectrum of moonlight or starlight, but the practical difficulties are much greater than with sunlight, owing to the much lower intensity of such light.

19. Emission-Spectral Evidence Regarding Upper Atmospheric Composition. The spectrum of the night sky shows prominently the green and red lines of atomic oxygen and the yellow line of atomic sodium (§ 12). In the infrared there are strong hydroxyl (OH) bands [33] (probably including bands at 10,400 A originally ascribed to N_2' by a process depending on recombination of N atoms) and also Kaplan-Meinel bands of O_2 (1.7 ev). These are the parts of the spectrum as yet most certainly identified. There is also a continuous background throughout the visible region, fading in the ultraviolet at λ 3900; somewhat lost in the blue and violet part of this continuum ($\lambda > 3900$) there are very many lines and bands, and there are others, very distinct, in the ultraviolet ($\lambda < 3900$). Of the latter, the strongest have been ascribed to the Herzberg bands of O_2 (excitation energy 4.7 ev), and others, weaker, as the Schumann-Runge bands of O_2 (6.2 ev). In the visible region many bands are ascribed to the Vegard-Kaplan bands of N_2 (>7.0 ev). Barbier has ascribed some of the visible bands to CO. Not all these ascriptions are certain.

The OH emission may be due to the reaction

$$H + O_3 \rightarrow OH' + O_2 \ (+76 \text{ kilocalories}),$$

the H atoms being produced by dissociation of water vapor by sunlight.

The excitation energies indicated for the oxygen green and red lines respectively are 4.2 ev and 2.0 ev, and 2.1 ev for the sodium yellow (or D) line.

It is likely that the spectrum is mainly what is called a *recombination spectrum*, the energy being provided by the recombination of atomic O to form O_2, and (less probably) of atomic nitrogen to form N_2. One argument for the latter hypothesis is that the recombination of oxygen provides no more than 5.09 ev, whereas that of nitrogen can provide at least 7.38 and possibly 9.76 ev. (The dissociation energy of N_2 is still doubtful.) The energy provided by the recombination of ions and electrons will also contribute to the night-sky spectrum to a smaller degree.

At twilight (dawn and sunset), part of the emission spectrum is enhanced: chiefly the red oxygen line and the sodium yellow line. At these times there also appears emission of light in the band spectrum of ionized nitrogen (N_2^+), in the region 3914 A, correlated with the enhanced emission of the red oxygen line. This betokens higher energy absorption from sunlight in the high atmosphere then irradiated, and consequent emission, some of which is visible from places on the ground where the sun has already set or not yet risen.

The auroral spectrum shows chiefly the green and red lines of atomic oxygen, bands of N_2 and N_2^+; it seems also to indicate the presence of atomic nitrogen, hydrogen, helium, and perhaps He^+. The energy of excitation is much greater than for the night-sky spectrum, and can reasonably be attributed to impacts caused by fast-moving particles coming from the sun and entering the atmosphere from above; they will "knock on" many atmospheric particles, which will thus be secondarily responsible for the impacts causing most of the observed excitation effects. Auroral emission caused by weak corpuscular streams impinging on the upper atmosphere may at times be mingled with the true night-sky recombination spectrum, even when there is no obvious visual sign of the presence of an aurora.

Recently it has been found that the atomic hydrogen lines in the auroral spectrum, when their light is received at a small inclination to the auroral rays, may show much broadening and large Doppler displacements towards the ultraviolet. This indicates that the emitting H atoms are travelling downwards through the atmosphere with speeds up to some thousands of kilometres per second.

20. The Heights of the Absorbing and Emitting Layers. By observing the intensity of absorption or emission of any particular kind along paths of different inclination to the vertical, or (by day) of different zenith distances of the sun, it is possible to estimate the level of absorption or emission, and in some cases to infer also the height-distribution of the atmospheric constituent concerned. In this way it has been possible to determine the height-distribution of atmospheric ozone,[5]

5. Consult "Ozone in the Atmosphere" by F. W. P. Götz, pp. 275–291 in this Compendium.

and the results have been substantially confirmed by more direct measurements, namely, by obtaining the solar spectrum at different altitudes, in and above the maximum ozone level, using instruments carried in balloons or rockets.

The height of the auroral emission has been much studied, chiefly by Störmer; and the spectrum has shown the presence of atomic oxygen and molecular nitrogen (neutral and ionized) between heights of 100 and 1000 km, the greater heights being associated with sunlit aurorae. The determination of the levels of emission of the various spectral components of night-sky light is much more difficult, because of the faintness of the light. The level of the sodium emission is estimated as mainly between 70 and 100 km, though part of the emission has been acribed to greater heights. The red oxygen light is ascribed to a level extending upwards from 100 km. A. and E. Vassy conclude that there is a second layer of emission at 800 to 1000 km. The green oxygen light is estimated to proceed partly from 75 to 100 km, and partly from a much higher level. These results must be considered as still provisional.

21. The Amounts of the Absorbing and Emitting Constituents. The amount of a gas which contributes to the terrestrial part of the absorption spectrum of sunlight can be estimated therefrom in cases in which the absorption coefficient is known through laboratory or other studies. It is in this way that the number of atmo-millimetres of ozone in the atmosphere is determined daily at several places.

In the case of emission it is possible to infer from the intensity the number of quanta received per second from within a vertical column of atmosphere of, say, an area of one square centimetre at the ground. This was first done by Rayleigh for the light of the green oxygen line in the night sky. He found that 2×10^8 photons of this light are radiated per square centimetre per second, and later studies have confirmed this. The corresponding number for the D line of Na is 8×10^7. There is a very strong infrared emission band of OH at about 10,400 A, for which the number of photons may be 10^{10} or even more. Studies of this kind are still in their infancy: they do not indicate the total amount of the emitting gas, but they give very valuable information to be fitted into our total conception of the state and phenomena of the upper atmosphere [24].

As an example of the type of argument that can be based on such data, consider the green oxygen emission; as it continues all night long with no very great change of intensity, the number of quanta emitted from a column one centimetre square per night (about 40,000 sec) is of order 10^{13}. It is not difficult to show that this exceeds the number that might be derived from ion-electron recombination in the ionized layers of the ionosphere, but that the number of the $O + O \rightarrow O_2$ recombinations is more than adequate. Hence this green light may draw its energy from the daily dissociation of oxygen by sunlight during the daytime, a reservoir of energy steadily drawn upon without apparent serious depletion throughout the night. The process suggested is a three-body collision, in which the third body is the oxygen atom which is thereby excited so as to emit the green light.

22. Ozone and Atomic Oxygen [26]. The ozone layer is produced by photolysis of O_2 into atomic oxygen. The O_2 absorption of the dissociating radiation occurs at different wave lengths and extends from more than 100 km down to 20 km or so. Let Q_2 and Q_3 denote the respective numbers of O_2 and O_3 molecules dissociated per cubic centimetre per second. Let n, n_1, n_2, n_3 denote the number per cubic centimetre of air molecules in general, and of O, O_2, and O_3 respectively, per cubic centimetre. The ozone is formed by attachment (in three-body collisions $O_2 + O + X \rightarrow O_3 + X$), and is destroyed by the reactions $2O_3 \rightarrow 3O_2$ (probably negligible in the atmosphere) and $O + O_3 \rightarrow 2O_2$, as well as by photodissociation. The atomic oxygen is removed by attachment and also by combination ($2O + X \rightarrow O_2 + X$). These various processes lead to the equations

$$dn_1/dt = 2Q_2 + Q_3 - k_{12}n_1n_2n - k_{13}n_1n_3 ,$$
$$dn_3/dt = k_{12}n_1n_2 - Q_3 - k_{13}n_1n_3 ,$$
$$dn_2/dt = -Q_2 - k_{12}n_1n_2n + k_{11}n_1^2n + 2k_{13}n_1n_3 ,$$

where the k's denote coefficients of combination or reaction. One difficulty in using these equations to determine n_1 and n_3 as functions of height and time is that Q_3 is known only when n_3 is known.

It is possible to explain the seasonal variation and latitudinal distribution of ozone on the basis of these formulas, using the known type of seasonal variation and latitude distribution of Q_2 and Q_3, depending merely on the changing zenith distance of the sun. The constants k are not directly known, but may be chosen (with values reasonable from a photochemical standpoint) consistent with the attempted explanation.

The primary process in ozone formation is O_2 dissociation, with the formation of atomic oxygen. With increasing height above the region of maximum ozone, the rate of attachment of O atoms to O_2 molecules, which will be mainly by three-body collisions, will decrease upwards as $e^{-2h/H}$ (see § 11). As the rate of ozone formation thus decreases rapidly upwards, the life of the free oxygen atoms will increase upwards. By simple quantitative arguments it may be inferred that at about 100 or 120 km the value of n_3 will have decreased to less than 10^5 (from over 10^{12} at the maximum level, about 25 km), whereas n_1 will be of the order 10^{12}, comparable with n_2; and that not far above this level n_1 will exceed n_2, the ratio n_1/n_2 increasing upwards.

Thus from the observed presence of the ozone layer, and from the theory of the associated reactions, the increasingly complete dissociation of oxygen as we ascend to high levels can be inferred independently of, though in accordance with, the spectral evidence for the presence of atomic oxygen as a permanent constituent of the upper atmosphere.

23. Atmospheric Sodium. Our knowledge of the presence of atomic sodium in the high atmosphere comes solely from the emission spectra, night-sky, twilight-

sky, and auroral. Intensity measurements at different altitudes have led to widely different estimates of the heights of emission of the sodium light. Bricard and Kastler have shown that for the twilight emission the spectral line is very narrow, whereas it is much wider in the night-sky emission.

The problems connected with the sodium light concern not only the levels and processes of emission; they concern also the amount of invisible sodium that may be present in the upper atmosphere, that is, the amount of sodium not in a state to emit the yellow (or other observable) light. Sodium is easily ionized by sunlight in the Hartley band, and must intercept some of this light, though most of it is absorbed lower down, in the ozone layer. The presence of some ionized sodium atoms must be expected in the high atmosphere, though their presence is not likely to make an important contribution to any of the ionized layers except possibly to the D-layer. Further, sodium forms oxides, NaO and NaO_2, and these also would not be likely to give direct spectral indication of their presence.

The proportions of Na, Na^+, NaO, NaO_2 and any other forms in which sodium may exist in the upper atmosphere will differ with differing height, and the distribution and reaction processes of the sodium present a fascinating problem of atmospheric photochemistry, on which some speculation has been exercised, though as yet without any clear and definite conclusions being established. The oxides of nitrogen (as well as oxygen itself) may play a significant part in the reactions associated with sodium. The discussion of these oxides, as well as of the sodium itself, demands much knowledge that is still lacking, concerning, for example, energy data and reaction coefficients. As an illustration, one process suggested for the excitation of sodium atoms to the state $^2P_0$ from which they can emit the D light is $Na + O + X \rightarrow NaO + X$, $NaO + O \rightarrow Na' + O_2$; for the excited sodium atom Na' in the latter equation to be in the $^2P_0$ state, the energy 2.1 ev must be derivable from the oxygen dissociation energy gained (5.1 ev) less the energy needed to dissociate NaO; this is about 3 ev, but it is not yet certain that it is not slightly too great for this process to provide the excitation energy, unless the colliding particles NaO and O have kinetic energies substantially greater than their average thermal energy (see § 14). This suggested process is based on the dissociation energy of oxygen stored up in the daytime from absorbed sunlight. Another possible process, involving only two-body collisions, is $Na + O_3 \rightarrow NaO + O_2$, $NaO + O \rightarrow Na' + O_2$; it involves the same doubt as to the latter reaction.

The photochemical problems briefly discussed here have an intrinsic interest and also some practical importance in that their solution is likely to aid in a full understanding of the phenomena of the ionosphere. At present the formation of the various ionized layers, themselves so vital to radio communication, is far from being properly understood, and every additional clue or sidelight concerning the electrical and chemical processes in those regions is likely to prove valuable.

SUGGESTIONS FOR FUTURE WORK

On the observational side the study of the photochemical processes in the upper atmosphere has vast scope. A basic necessity is the determination of the composition of the upper atmosphere as well as the height-distribution of temperature and density. Rocket investigations may aid greatly in this field, particularly with regard to the composition, which indicates the extent to which diffusive separation occurs. Such investigations will also provide basic knowledge as to the solar radiation received at high levels in the atmosphere.

Spectroscopic studies, from the ground and *in situ* (by means of balloons and rockets), have much to add concerning the nature and distribution of the constituents of the upper atmosphere, including water vapor, carbon dioxide, ozone, atomic oxygen and nitrogen, the oxides of nitrogen, hydroxyl (OH), and sodium. The study of the absorption and emission by these and other gases in the high atmosphere, by night, at twilight, and during the day will enhance our understanding of the nature and processes of chemical, electrical, and energetic change there. For this work it is necessary to develop instruments of enhanced light-gathering power, spectral range, and dispersion, in particular for the auroral and night-sky luminosity.

Radio investigations will greatly extend our knowledge of the energy absorption, photoelectric processes, and motions in the ionosphere, both at normal times and during magnetic disturbances, when corpuscular energy as well as wave energy is operative.

Such observations need to be made not only at one or two places on the earth, but at a network of stations, so that the world distribution of the phenomena may be known—for its own interest, and also to aid in understanding the processes occurring in any one place. Continued observations to determine seasonal changes and changes throughout the sunspot cycle are desirable in many cases.

On the theoretical side the subject is still young. Many additional purely chemical and physical data are required as a basis for atmospheric theories, namely data on the rates of various types of atomic and molecular processes, and in some cases on the temperature dependence of these rates. These data must be obtained partly by laboratory measurements, and partly by theoretical calculations in atomic and molecular physics. Such data are needed to elucidate the main processes of photochemical and photoelectrical change in the upper atmosphere, and to determine their height-distribution, distinguishing, among the vast number of possible reactions and processes, those whose rate and number give them real significance.

REFERENCES

I. Books on chemistry, particularly photochemistry.
1. Bonhoeffer, K. F., und Harteck, P., *Grundlagen der Photochemie*. Dresden & Leipzig, T. Steinkopff, 1944.
2. Hinshelwood, C. N., *The Kinetics of Chemical Change in Gaseous Systems*, 4th ed. Oxford, Clarendon Press, 1938.

3. Jost, W., *Explosion and Combustion Processes in Gases*. New York, McGraw, 1946.
4. Kassel, L. S., *Kinetics of Homogeneous Gas Reactions*. New York, Reinhold, 1932.
5. Noyes, W. A., Jr., and Leighton, P. A., *The Photochemistry of Gases*. New York, Reinhold, 1941.
6. Rollefson, G. K., and Burton, M., *Photochemistry and the Mechanism of Chemical Reactions*. New York, Prentice-Hall, Inc., 1939.
7. Semenoff, N. N., *Chemical Kinetics and Chain Reactions*. Oxford, Clarendon Press, 1935.

II. Books on atomic and molecular physics and spectra.

8. Arnot, F. L., *Collision Processes in Gases*. London, Methuen, 1933.
9. Bacher, R. F., and Goudsmit, S. A., *Atomic Energy States as Derived from the Analysis of Optical Spectra*. New York, McGraw, 1932.
10. Brode, W. R., *Chemical Spectroscopy*. New York, Wiley, 1943.
11. Foote, P. D., and Mohler, F. L., *The Origin of Spectra*. New York, Chemical Catalog Co., 1922.
12. Gaydon, A. G., *Dissociation Energies and Spectra of Diatomic Molecules*. London, Chapman, 1947.
13. Jeans, J. H., *Dynamical Theory of Gases*, 4th ed. Cambridge, University Press, 1925.
14. Jevons, W. S., *Report on Band-Spectra of Diatomic Molecules*. Phys. Soc., London, 1932.
15. Massey, H. S. W., *Negative Ions*. Cambridge, University Press, 1938.
16. Moore, C. E., *Atomic Energy Levels*. Washington, U.S. Bureau of Standards, 1949.
17. Mott, N. F., and Massey, H. S. W., *The Theory of Atomic Collisions*. Oxford, Clarendon Press, 1949.
18. Pearse, R. W. B., and Gaydon, A. G., *The Identification of Molecular Spectra*. London, Chapman, 1941.
19. Sponer, H., *Molekülspektren und ihre Anwendung auf chemische Probleme*, 2 Bde. Berlin, Springer, 1935–36.

III. General references.

20. Barbier, D., et Chalonge, D., *De la stratosphère à l'ionosphère*. Paris, Presses Universitaires de France, 1942.
21. Fleming, J. A., ed., *Terrestrial Magnetism and Electricity*. New York, McGraw, 1939.
22. Kuiper, G. P., ed., *The Atmospheres of the Earth and Planets*. Chicago, Chicago University Press, 1949.
23. Mitra, S. K., *The Upper Atmosphere*. Calcutta, Royal Asiatic Society of Bengal, 1948.

IV. Summarizing reports and discussions, containing extensive detailed bibliographies.

24. Bates, D. R., "The Earth's Upper Atmosphere." *Mon. Not. R. astr. Soc.*, 109: 215–245 (1949).
25. Déjardin, G., "The Light of the Night Sky." *Rev. mod. Phys.*, 8: 1–21 (1936).
26. Gassiot Committee, *Reports on Progress in Physics*, 9: 1–100. Phys. Soc., London, 1943.
27. Gassiot Committee, *The Emission Spectra of the Night Sky and Aurora*. Phys. Soc., London, 1948.
28. Paneth, F. A., "The Chemical Composition of the Atmosphere." *Quart. J. R. meteor. Soc.*, 63: 433–438 (1937).

V. Articles cited in text.

29. Bates, D. R., and Massey, H. S. W., "The Basic Reactions in the Upper Atmosphere. II—The Theory of Recombination in the Ionized Layers." *Proc. roy. Soc.*, (A) 192: 1–16 (1947).
30. Bates, D. R., and Seaton, M. J., "The Quantal Theory of Continuous Absorption of Radiation by Various Atoms in Their Ground States, II." *Mon. Not. R. astr. Soc.*, 109: 698–704 (1950).
31. Hirschfelder, J. O., "Semi-empirical Calculations of Activation Energies." *J. Chem. Phys.*, 9: 645–653 (1941).
32. Livingston, M. S., and Bethe, H. A., "Nuclear Physics. C. Nuclear Dynamics, Experimental." *Rev. mod. Phys.*, 9: 245–390 (1937). (See p. 370, Table LXX)
33. Meinel, A. B., "*OH* Emission Bands in the Spectrum of the Night Sky." *Astrophys. J.*, 111: 555–566 (1950).

OZONE IN THE ATMOSPHERE*

By F. W. PAUL GÖTZ

Lichtklimatisches Observatorium Arosa and University of Zürich

INTRODUCTION

As early as 1845 the chemist Schönbein, discoverer of ozone, attempted to prove that traces of this gas are a constant constituent of the atmosphere. However the ozone problem in its entire scope was revealed only when Fabry and Buisson [32] recognized by spectrographic methods that the principal quantities of ozone are present in the upper strata of the atmosphere. Although the ozone amount, that is, the total quantity of ozone present when reduced to standard conditions, amounts to but a few millimeters of the 8-km height to which the homogeneous ocean of air rises, this small trace is nevertheless sufficient (because of the enormous absorption in the ultraviolet) for intercepting the entire extraterrestrial radiation below 2900 A, for heating the "warm layer" at a height of 50 km, and for protecting the biosphere against an excess of short-wave radiation. Only during recent years has it been possible to pass through the high ozone layer with the aid of V-2 rockets, a feat which opened an entirely new spectral range of solar radiation to astrophysical investigations. But how is it possible for such a relatively heavy component of air to maintain itself high in the atmosphere? Only if it is constantly being regenerated, if short-wave radiation from the sun is capable of transforming oxygen to ozone. The opposing effect of long-wave ultraviolet which is again absorbed by ozone thereby destroying it (E. Regener), produces a spontaneous chemical ozone equilibrium in the upper layers of the atmosphere; below 35 km, however, this latter process is so slow that ozone which, by some meteorological processes, has been transported to these altitudes is largely "protected." In lower layers ozone becomes an important conservative property of the air until the stream from the high altitude ozone source is finally destroyed near ground level due to photochemical, chemical, and catalytic decomposition. In view of its intimate connection with weather processes (Dobson), the study of ozone provides methods of indirect aerology; there is an ever-increasing tendency to supplement a mere observation of total ozone amount with a detailed measurement of its vertical distribution and changes thereof. Only by investigation of this latter type will it be possible to understand the effect of ozone on the flux of radiation.

An intriguing aspect of the ozone problem is that the most diverse scientific fields converge at it. Two international conferences [16, 31] and several national conferences [71] have so far been devoted to this problem. The Meteorological Association of the International Union for Geodesy and Geophysics has a special commission on ozone.

* Translated from the original German.

METHODS OF OBSERVATION

Absorption Coefficients. *Ozone* has its principal absorption in the Hartley band [68] extending from 3200 to 2000 A. At $\lambda = 2553$ A the decadic absorption coefficient is 145, that is, a layer as thin as $\frac{1}{145}$ cm reduces the intensity of radiation to one tenth. On the short-wave side, in the region of the ozone-oxygen gap which is so important for the theory, Mme. A. Vassy [96] satisfactorily verified the older measurements of Edgar Meyer. The long-wave side of the Hartley band connects with the Huggins bands which extend up to 3690 A, and these bands are temperature dependent. However, new determinations [4] yielded the opposite finding that the absorption maxima too are influenced by temperature. The Chappuis bands, although relatively weak, coincide with the region of maximal solar energy; because of the superposition of the vapor bands in the solar spectrum, they are not suitable for ozone determination, according to Mme. Vassy. Absorption in the long-wave region is important for the problem of heat balance; the absorption band between 9.0 and 9.7 μ seems particularly significant because it lies in the otherwise very transparent atmospheric window between 9 and 13.5 μ. According to Strong [92], and Adel and Lampland [1], absorption in the 9.6-μ band exceeds all previous assumptions by a factor of about ten and is very dependent on pressure, approximately proportional to the fourth root of pressure, a fact which Strong uses for determining the average height of the ozone layer by simultaneous determination of the ozone absorption in the ultraviolet and the infrared.

Oxygen has absorption properties which form the basis for the theory of the ozone layer. In this connection, the region of maximal absorption between 1300 and 1750 A [56] is of less interest than the wave lengths around 2000 A [96] which overlap the ozone absorption. Deviations from Beer's law are significant. I am indebted to W. Heilpern and E. Meyer [51] for information regarding the current state of this question. If the absorption of light is given by $J = J_0 \times 10^{-\epsilon l}$, where ϵ denotes the extinction coefficient for a pressure P in kilograms per square centimeter, and if the thickness of the layer $l = 1$ cm, then ϵ for pure oxygen is given by

$$\epsilon(O_2) = \epsilon_1 P + \epsilon_2 P^2. \qquad (1)$$

For an arbitrary mixture of O_2 and N_2, where c_1 denotes the volume concentration of O_2 and c_2 that of N_2, ϵ is given by

$$\epsilon(G) = \epsilon_1 P c_1 + \epsilon_2 P^2 c_1^2 + \epsilon_3 P^2 c_1 c_2 , \qquad (2)$$

where P denotes the total pressure. The third term represents the effect of the "extraneous gas." In the region investigated, the constants ϵ_1, ϵ_2, and ϵ_3 have

the following values:

λ(A)	$\epsilon_1 \times 10^5$	$\epsilon_2 \times 10^{-5}$	$\epsilon_3 \times 10^{-5}$
2100	12.82	10.50	4.07
2144	6.17	9.30	4.01
2150	6.70	8.81	3.70
2200	5.49	7.21	3.25
2250	4.10	5.73	2.86
2300	2.63	4.31	2.57
2350	2.69	3.10	1.87
2400	1.47	2.10	1.37

The formulas given above agree surprisingly well with observed data and are valid in the large pressure interval between $P = 0.20$ and 130 kg cm^{-2}, in which ϵ varies by four orders of magnitude.

Determination of the Ozone Amount. At first, we are interested in the total atmospheric ozone amount, which is customarily expressed in terms of the thickness x of a layer under standard conditions of temperature and pressure. The decadic extinction coefficient of the total, pure atmosphere above the observer, according to Rayleigh, may be denoted by β; the decadic extinction coefficient of the total turbid atmosphere may be denoted by δ, and that of ozone by αx; correspondingly, let the total path of the sun's rays, having a zenith distance z be m according to Bemporad (for the turbid layer, $m = \sec z$ because the turbid layer generally lies on the ground), and let $\mu = \sec \zeta$ for the ozone layer, with

$$\sin \zeta = \frac{\sin z}{1 + h/R}, \tag{3}$$

where h is the height of the ozone layer and R the radius of the earth. For radiation of the extraterrestrial intensity I_0 subject to ozone absorption in the Hartley band, we obtain

$$\log I = \log I_0 - \beta m - \delta \sec z - \alpha x \mu; \tag{4}$$

for a neighboring wave length ($\lambda' > \lambda$) subject to lesser absorption by ozone, we find

$$\log I' = \log I_0' - \beta' m - \delta' \sec z - \alpha' x \mu. \tag{5}$$

From equations (4) and (5), we obtain the amount of ozone as

$$x = \frac{\log I_0/I_0' - \log I/I'}{(\alpha - \alpha')\mu} - \frac{(\beta - \beta')m}{(\alpha - \alpha')\mu} - \frac{(\delta - \delta')\sec z}{(\alpha - \alpha')\mu}. \tag{6}$$

The quantity $(\delta - \delta')$ may be neglected except in the case of dense haze [42, 79]. The constant $\log (I_0/I_0')$ is then obtained by plotting diurnal measurements of $\log (I/I') - (\beta - \beta')\, m$ as ordinate versus μ as abscissa and extrapolating to $\mu = 0$.

The experimental apparatus must be of such construction that scattering within the spectrograph is avoided since this would be a serious disturbance, particularly in view of the pronounced intensity drop at the end of the spectrum. Dobson's spectrophotometer [22] is particularly designed to avoid such errors. By means of a rotating sector, I and I' are made to fall, in rapid alternation, upon a photoelectric cell connected to an amplifying device. By a measured attenuation with an optical wedge, I' can be made equal to I, a condition indicated by a zero amplifier output; in this manner, the ratio I/I' is measured. As calibration runs in Arosa and on the Jungfraujoch have indicated, the wave-length adjustment of the instrument depends not only on temperature but also on altitude, since the refractive index of quartz with respect to air changes with air pressure. Values obtained with the old photographic Dobson instrument [24] must be reduced to 88 per cent of the indicated value [40]. For the purpose of increasing the sensitivity of the Dobson spectrophotometer, it has recently been the practice to replace the photoelectric cell by a photo-multiplier. Figure 1 shows the latest model of the instrument according to the catalog of its manufacturer, Beck of London [70].

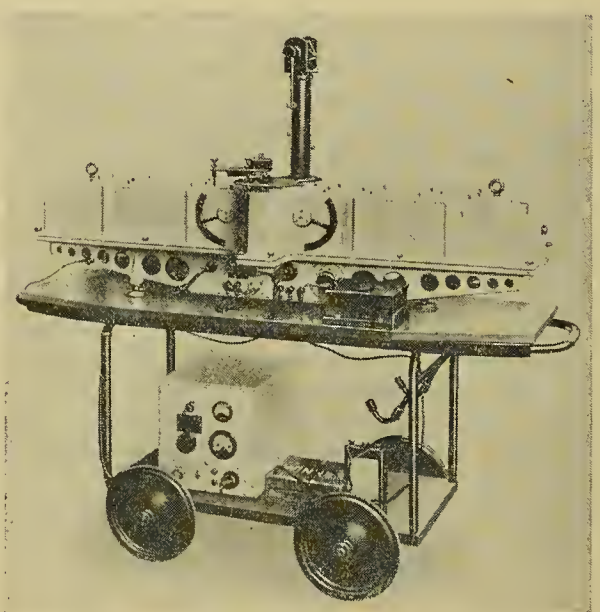


FIG. 1.—The Dobson Spectrophotometer.

The ozone near ground level is determined by essentially the same method, employing in place of the sun, moon, or stars [13] an artificial light source [44] rich in ultraviolet and located at a distance of several kilometers. Improved chemical ozone determinations have recently gained extensive use [12, 20, 29, 73, 87, 91].

Determination of the Vertical Distribution. Two methods are available for determining the vertical distribution of ozone.

1. *The Umkehr Effect.* In the case of solar radiation, ozone absorption as well as scattering is responsible for the fact that with increasing zenith distance z, the intensity I of a shorter wave length (for instance $\lambda = 3110$ A) decreases more strongly than the intensity I' of a longer wave length (for instance $\lambda' = 3290$ A); the composition of light I/I' which we may also designate as light quality, decreases continuously as the zenith distance of the sun increases. For zenith radiation i, the ratio i/i' shows a similar behavior at first, as long as the zenith distance of the sun is relatively small;

as z continues to increase, however, the function i/i' reaches a minimum at about $z = 85°$, and shows an inversion (Umkehr) [36] at $z > 85°$, that is, i/i' increases again at very low elevations of the sun. This experimental result (Fig. 2) is explained on the basis of

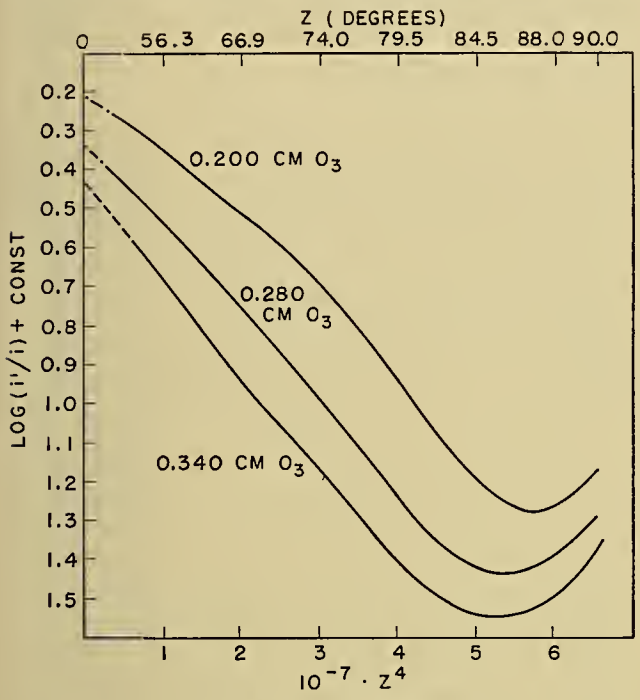


Fig. 2.—Umkehr curves.

the distribution of ozone in the scattering atmosphere. Zenith light is the sum of the components scattered at various altitudes corresponding to the prevailing air density. Because of its smaller density, the atmosphere above the ozone layer scatters to a lesser extent than the atmosphere below this layer. However, the short-wave radiation scattered vertically downward traverses the attenuating ozone layer by the shortest path and, as the zenith distance increases, it attains ever more significance in comparison with the radiation scattered below the ozone layer; in this case scattering power would indeed be greater, but the sunlight which is to be scattered has been considerably weakened by traversing the ozone layer by the long, oblique path. The stronger the absorption, the greater the tendency for the sky radiation of the shortest wave length to find the shortest path directly from above through the ozone layer. The variation of the Umkehr function i/i' with z depends on the type of stratification, that is, the vertical distribution of ozone. It thus provides a means for measuring the vertical distribution.

We divide the atmosphere into a number of shells within each of which we assume the ozone content (expressed in centimeters of ozone per kilometer) to be constant. We might assume the following shells [39, 45]:

Altitude	Ozone amount
65–50 km	0, but scattering not zero
50–35 km	x_1
35–20 km	x_2
20– 5 km	$x - (x_1 + x_2 + u)$ since x is known.
5– 0 km	u, where u is assumed to be known.

Thus the characterization of the vertical distribution involves only two unknowns, x_1 and x_2, which is desirable in order that the numerical solution of the equations may not be too difficult. The unknowns x_1 and x_2 enter into the expression for the thickness l cm of an ozone layer which is traversed by a sun ray incident at a zenith distance z. The ray is scattered at point A where the barometric pressure is b and reaches the instrument at the point B. The path length in each ozone shell is determined trigonometrically; it suffices to tabulate the distance s of Fig. 3 as a function of the

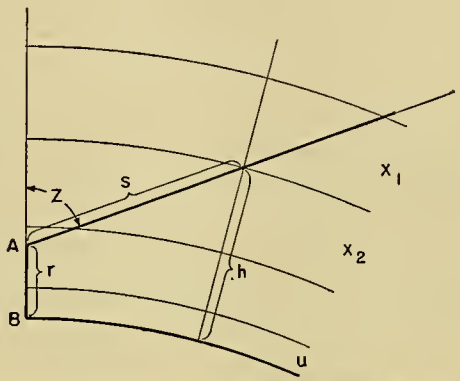


Fig. 3.—Subdivision of the atmosphere into layers.

quantity $h - r$ (*i.e.*, independently of r). Of significance for the Rayleigh scattering is the air mass L traversed by the ray, which is easily found to be

$$L = 1 + (b/760)\,(m - 1). \tag{7}$$

We sum the light scattered by each kilometer A from the ground to an altitude of 65 km and find the intensity of zenith radiation i to be

$$i = \Sigma 10^{-\beta L} b 10^{-\alpha l} \text{const} \tag{8}$$

and similarly for i' and i/i' or i'/i. The corresponding expression for i'/i is the same as the result of observations (Fig. 2). It is to be noted that the undetermined constant in the observed curve may be determined by calculating i'/i for $z = 0$, since the vertical distribution of ozone has no effect for $z = 0$. The two unknowns x_1 and x_2 require two equations, such as an equation for $z = 90°$ and one for $z = 80°$. The solution is obtained graphically by plotting x_2 as a function of x_1 for both cases and determining the intersection of the two curves. Choice of a third value of z, such as $z = 86.5°$, provides a convenient check. The values of x_1 and x_2 then characterize the vertical distribution stepwise.

In addition to this analytical method A, use has also been made of a more synthetic method B, which consists of starting with an assumed vertical ozone distribution and varying it until calculated and measured Umkehr curves are in agreement. In both methods secondary scattering should be further considered.

In order to obtain a more detailed ozone distribution curve it is, of course, desirable to choose our shells as thin as possible. It has been verified that essentially no ozone is present above 50 km. Between 50 and 35 km the ozone may be assumed constant as will be seen

later. From the ground upwards, the region of known ozone concentration could today easily be extended from 5 to 10 km above ground by modern aerological methods (chemical measurements by airplane). A subdivision of the region between 10 and 35 km into three shells would then remain for the Umkehr measurement. Day-to-day measurement of these three layers would be of the most important meteorological interest.

2. *Techniques Using Recording Balloons, Stratospheric Balloons, and V-2 Rockets.* In addition to the rather indirect method involving the Umkehr effect, there is of course the challenge of sending spectrographs to ever higher altitudes until finally the ozone layer is penetrated. The first great success in this respect was achieved by E. and V. H. Regener on July 31, 1934 when they sent a recording spectrograph to an altitude of 31 km by means of a balloon [85]. The lightweight spectrograph was pointed downward against a horizontal magnesium oxide disk exposed to sunlight; magnesium oxide effectively reflects ultraviolet. The apparatus is installed in a light wooden frame, the lower half of which is covered with aluminum foil, the upper half with cellophane, an arrangement which affords excellent protection against cold. The photographic plate Fig. 4) is rotated through a small angle every ten

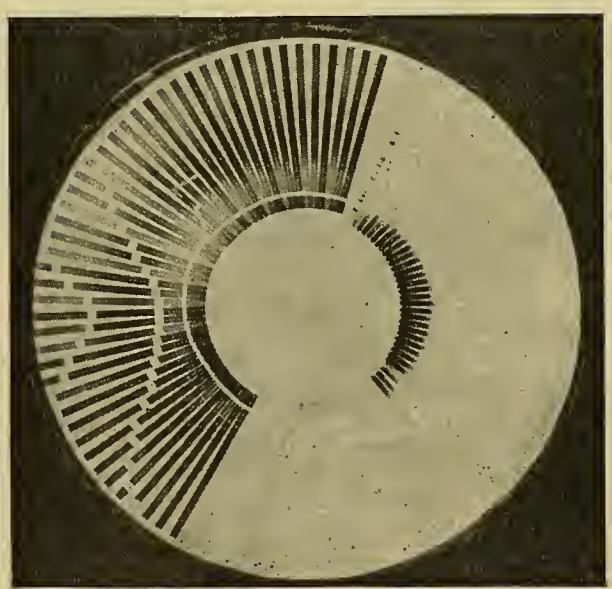


Fig. 4.—Regener-photographs of the ultraviolet spectrum taken in the stratosphere.

minutes, being continuously exposed in the meantime. At the instant the plate is advanced, a small light bulb is turned on by which the readings of two aneroid gauges (standard barograph and low-pressure barograph) as well as of a bimetallic strip to indicate the temperature of the instrument are recorded. On the plate which has a diameter of 10 cm the step-shaped interruptions of the long rays are the shadows of the two indicators connected to the aneroids; they represent lower pressures the closer they approach the center. The clear ultraviolet spectra are located diametrically opposite the pressure indications and they are seen to extend to shorter wave lengths the greater the altitude and the smaller the quantity of ozone traversed by the solar radiation. In addition, the plate also bears a mercury calibration spectrum. The entire device, including a protective frame, weighs 2.7 kg. If λ_n represents the shortest wave length of the solar spectrum at an altitude h_n and zenith distance z_n, and if x_n represents the total quantity of ozone located above the instrument, then, approximately,

$$x_2\alpha_2 \sec z_2 = x_1\alpha_1 \sec z_1 . \qquad (9)$$

Further improvements are obtained with an accurate intensity measurement at the end of the spectrum and elimination of diffuse sky light, for instance with the aid of a magnetically oriented hemispherical stop [86].

Coblentz and Stair [15] and Stranz [91] attempted to simplify the methods by replacing the spectrograph by a cadmium cell and filter as was done in the first ozone measurements in 1921 at Arosa. The intensity values are transmitted radiotelegraphically as in the radiosondes.

With the aid of a free balloon, A. Wigand as early as 1913 measured the ultraviolet end of the solar spectrum up to a height of 9 km. Shortly after the Regener experiments, the stratosphere balloon, which is capable of carrying aloft high quality instruments, was employed for this purpose. Under the leadership of Stevens and Anderson, *Explorer II* reached an altitude of 22 km in November 1935; ozone at even higher altitudes was determined indirectly from the sky radiation [39, 69]. Ozone determination by means of captured V-2 rockets [95] represents the climax of this brilliant development. It was the fulfillment of a dream when on October 10, 1946 in White Sands the ozone layer was penetrated and a large unknown range of the solar spectrum became accessible (Fig. 5).

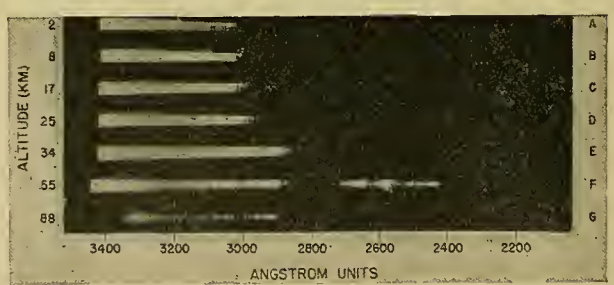


Fig. 5.—Short-wave end of the solar spectrum taken from V-2 rocket.

RESULTS OF OBSERVATIONS

The Ozone Amount. The first ozone determinations made in Marseille [32], as well as the measurements made at Arosa with filtered cadmium cells since 1921, revealed irregular ozone fluctuations from *day to day.* Dobson [24] discovered their connection with the general distribution of air pressure so that they may be designated as *meteorological* fluctuations. The interdiurnal variability of the ozone amount at Arosa was found to be:

	cm		cm		cm
January	0.017	May	0.011	September	0.008
February	0.017	June	0.010	October	0.009
March	0.015	July	0.010	November	0.010
April	0.015	August	0.009	December	0.013

During the summer season there are no great meridional differences in the interdiurnal variability. In July and August, in India it is 0.005 cm, at Tromsö 0.010 cm, at Spitsbergen 0.007 cm; in December it rises to 0.056 cm at Tromsö. The fluctuations frequently become manifest even during a single measuring day, in the form of an "ozone cloud" [55]. In one case in which ozone was determined by sighting on stars in various directions of the sky, it was possible to delineate such

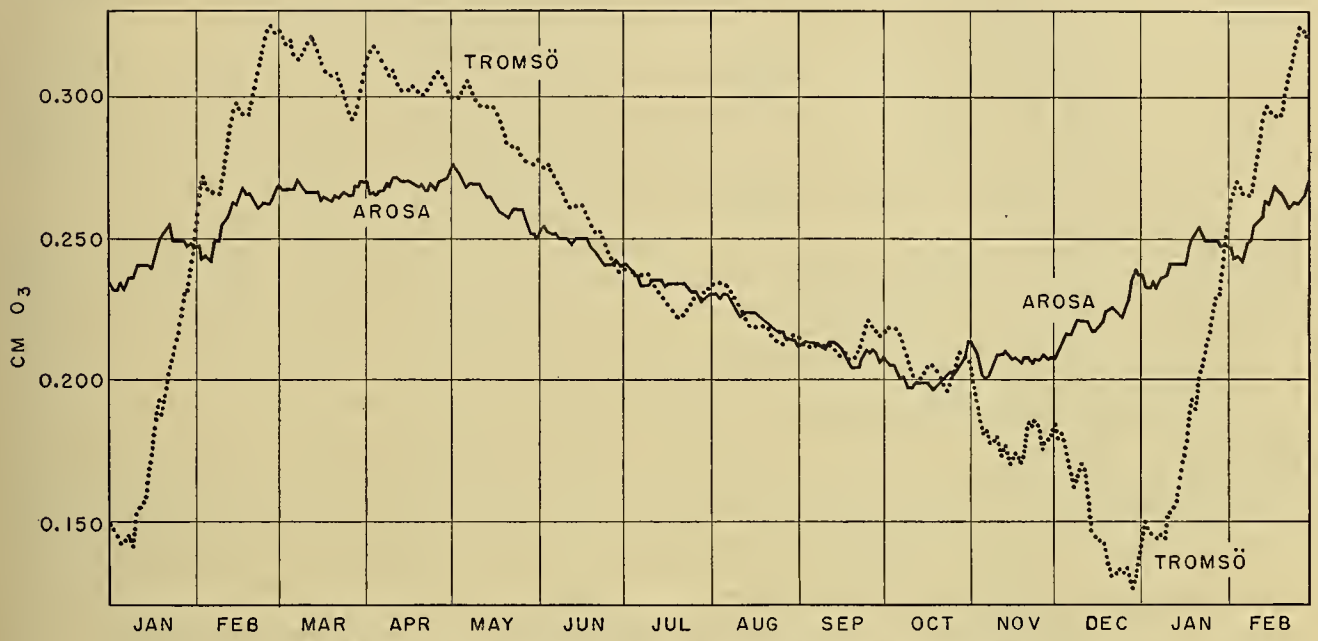


Fig. 6.—Annual variation of the ozone amount at Arosa and Tromsö.

an ozone cloud in space [3] as a forerunner of a considerable increase in ozone. A regular, slight *daily* variation of ozone has so far been reported only from Delhi [53]; at Arosa, a change in the temperature correction of the Dobson spectrophotometer simulated a diurnal variation for some time. Differences between day and night have not yet been confirmed. Frequently ozone fluctuations are of periodic nature [43] with periods up to 36 days, as has been found for the barometric pressure at the ground.

Concerning the *annual* variation of ozone amount, we are quite well informed by the extensive observational network of Dobson [21]. The representation as a yearly sine curve, with an amplitude which increases with latitude from a value of zero at the equator, is a considerable simplification even at middle latitudes. Figure 6 represents the overlapping five-day averages of the extensive measurements at Arosa (1926–1946) and Tromsö (1939–1948). At Tromsö an almost sudden linear increase in January and February is followed by a gradual decrease terminating in an "ozone gap" at the end of December. It seems doubtful that all the peaks in the Arosa curve, such as those found at the end of April or the end of October, will be smoothed out by averaging more extensive observational material [43].

The latitudinal dependence of the annual averages of ozone amount is revealed by an increase from 0.17 cm at the equator to an "ozone belt" (Götz) of 0.26 cm at latitude 60°N; the ozone amount again decreases toward higher latitudes. In the Southern Hemisphere, the gradient is more pronounced, and here again it is the maxima rather than the minima that are significant. The variation is best described by an isoplethic representation (Fig. 7), in which the ozone belt is shown by means of a dot-dash line.

As regards the *secular* variation, that is, ozone variation from year to year, Fowle's hypothesis concerning the correlation with the relative sunspot number was

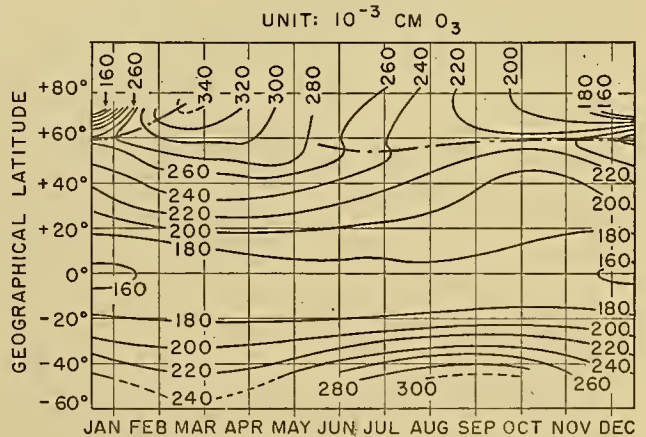


Fig. 7.—Isopleths of the average ozone amount. The ozone belt is indicated by dot-dash lines.

untenable. Figure 8 shows the secular variation for the twenty-year series at Arosa, which reveals waves of several years' length. For our further conclusions we might wish to have such figures back to ice ages! The mean monthly fluctuation about the over-all monthly

average is given below:

	cm		cm		cm
January	±0.014	May	±0.008	September	±0.007
February	±0.014	June	±0.006	October	±0.007
March	±0.012	July	±0.005	November	±0.005
April	±0.012	August	±0.006	December	±0.008

It is apparent that the fluctuation is twice as great in winter as it is in summer and fall; in winter, as well as in spring, it is considerably greater at Tromsö than it

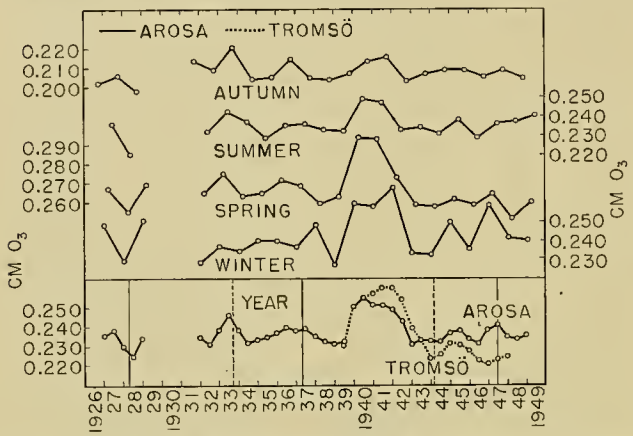


FIG. 8.—Secular fluctuation of the ozone amount. Vertical broken lines represent sunspot minima; vertical solid lines, sunspot maxima.

is at Arosa; the series is thus subject to the same influence which also causes the meteorological (interdiurnal) variability. If one were to seek a direct correlation between ozone and solar activity, then the double fluctuations of the Eurasian large-scale weather within the sunspot cycle according to Baur [7] would be of interest. In the lower yearly average curve of Fig. 8, the vertical solid lines represent sunspot maxima, the vertical broken lines represent sunspot minima. For the yearly averages the following correlation coefficients are obtained.

	Correlation coefficient
Ozone and relative sunspot number	+0.01
Ozone and air pressure	−0.43
Ozone and solar constant	+0.48

As regards the analysis of other long series of observations, there exists only the attempt [11, 93] to determine ozone by means of the Chappuis band from the Smithsonian measurements. Even though wide scattering of these values makes them seem rather unreliable, it is nevertheless interesting to note the very slight ozone amount during the year 1912 since it has been ascribed to an effect of the eruption of the Katmai Volcano [37]. Such a possibility has also been discussed by E. Regener [84].

The Vertical Distribution of Ozone. The vertical distribution of ozone, first of all shows a high primary layer of maximum ozone content at an altitude of 20 to 25 km. Its stratified character is more pronounced the smaller the total ozone amount. The layer thickens as the ozone amount increases, owing to meteorological

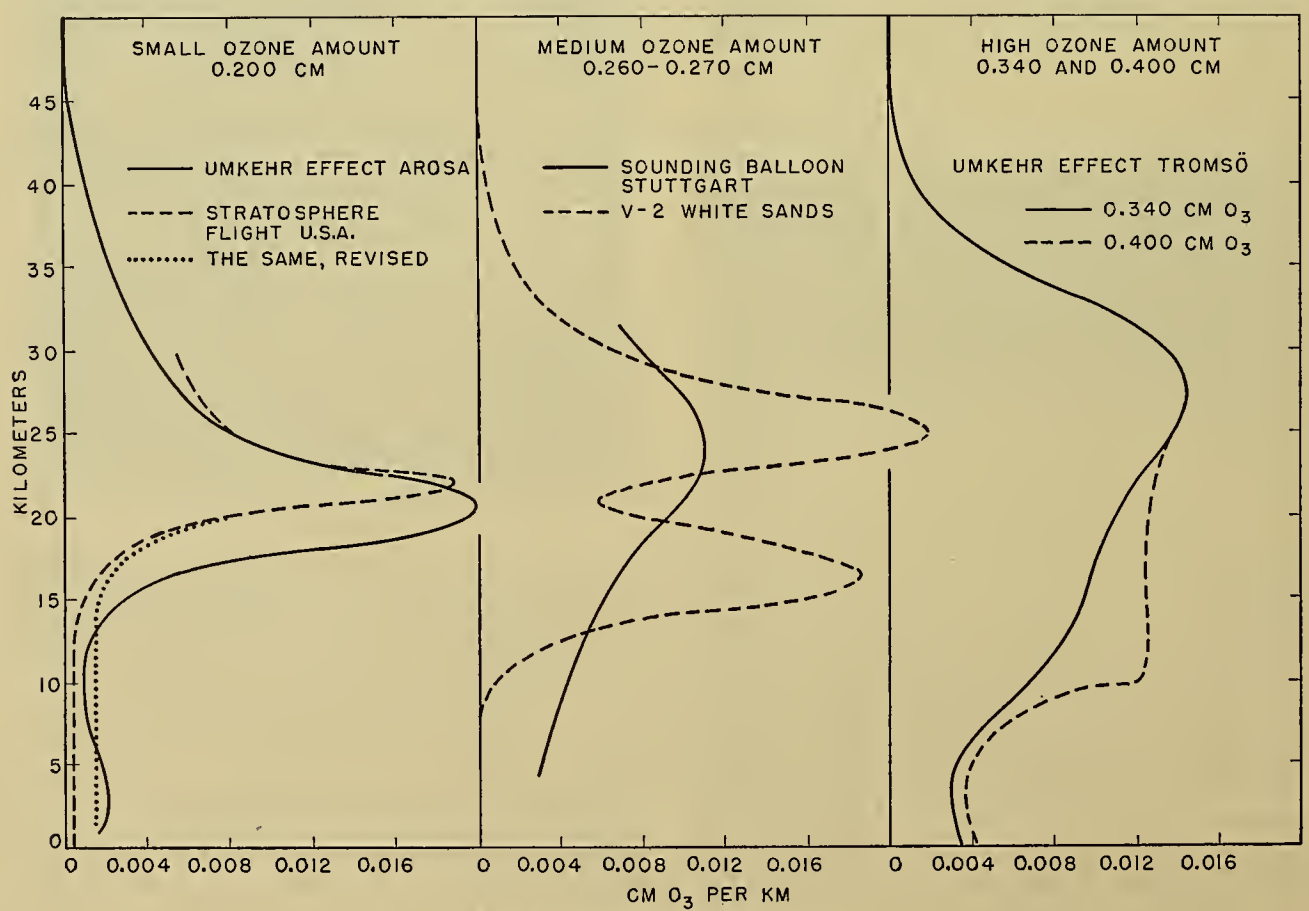


FIG. 9.—Vertical ozone distribution.

processes. It is not impossible that, on the one hand, the primary layer moves to slightly higher altitudes; on the other hand, it is principally of meteorological significance that a lower secondary layer [39] appears which swells with increasing ozone amount, so that the original separation of the two layers becomes more and more indefinite. Representation by the direct stepwise distribution according to the Umkehr method may be most instructive in this connection [39]. Figure 9 shows smoothed curves for several examples, including the result of the first vertical ozone determination with V-2 rockets in which the lower secondary layer was also found.[1] Table I summarizes in customary manner

TABLE I. ALTITUDE OF THE MASS CENTER OF OZONE (in km)

Station	Lat.	Method	Ref.	Ozone amount (cm O_3)				
				0.160	0.220	0.280	0.340	0.400
Poona	18°	*B*	[54]	28.1				
Delhi	29°	*B*	[54]	26.2	25.1			
White Sands	35°	*V*-2	[66]			22.5		
Arosa	47°	*A*	[45]		23.4	21.5	21.3	
Arosa	47°	*B*	[45]		22.6	21.4		
Tromsö	70°	*B*	[62]		21	21		
Tromsö	70°	*A*	[94]	26.7	25.7	24.7	23.0	20.8

the mass center of ozone; there *A* denotes the analytical, *B* the synthetic evaluation of the Umkehr curve. The variation of the center of mass of ozone with latitude evidently requires additional measurements.

The lowest portion of the vertical ozone distribution may be determined by direct measurement of the ozone content near ground level. Whereas in the lowlands the ozone near ground level is subject to considerable

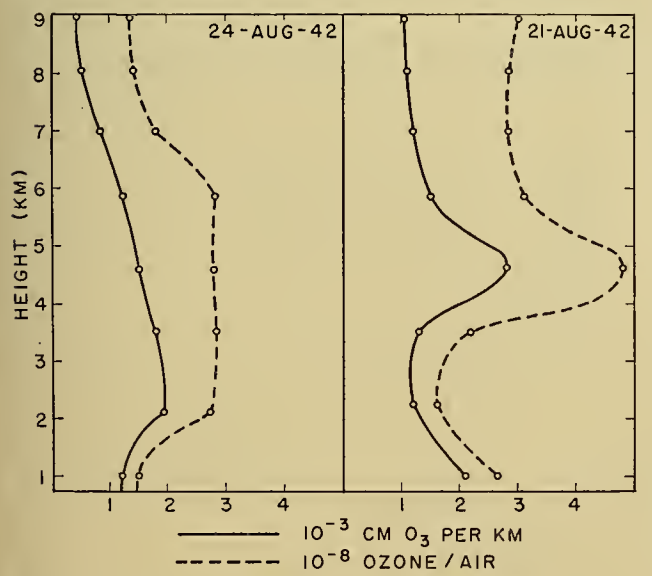


FIG. 10.—Tropospheric ozone distribution.

fluctuation [2], it shows considerable constancy at Arosa which is situated at an altitude of 2 km; thus there appears to be a weak tertiary ozone layer at this altitude. As shown by aircraft measurements by Ehmert [28] the altitude of this tertiary tropospheric layer may vary in individual cases, depending on meteorological conditions (Fig. 10).

THE THEORY OF ATMOSPHERIC OZONE

The Ozone in Undisturbed Photochemical Equilibrium. The existence of a layer of a heavy gas in the upper atmosphere necessitates the assumption of continuous regeneration and destruction with consequent equilibrium. As early as 1906, E. Regener demonstrated by means of laboratory experiments the photochemical equilibrium of ozone under the influence of the short-wave ultraviolet radiation which, when absorbed by oxygen, generates ozone, and also under the influence of a radiation that is then again absorbed by the ozone. This clearly provides a basis for a photochemical equilibrium in the atmosphere under the action of solar radiation. No other theory provides an equally adequate explanation even though it cannot be denied that occasionally electrical storms and cosmic radiation might contribute some ozone. At the altitude of the equilibrium layer, radiation in the wave length region between the Schumann-Runge and Herzberg bands produces ozone, whereas the wave length of the Hartley band particularly, but also those of the Chappuis band, destroy ozone. The various wave lengths involved are absorbed very differently and therefore penetrate the atmosphere to different depths.

We are indebted to S. Chapman for the first theory of atmospheric ozone based on these considerations. Wulf and Deming, to whom more extensive data were available [103], worked on the basis of Chapman's fundamental reactions. There are two primary photochemical reactions, namely,

$$O_2 + h\nu(\lambda < 2420) \rightarrow O + O, \tag{10}$$

involving the number Q_2 of oxygen-dissociating quanta $h\nu$, and

$$O_3 + h\nu(\lambda < 11540) \rightarrow O_2 + O, \tag{11}$$

involving the corresponding number Q_3 of ozone-destroying quanta. The quantities Q_2 and Q_3 are the numbers of quanta absorbed by oxygen and ozone, respectively, in a given unit volume of air. For example, Q_2 is given by

$$Q_2 = \int_\lambda \alpha_{2\lambda} I_\lambda [O_2]\, d\lambda,$$

where $\alpha_{2\lambda}$ is the absorption coefficient of oxygen, I_λ is the number of quanta incident on the volume, and the brackets about O_2 indicate (here and in subsequent discussion) the concentration of oxygen. In addition to (10) and (11) there are the secondary reactions.

$$O_2 + O + M \rightarrow O_3 + M, \tag{12}$$

which requires a triple collision with an arbitrary collision partner M for conserving energy and momentum, and for which k_2 represents the coefficient of the reaction rate, and finally,

$$O_3 + O \rightarrow 2O_2^*, \tag{13}$$

1. The reality of this lower-altitude maximum has been questioned in a recent paper: "Upper Air Research by Rockets," by H. E. Newell, Jr., in *Trans. Amer. geophys. Un.*, 31: 25-34 (1950). (See p. 31)

with k_3 as the constant of the rate of reaction (O_2^* is an excited molecule). The constants k_2 and k_3 enter only in the form of the ratio $k_2/k_3 = k$; the data for k are as yet somewhat variable. If the concentration [ ] is referred to the number of molecules per cubic centimeter, the extrapolated measurements by Eucken and Patat [30] yield the values of k shown in Table II.

TABLE II. VALUES OF k

deg C	k	deg C	k
100	1.95×10^{-21}	0	4.00×10^{-20}
80	3.16×10^{-21}	−20	9.95×10^{-20}
60	5.27×10^{-21}	−40	2.79×10^{-19}
40	9.51×10^{-21}	−60	9.85×10^{-19}
20	1.88×10^{-20}	−80	4.35×10^{-18}

From the equations (10) to (13) and the reaction constants we obtain the changes in concentration per second:

$$\frac{d[O_3]}{dt} = k_2\,[O_2]\,[O]\,[M] - k_3\,[O_3]\,[O] - Q_3 \qquad (14)$$

and

$$\frac{d[O]}{dt} = 2Q_2 + Q_3 - k_2\,[O_2]\,[O]\,[M] - k_3\,[O_3]\,[O]. \qquad (15)$$

If we calculate $[O]$ from the condition of equilibrium $\frac{d[O]}{dt} = 0$, we obtain

$$\frac{d[O_3]}{dt} = \frac{2Q_2\,k_2\,[O_2]\,[M] - 2Q_2\,k_3\,[O_3] - 2Q_3\,k_3\,[O_3]}{k_2\,[O_2]\,[M] + k_3\,[O_3]}, \qquad (16)$$

whence, under equilibrium conditions, $\frac{d[O_3]}{dt} = 0$, it follows that

$$[O_3] = \frac{k_2}{k_3}\frac{[O_2]\,[M]}{1 + Q_3/Q_2} = k\,[O_2]\,[M]\,\frac{1}{1 + Q_3/Q_2}. \qquad (17)$$

Wulf and Deming base their numerical calculations of the vertical ozone distribution upon this equation, in which Q_3 itself is a function of $[O_3]$, so that one can only proceed by a method of successive approximations.

In addition to oxygen, the third collision partner M is primarily nitrogen which, however, is only about one-half as effective as oxygen so that

$$[M] = [O_2] + \tfrac{1}{2}[N_2] \sim 3[O_2]. \qquad (18)$$

We may replace Q_3, the number of quanta absorbed per second per cubic centimeter (which is also proportional to $[O_3]$), by

$$Q_3 = [O_3]f_3\,, \qquad (19)$$

where f_3 is the number of quanta absorbed per molecule, and correspondingly

$$Q_2 = [O_2]f_2\,. \qquad (20)$$

Dütsch [27], using this resolved form (which facilitates subsequent treatment of nonequilibrium conditions), obtains a quadratic equation for $[O_3]$:

$$[O_3]^2 f_3 + [O_3][O_2]f_2 - 3k[O_2]^3 f_2 = 0. \qquad (21)$$

with the solution

$$[O_3] = [O_2]\,\frac{-f_2 + \sqrt{f_2^2 + 12\,f_3\,f_2\,[O_2]k}}{2f_3} \qquad (22)$$

and, with the justified approximation $f_2 \ll 12\,f_3 f_2 [O_2]k$,

$$[O_3] = [O_2]\sqrt{3kf_2[O_2]/f_3}. \qquad (23)$$

This equation does not differ much from the form

$$[O_3] = [O_2]^2\sqrt{Cf_2/f_3}, \qquad (24)$$

which served as basis for Mecke's theory [60] as early as 1931. Mecke replaced the numerical integrals f_2 and f_3 by $\alpha_2 I_2$ and $\alpha_3 I_3$, thus taking not only average radiation intensities but also average constant mean values for α_2 and α_3. This enabled him to integrate over limited intervals, resulting in the following ozone distribution:

$$[O_3] = [O_3]_{\max}\left(\frac{p}{p_{\max}}\,e^{1-(p/p_{\max})}\right)^2. \qquad (25)$$

In this equation, according to recent analyses, the power 2 is to be replaced by the power $3/2$ which furnishes a somewhat broader maximum (Fig. 11). The symbol $p_{\max}$ denotes the pressure p at the altitude of maximum ozone concentration $[O_3]_{\max}$. When Mecke introduced the total amount x of ozone (instead of the maximum ozone content) he obtained for an altitude H of the homogeneous atmosphere the ozone content

$$\epsilon 10^{-5} = \frac{x}{1.85H}\left(\frac{p}{p_{\max}}\,e^{1-(p/p_{\max})}\right)^2. \qquad (26)$$

We might mention one conclusion [39] derived from Mecke's theory. If we start with the altitude of maximum oxygen absorption Q_2 (which, incidentally, is still dependent upon the zenith distance z of the incident solar radiation), then we find the altitude of maximum ozone content to be lower by the constant amount $H \ln 3 = 7$ km, independently of z. Figure 11 gives

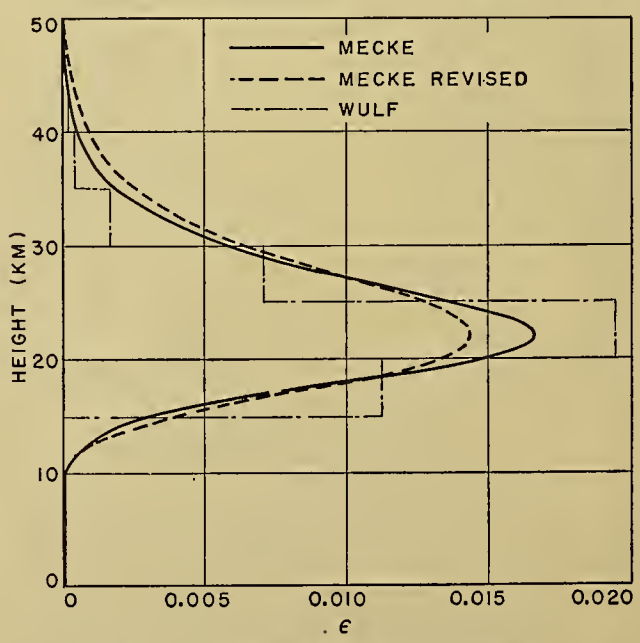


FIG. 11.—Theoretical ozone distribution.

Mecke's result, for which the total ozone amount and altitude of the maximum content are given, according to the original and to the revised formula, as well as according to Wulf and Deming's result, valid only for the sun at the zenith. Below 10 km the ozone content approaches zero because at lower altitudes the deozonizing effect of light becomes prevalent. Above 50 km the ozone content again approaches zero because the frequency of ozone-forming collisions decreases much more rapidly with altitude than does the oxygen concentration. The highly gratifying results from the analytic integration should not lead us to forget that Mecke's theory cannot be developed further in view of its inherent simplifications.

Wulf's treatments have been independently expanded by Schröer [89] and Dütsch on the basis of more modern theories. These make allowance first of all for the sun's elevation, that is, the effect of season and latitude. The ozone calculation is again performed layerwise, from the top downward with the aid of equation (23). New data are used for the seasonal and zonal vertical temperature variation in the stratosphere, for the ozone absorption coefficients in the vicinity of 2100 A, for the deviations of oxygen absorption from Beer's law, and for extraterrestrial solar radiation. Whereas the results concerning vertical ozone distribution and total ozone amount are satisfactory, one finds a decrease of ozone from equator to pole and from summer to winter. Even if the effect of the daily variation of the sun's position as well as of Rayleigh scattering is considered, it is not possible to reverse this erroneous trend completely. It follows that the observed ozone distribution cannot be explained by assuming pure radiation equilibrium at every point of the atmosphere. Recently, Craig [17] has reached the same conclusion.

An explanation for the seasonal variation of ozone solely on the basis of temperature variation [49, 98] is not convincing.

Ozone under Conditions of Disequilibrium. The period of time required for photochemical equilibrium to be established was investigated quantitatively by Wulf [104]. Dütsch [27] gives the temporal trend of ozone concentration in the case of a disturbance; during the time $t_R = \frac{1}{4}\,(f_2 f_3/k[M])^{-\frac{1}{2}}$, the ozone concentration is reduced to approximately $1/e$ of its original value, corresponding to the "time of half restoration" used by Wulf. Within high altitude layers ($\geqslant$ 35 km) correction of the disturbed equilibrium takes place almost immediately; between 30 and 25 km from several days to several months are required; at low altitudes ("protected regions") equilibrium is practically never regained. At high elevation of the sun, equilibrium is established more rapidly than in the case of low sun. Schröer, too, emphasizes the conclusion that (in middle latitudes) photoequilibrium occurs only down to about 33 km, and that the ozone distribution below this altitude is determined by transport processes (turbulence).

The Effect of Air Motion on Ozone. Dütsch lets the subscript s denote the effect of disturbances, R represent the absolute gas constant, takes the molecular weight of air as 28.9, and finds the effect of air currents on ozone concentration to be

$$\left(\frac{\partial[O_3]}{\partial t}\right)_s = -\,v_z\left(\frac{28.9g}{RT}\,[O_3] + \frac{\partial[O_3]}{\partial z}\right) - \left(v_x\frac{\partial[O_3]}{\partial x} + v_y\frac{\partial[O_3]}{\partial y}\right). \tag{27}$$

This derivation is based on the simplifying assumptions that the air density varies only with altitude and that the vertical temperature gradient is zero, an assumption that is justified only as applied to the lower stratosphere. Reed [80] therefore took the vertical temperature gradient into account and obtained the expression

$$\left(\frac{\partial[O_3]}{\partial t}\right)_s = -\,v_z\left\{[O_3]\left(\frac{28.9g}{RT} + \frac{\partial \ln T}{\partial z}\right) + \frac{\partial[O_3]}{\partial z}\right\}. \tag{28}$$

He uses this equation primarily in treating the diurnal meteorological ozone fluctuations and in establishing a theory concerning seasonal and zonal ozone fluctuations. Dütsch, moreover, finds the effect of turbulent exchange (austausch) to be

$$\left(\frac{\partial[O_3]}{\partial t}\right)_a = \frac{[O_2]}{\rho}\,\frac{\partial}{\partial z}\left\{A\,\frac{\partial}{\partial z}\,\frac{[O_3]}{[O_2]}\right\}, \tag{29}$$

where the subscript a denotes the effect of the austausch and A represents the austausch coefficient. Diffusion may be neglected. Because of the long time required for establishment of photochemical equilibrium at lower altitudes, changes in the concentration owing to flow and turbulence play an essential role. If B is the change in ozone concentration per second caused only by air motion (flow and austausch) then equation (23) becomes

$$[O_3] = [O_2]\sqrt{\frac{f_2 + \dfrac{B}{2[O_2]}}{f_3}\,k\,[M]}, \tag{30}$$

and the equation for disequilibrium becomes

$$[O_3](t) = \frac{[O_3]\,(t_0)\sqrt{\dfrac{4\left(f_2 + \dfrac{B}{2[O_2]}\right)f_3}{K[M]}} + 2\left(f_2 + \dfrac{B}{2[O_2]}\right)[O_2]\tanh\left(\sqrt{\dfrac{4\left(f_2 + \dfrac{B}{2[O_2]}\right)f_3}{K[M]}}\,(t - t_0)\right)}{\sqrt{\dfrac{4\left(f_2 + \dfrac{B}{2[O_2]}\right)f_3}{K[M]}} + \dfrac{2f_3[O_3](t_0)}{K[O_2][M]}\tanh\left(\sqrt{\dfrac{4\left(f_2 + \dfrac{B}{2[O_2]}\right)f_3}{K[M]}}\,(t - t_0)\right)}. \tag{31}$$

Thus, equilibrium can be established only if $f_2 > B/(2[O_2])$; equation (31) is always valid.

Dütsch used this expression to recompute the seasonal variation of ozone at various latitudes. For the layers above 16 km, the austausch coefficient is calculated on the basis of experimental results of Paneth and

E. Regener concerning separation in the atmosphere; for the equatorial zone it is assumed that the turbulence extends to higher altitudes corresponding to the greater altitude of the tropopause. The result of this calculation indicates a constant downward stream of ozone caused by turbulence. It follows that ozone must constantly be destroyed in the layers near ground level. The mean ozone distribution in the lowermost 15 to 20 km is determined essentially by the austausch. The small amount of total ozone in the tropical zone (O_3 = 0.185 cm) is due to high-reaching turbulence combined with a rapid destruction of the downward transported ozone in the layers near the ground level, which seems plausible for tropical conditions. Disregarding the possibility of achieving a better agreement between theory (as indicated in Table III) and observation by making

Table III. Amount of Ozone According to Equation (31) (mm O_3)

ϕ	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
45°	2.02	2.05	2.09	2.11	2.07	2.00	1.97	1.96	1.96	1.99	2.01	2.00
60°	2.16	2.17	2.20	2.22	2.22	2.16	2.14	2.15	2.15	2.17	2.18	2.16
80°	1.73	1.73	1.77	1.85	1.88	1.91	1.86	1.89	1.76	1.75	1.73	1.72
Corresponding values, assuming a hypothetical circulation:												
60°	2.28	2.53	2.69	2.68	2.53	2.18	1.88	1.75	1.68	1.69	1.78	2.02

different assumptions concerning several as yet uncertain quantities entering the calculations, it is possible to improve agreement by assuming a meridional circulation. The last line of Table III is calculated under the assumption of an ascending flow in the higher latitudes of the summer hemisphere, and a descending flow in the higher latitudes of the winter hemisphere, the seasonal variations occurring primarily in the lower layers.

Another circulation hypothesis by Craig [17] has not yet been calculated. Reed [80], on the other hand, suggests an explanation of the seasonal and zonal ozone fluctuations based on the variable large-scale atmospheric turbulence with a maximum in winter and at a latitude of 60°. Between fall and spring, this is supposed to cause more ozone to be transported down into the protected regions than can be destroyed at ground level. During this period an increase in the total amount of ozone would consequently occur since the ozone transported to low altitudes is replaced at high altitudes. Between spring and fall, on the other hand, decomposition at ground level presumably predominates so that there is a drop in the total ozone content. The discussion by Schröer and Moser [71] will be dealt with in the following section.

OZONE AND WEATHER

Ozone in the Stratosphere. According to Dobson, the amount of ozone varies with the atmospheric pressure distribution, that is, the synoptic situation. In western Europe a maximum in the ozone amount is found on the rear side of a cyclone, a minimum of ozone is found above the southwest side of a high [25]. More recent studies concerning the connections between ozone amount and the surface weather map and fronts have supported this observation [23, 94] and have shown that the center of high ozone amount on the rear side of a cyclone is particularly the property of a newly developing depression (Fig. 12). Since its axis is inclined back-

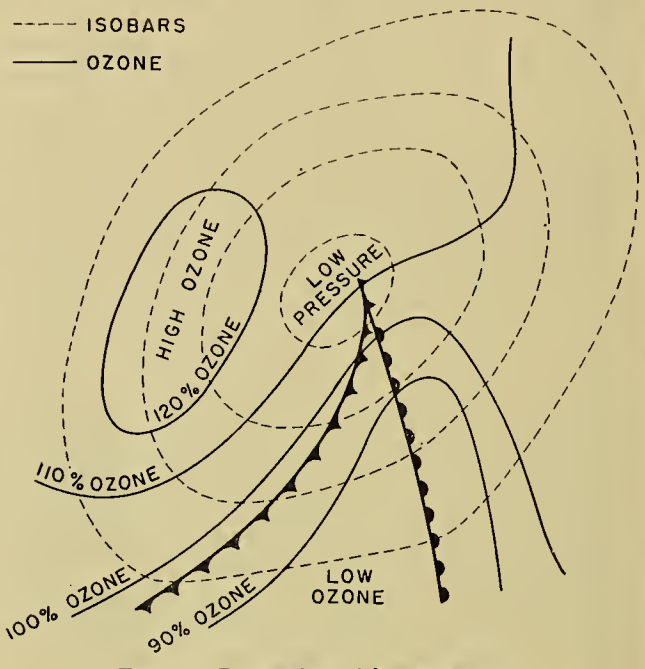


Fig. 12.—Depression with warm sector.

wards, Palmén [72] was able to point out that the region of maximum ozone coincides with the tropopause vortex, that is, the trough at the tropopause, and that the ozone amount is symmetrical to the pressure distribution at that height. Tropopause topography as well as ozone is explained as a common consequence of the three-dimensional air flow within a deepening cyclone. Thus, at the time at which one was still dependent upon statistical studies between ozone amount and the other atmospheric elements, Meetham [61] found the best correlations when referring all elements (variables) to a "pseudo-center" 300 to 350 km to the southwest of the center of the surface pressure system; this pseudo-center corresponds to the tropopause vortex. According to Meetham, a 0.01-cm increase in the amount of ozone is accompanied by a 3C rise of the potential temperature at an altitude of 18 km and by a lowering of the tropopause by 1 km. The correlation coefficient between the ozone amount and potential temperature at an altitude of 18 km has the high value of 0.8, a fact which however should not induce us to seek the seat of ozone fluctuations at this altitude. Johansen [52], who conveniently groups the ozone measurements according to rise and fall regions of atmospheric pressure at ground level, also finds high ozone amounts with a low tropopause over Tromsö.

According to the preceding section, these conditions must be explained by the fact that meteorological transport processes displace the ozone from its source. Two main possibilities exist; the decision between them requires modern upper-air weather maps on the one

hand, and continuous measurements of the vertical ozone distribution on the other hand.

As a first possibility, one considers the meteorological ozone distribution as a result of horizontal advection. Every veteran observer of ozone has long been familiar with the fact that invasions of arctic air extending to high altitudes are accompanied by an increase in ozone amount, whereas the ozone amount is slight in the case of European anticyclones composed of air of southern origin. The increased ozone amount accompanying invasions of cold air has been particularly noted by various expeditions [5, 97]. Lejay [58] points out that in the Siberian anticyclone in which the ozone amount is high, cold northern air masses advance so that the "transportation theory" removes the apparent contradiction with European findings. Penndorf [76] finds the 96-mb surface as suitable for indicating the flow conditions. Moser [65], in examining drawings of the trajectories at 5, 11, and 16 km, finds the best correlation between the trajectories at 11-km height and the ozone amount; he therefore assumes the main location of ozone fluctuations to be in the stratosphere near the tropopause where the maximum in wind velocity is located. This ozone maximum coincides with the lower secondary ozone layer which was first suggested, at least intermittently, by the investigations at Arosa [39]. This lower secondary layer is generally separated from the very constant high-altitude layer of ozone by a minimum in wind velocity between 15 and 22 km. The trajectories of Moser (Fig. 13) illustrate well the

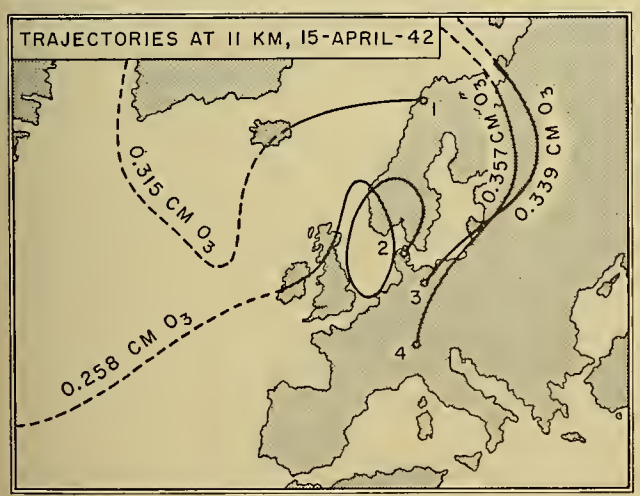


Fig. 13.—Air trajectories at 11 km.

temporal constancy of ozone amount at various points situated in the same air current, indicating that ozone is a conservative property of the air and may thus serve as an important indicator for air-mass determinations. On the other hand, in Fig. 14, Moser shows that at Arosa the ozone amount is higher the more northerly the latitudes from which the air current comes. The source of ozone, in this case, is the ozone belt located by us (Fig. 7). But how is this secondary source of low-altitude ozone supplied? Moser and Schröer suggest that the great temperature gradient at the shadow boundary between polar night and the sun-illuminated

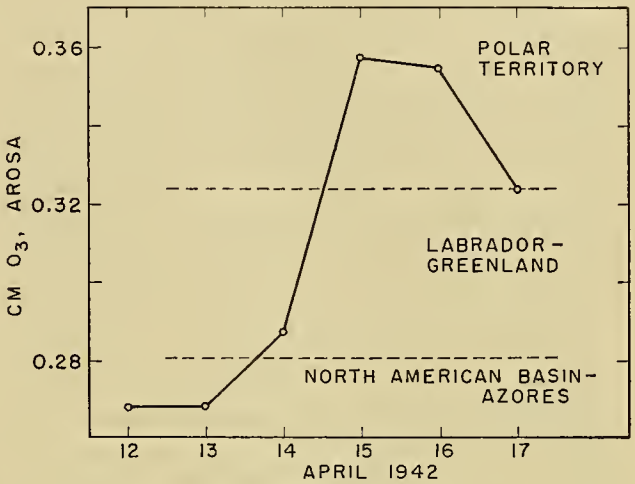


Fig. 14.—Ozone amount and origin of air at 11 km.

atmosphere leads to increased shear-turbulence between the layer of photochemical equilibrium and the lower ozone layer at an altitude of 23 km. They also suggest that a further connection with the lower stratosphere is provided by high-reaching cold-core lows within which the barrier by the minimum in wind velocity is lacking. This theory needs corroboration by additional measurements. Flohn [33] sees a possible source of the winter singularities in the 30.5-day oscillations of the winter circumpolar vortex of the warm layer [43]. Craig also suggests continuous subsidence of ozone in the polar cap during the cold months.

This leads us to the second possibility for ozone distribution, that of transport by vertical flow. A sinking column of air, such as exists in low-pressure regions at least at the height of the tropopause, is lengthened (stretched) in its upper ozone-containing portion and this is accompanied by a horizontal convergence or contraction. This is equivalent to an increase in ozone amount. An ascending current has the opposite effect. Thus, the explanation of the day-to-day changes does not even require the assumption that sinking air comes directly from the layer of photochemical equilibrium and is there replenished with ozone. Dobson [25] pointed out long ago "that the air immediately in the rear of a cyclone often has a higher ozone value than in the same general air stream further to the north or northwest." Haurwitz [50] emphasizes that horizontal advection cannot explain closed lines of equal ozone deviation such as those in Fig. 12. Moreover, the ozone content at the rear of a cyclone is sometimes higher than it is during the same season in the northern ozone source. Reed [81] emphasizes:

> Particularly noteworthy is the fact that closed isopleths of positive and negative ozone deviations bear the same relationship to surface cyclone centers as the closed isotherms which are invariably observed on the 200 mb chart (approximately

12 km), the warm pocket at this level coinciding with the positive deviations, the cold with the negative. Since these warm and cold areas are unquestionably due to subsidence and lifting respectively, it seems reasonable to suspect that the vertical displacements also in some fashion affect the ozone concentration.

Reed, moreover, recently calculated the change in ozone content during vertical displacements on the basis of the fact, already emphasized by Regener, that in strong vertical mixing the ozone mixing ratio (grams of ozone per gram of air) is independent of altitude. To begin with, the curve of ozone content versus altitude is transformed into a curve of ozone mixing ratio versus altitude. The curve is thereupon raised or lowered and finally reconverted back to an ozone content curve. In Fig. 15 it is assumed, at first, that the vertical displacement extends through the entire column of air [80]. More recently [81] the assumption was made that, in the closed region of high ozone amount to the west and southwest of an intense surface cyclone, the air subsides most markedly near the tropopause and that the sinking motion becomes unnoticeable above the tropopause between 16 and 18 km and below it at 7 to 8 km. Thus, the vertical displacements take place only within the secondary advection layer. Figure 16 is based on an Arosa curve for an ozone amount of 0.131 cm between altitudes of 5 and 20 km and an assumed maximum downward displacement of 1.4 km and shows an ozone increase of 0.025 cm. A corresponding example for lifting motion would yield a decrease of 0.017 cm. According to this new estimate, the contribution of the vertical motion in the interplay between advection and vertical motion is less significant than in previous assumptions, for instance that of Nicolet [67]. According to Reed, vertical motions account for, at most, about ⅓ of the total range. In summer and in fall, when the ozone content of the lower stratosphere is considerably less and vertical motion much weaker, the vertical motion effect will be almost negligible.

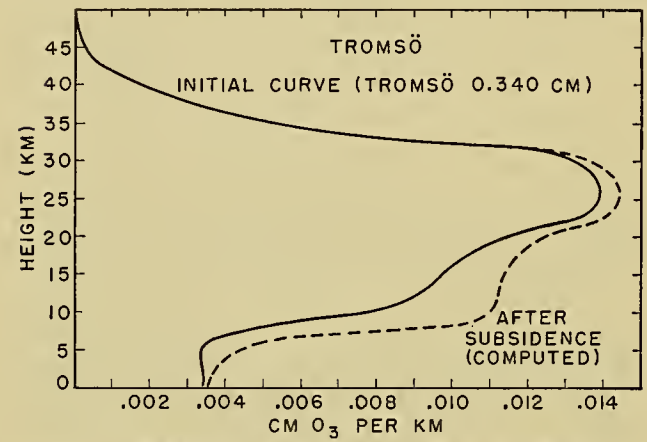


Fig. 15.—The effect of subsidence on the ozone distribution.

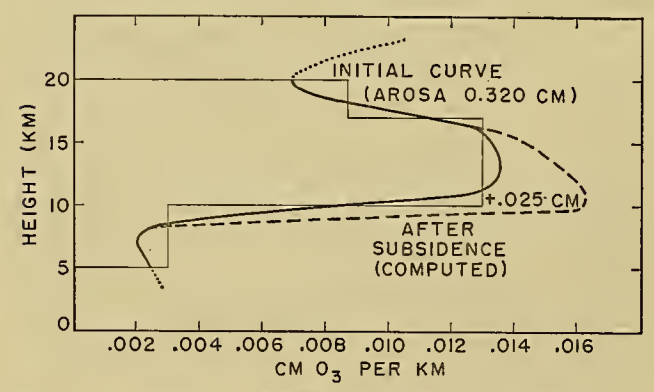


Fig. 16.—The effect of subsidence on the ozone distribution.

According to the upper-level weather maps, on which isobars generally have sinusoidal shape, advection and vertical motion always affect ozone changes in the same sense. In an influx of air from a ridge to a trough, the air comes predominantly from northern latitudes, and the consequent higher ozone amount is further increased by subsidence. In the example studied by Reed, the higher ozone content of 0.320 cm in an advectively transported air mass increases to 0.345 cm owing to subsidence; by contrast, the ozone content of 0.260 cm is decreased to 0.243 cm in an air mass that is lifted in the northeast quadrant of a surface cyclone.

Ozone Conditions in the Troposphere. Ozone conditions in the troposphere are evidently quite variable (Fig. 10), depending on the interplay of advection, which makes ozone suitable for air-mass analysis, with the turbulent ozone supply from the secondary tropopause layer ("flux d'ozone" according to Jaumotte, "Ozonstrom" according to V. H. Regener [86]). Whereas mixing produces an increase of ozone content at lower altitudes, it may happen in layers near ground level, in which ozone is consumed at a rapid rate by dust and oxidizable organic substances, that the ozone flow completely disappears in the case of stagnant air [83]. As a result, a weak tertiary layer of ozone develops at an altitude of several kilometers which is of climatic significance in high mountain regions. Starting from ozone near ground level, the significance of turbulence has been successfully demonstrated by E. Regener [82, 83] and his collaborators. The ozone stream treated theoretically by Lettau [59] has only on the average a downward direction in the troposphere because of advective superposition of air masses and shows relatively pronounced irregularities in space and time. Glueckauf [35] explains the region of diminished ground-level ozone ahead of a warm front on the basis of an interruption of the turbulent air exchange by the barrier layer of the front.

Ozone and Radiation Flux. A direct physical influence of ozone on meteorological processes through the medium of radiation flux has not been ascertained. As Lord Cherwell (Lindemann) suggested as early as 1919, ozone is as important for radiation balance in the lower stratosphere as are water vapor and carbon dioxide. However, any temperature rise due to increased ozone evidently takes place so slowly that it cannot be detected in meteorological circulation processes. However, in the case of a prolonged influence, the effect of ozone on incident and emitted radiation can

hardly be denied. Möller [64] explains the stratospheric temperature differences between tropics and temperate latitudes on the basis of the variable protective action of the ozone located above the emitting carbon dioxide; Dobson [23] arrives at an identical result even more directly. For the lower stratosphere he estimates the temperature of radiative equilibrium of H_2O to be 200C, that of CO_2 to be 200C, and that of O_3 to be 260C, which, for equal portions of these three absorbers, would lead to the plausible mean value of 220C. In comparison with higher latitudes, the lower secondary advection layer is absent in the vicinity of the equator, and ozone has correspondingly less effect on temperature. But even considering only the indirect action due to advection of air masses of different temperature, ozone has a significant effect on our weather.

ADDITIONAL REMARKS CONCERNING THE SIGNIFICANCE OF ATMOSPHERIC OZONE

Ozone and the Constitution of the Upper Atmosphere. *The Warm Layer.* Absorption of the short-wave radiation from the sun at the upper boundary of the ozone layer results in a temperature of +50 to +80C at an altitude of 55 km. This region is known as the warm layer [34]. This warm layer was first encountered in 1922 by Lindemann and Dobson in connection with the determination of the altitude at which meteors become incandescent. The anomalous propagation of sound, which results when sound encounters the warm layer and is refracted back to the earth [101], leads to good temperature data for this high layer; taking an average of the values given by Duckert [26], Gutenberg [48], and Whipple [26], the temperature in summer is −50C at an altitude of 30 km, +28C at 40 km, and +70C at 50 km. If a gas, such as ozone at its upper boundary, absorbs mainly in the ultraviolet part of the spectrum but emits according to the laws of temperature radiation, then its temperature must rise to a rather high value before incident and emitted radiation will attain equilibrium. This is even more applicable, incidentally, to the case of oxygen absorption in the ionosphere [38]. On the other hand, water vapor with its absorption in the infrared tends to lower the equilibrium temperature. Considering the fact that the sun's temperature is lower in the short-wave continuum, Gowan [46] finds that his calculations show satisfactory agreement between ozonosphere temperature and the data on record. According to his calculations [47], the warm layer persists even during the night. Barbier and Chalonge, E. Vassy and Déjardin (for references see [19]) attempted to calculate the mean temperature of the ozone layer on the basis of the temperature dependence of absorption in the Huggins bands by methods of indirect aerology; they also attempted to calculate the mean temperature of ozone above 30 km on the basis of the temperature distribution up to 30 km. In order to analyze the warm layer proper, Götz suggests that, instead of measuring the ozone bands in sunlight as customary, the Umkehr effect should be used in zenith light at low elevation of the sun. Indirect conclusions concerning the vertical temperature distribution have recently been verified [66] quite satisfactorily by means of V-2 rockets.[2]

Pekeris [74] has shown that on the basis of the warm layer it is possible to calculate a free oscillation of the atmosphere with a period of very nearly 12 hours, such as is necessary for the resonance theory of the migratory pressure wave of a half day's duration. As a heat source at high altitudes, the warm layer is capable of damping the circulation currents in the troposphere and is thus of climatic significance [8]. Many years ago, Wegener [100] deduced the existence of a high altitude troposphere (*Hochtroposphäre*) above 60 km from the existence of noctilucent clouds at an altitude of 82 km; we now know that the heat source for this phenomenon is the warm layer caused by ozone.

The Screening Effect of the Ozone Layer. The atmosphere of the earth forms an opaque screen for short-wave ultraviolet, so that the shadow cone for this radiation is not tangent to the earth's surface but rather to a sphere of enlarged radius. Götz [37] has therefore suggested that the upper boundary of the ozone layer be determined from ultraviolet observations during lunar eclipses. Barbier, Chalonge, and Vigroux [6] have accomplished this up to an altitude of 16 km [71*a*]. According to Götz this screening effect is of particular significance if the altitude of atmospheric dissociation, excitation, and ionization processes is to be calculated from the time of onset of these processes at dawn or the time at which they cease in the evening. When the altitude of sodium luminescence was given as 60 km on the basis of the dawn effect, it was pointed out [39] that the altitude of the ozone layer would have to be added to this figure if photoluminescence were assumed. And indeed, Vegard and Tønsberg [99], by simultaneous measurement of the intensity drop of the sodium line at the zenith and at the horizon, found the upper sodium boundary to be at 116 km and the altitude of the screening layer at 56 km. Penndorf [75], rigorously defining the ozone shadow boundary, has carefully calculated the conditions in an effort to determine accurately the upper boundary of the ozone layer from such observations. The ozone shadow boundary seems also to be significant in connection with certain delay processes observed in the E-layer by Lugeon and Mitra [39]. Störmer [90] found that noctilucent clouds always appear only when the sunlight striking them is no closer to the earth's surface than 30 to 45 km. It does not follow, of course, that it is the ozone shadow which is significant in all similar cases. In the case of the photoluminescent effect of the red twilight line at 6300 A, it is the shadow of the oxygen sphere which is opaque up to an altitude of about 100 km [41].

Ozone and Bioclimatology. The biological significance of the ozone layer can be touched upon but briefly within the scope of the present article. The ozone layer is of the greatest importance in controlling the ultra-

2. Consult Fig. 8 in "Temperatures and Pressures in the Upper Atmosphere" by H. E. Newell, Jr., pp. 303 to 310 in this Compendium.

violet radiation climate [10, 37, 57, 84] in a spectral range distinguished by a number of biological effects such as the formation of erythema, the formation of vitamin D, and bactericidal effects. Radiation intensities in the ultraviolet range of the solar spectrum are highly variable from wave length to wave length; however, the ozone measurements can entirely account for these variations in intensity. Our present-day knowledge concerning ozone amount is sufficient to allow us to estimate the incidence of ultraviolet over the entire earth. Without the ozone layer, sunburn would easily be fifty times as effective as it is during the highest sun's elevation during summer in high mountains. On the other hand, an increase in the ozone screen would completely eliminate the stimulating radiation, that, in proper dosage, is essential to life, and would leave us in biological darkness. Therefore it is probably significant if, as Fig. 8 would indicate, the ozone amount increases over a period of years while the ultraviolet radiation decreases. The zoologist Rowan [88] has recently raised the question as to whether such fluctuations may not provide an explanation for the ten-year cycle in the abundance of certain species among the Canadian fauna. Equally stimulating from the bioclimatological viewpoint are speculations as to whether the protective effect of the ozone layer has undergone changes during the development of the earth. For a mature oxygen planet such as Mars, the photochemical equilibrium layer must rest on the ground, a fact which would substantiate the explanation of the red discoloration of extended portions of the surface of Mars [102].

This oxidizing effect leads us, finally, to the ground-level ozone of the earth's biosphere. Götz and Ladenburg [44] have emphasized the desirability of investigating any possible direct influences of ozone on the human body. As E. Regener has pointed out, ozone bioclimatologically plays the important role of an air purifier [82]. It may even be said that it is ozone which characterizes "living air." The physician Curry [18] draws very extensive conclusions concerning the effect of ozone, and in general, of active forms of atmospheric oxygen on the human body. Even though these assertions have aroused much interest in this problem, they have not yet been substantiated by the use of rigorous statistical methods.

CONCLUSION

If we consider the results obtained thus far, the data concerning ozone amount appear the most reliable. Of course, the observation network is still not sufficiently dense, particularly in the region of large vortices. It has justly been pointed out that the present state of synoptic meteorology would leave much to be desired if it were dependent on a dozen barometers distributed over the entire earth! The continuous series of observations which were started at several locations must be carried on for prolonged periods of time because secular fluctuations reflect long-term changes in atmospheric circulation. The most interesting annual curve of ozone amount was observed in Tromsö, a fact which urgently suggests that the observation network be extended to the polar regions proper, such as Spitsbergen, for example.[3]

Since the ozone problem is fundamentally a flow problem, ozone amount must also be observed continuously in the third dimension. Information now available concerning the vertical distribution of ozone is far from adequate and, above all, much too disconnected. Vertical ozone distribution, together with information concerning water vapor, could be exploited directly in connection with forecasting and would provide the material necessary for approaching the problem of radiation flow. Perhaps it is a desire for more simple and less expensive apparatus which is discouraging meteorologists from attempting large-scale experiments? Stratosphere flights and V-2 rockets are rare opportunities. The Umkehr-effect method could be refined considerably if it were supplemented by ozone determinations in the lower ten kilometers by flight measurements —if one does not expect to measure the Umkehr effect continuously from meteorological aircraft at cloud-free altitudes. Perfection of the radiosonde should meet with no further fundamental difficulties.

It is fascinating to observe how ozone pulsates through the atmosphere like blood circulating in an organism. Ozone is created by radiation in high-altitude layers. Primarily at the shadow boundary of the polar night, and at the altitude of the warm layer, it produces the temperature contrasts and resulting polar vortices which enable it to sink as a secondary layer down to the theater of meteorological activity. And finally, diffusing to the vicinity of ground level, the ozone re-enters the oxygen metabolism. Much work remains to be done before this sketchy picture is completed and verified by observation.

This task requires ever-increasing international cooperation. It would be appropriate to pool all available tools (apparatus as well as observers, including other aerological methods) during international ozone weeks and to distribute them over suitable regions. International cooperation has indeed always been exemplary in the ozone field. Typical single cases could then be treated in a united effort which would help us in a decisive manner to push forward to the last meteorological consequences.

REFERENCES

1. Adel, A., and Lampland, C. O., "Atmospheric Absorption of Infrared Solar Radiation at the Lowell Observatory." *Astrophys. J.*, 91: 481–487 (1940).
2. Auer, R., "Ueber den täglichen Gang des Ozongehalts der bodennahen Luft." *Beitr. Geophys.*, 54: 137–145 (1939).
3. Barbier, D., et Chalonge, D., "Recherches sur l'ozone atmosphérique." *J. Phys. Radium*(7) 10: 113–123 (1939).

3. (*Added in press.*) In the ozone network of the International Ozone Commission arrangements have been made for daily measurements at Tromsö or Spitsbergen, Norfolk, Cornwall, Oxford, Uppsala, Oslo, Trappes, Arosa, and in the Shetland Islands, Northern Ireland, the Azores, Iceland, and outside Europe in Canada, the United States, India, and New Zealand.

4. —— "Sur les coefficients d'absorption de l'ozone dans la région des bandes de Huggins." *Ann. Phys., Paris*, (11) 17: 272–302 (1942).
5. —— et Vassy, E., "Mesure de l'épaisseur réduite de l'ozone atmosphérique pendant l'hiver polaire." *C. R. Acad. Sci., Paris*, 201: 787–789 (1935).
6. Barbier, D., Chalonge, D., et Vigroux, E., "Utilisation des éclipses de lune à l'étude de la haute atmosphère." *C. R. Acad. Sci., Paris*, 214: 983–984 (1942); "Étude spectrophotométrique de l'éclipse de lune des 2 et 3 mars 1942." *Ann. Astrophys.*, 5: 1 (1942).
7. Baur, F., "Die doppelte Schwankung der atmosphärischen Zirkulation in der gemässigten Zone innerhalb des Sonnenfleckenzyklus." *Meteor. Rdsch.*, 2: 10–15 (1949).
8. Bjerknes, V., "Sur les relations entre l'ozone et les mouvements de la troposphère." (See [31])
9. Bolz, H. M., "Ueber die Wirkung der Temperaturstrahlung des atmosphärischen Ozons am Erdboden." *Z. Meteor.*, 2: 225–228 (1948).
10. Büttner, K., *Physikalische Bioklimatologie. Probleme der kosm. Physik*, Bd. 18. Leipzig, Akad. Verlagsges., 1938.
11. Cabannes, J., et Dufay, J., "Les variations de la quantité d'ozone contenue dans l'atmosphère." *J. Phys. Radium*, (7) 8: 353–364 (1927).
12. Cauer, H., "Bestimmung des Gesamtoxydationswertes des Nitris, des Ozons, und des Gesamtchlorgehaltes roher und vergifteter Luft." *Z. anal. Chem.*, 103: 395–416 (1935).
13. Chalonge, D., et Vassy, E., "Spectrographe astigmatique à prisme objectif." *Rev. Opt. (théor. instrum.)* 13: 113–126 (1934).
14. Chapman, S., "The Götz Inversion of Intensity-Ratio in Zenith-Scattered Sunlight." *Phil. Trans. roy. Soc. London*, (A) 234: 205–230 (1934–35).
15. Coblentz, W. W., and Stair, R., "Distribution of Ozone in the Stratosphere." *J. Res. nat. Bur. Stand.*, 22: 573–606 (1939).
16. *Conferenee on Atmospheric Ozone. Quart. J. R. meteor. Soc.*, Supp. to Vol. 62, 76 pp. (1936).
17. Craig, R. A., *The Observations and Photochemistry of Atmospheric Ozone and Their Meteorological Significance.* Thesis, Mass. Inst. Tech., 1948. (See also under same title, *Meteor. Monogr.*, Vol. 1, No. 2 (1950).)
18. Curry, M., *Bioklimatik, Die Steuerung des gesunden und kranken Organismus durch die Atmosphäre*, 2 Bde. Riederau, Ammersee, American Bioclimatic Research Institute, 1946.
19. Déjardin, G., "Nouvelle détermination des coefficients d'absorption et de la température de l'ozone atmosphérique." *Cah. Phys.*, No. 16 (1943).
20. Dirnagl, K., "Messmethoden zum Studium biologischer Wirkungen des bodennahen Ozons." *Z. Hyg. InfektKr.*, 127: 676–683 (1948).
21. Dobson, G. M. B., "Observations of the Amount of Ozone in the Earth's Atmosphere, and Its Relation to Other Geophysical Conditions, Part IV." *Proc. roy. Soc.*, (A) 129: 411–433 (1930).
22. —— "A Photoelectric Spectrophotometer for Measuring the Amount of Atmospheric Ozone." *Proc. phys. Soc. Lond.*, 43: 324–339 (1931).
23. —— Brewer, A. W., and Cwilong, B. M., "Meteorology of the Lower Stratosphere." (Bakerian Lecture) *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
24. Dobson, G. M. B., and Harrison, D. N., "Measurements of the Amount of Ozone in the Earth's Atmosphere and Its Relation to other Geophysical Conditions." *Proc. roy. Soc.*, (A) 110: 660–693 (1926).
25. —— and Lawrence, J., "Measurements of the Amount of Ozone in the Earth's Atmosphere and Its Relation to Other Geophysical Conditions, Part III." *Proc. roy. Soc.*, (A) 122: 456–486 (1929).
26. Duckert, P., "Über die Ausbreitung von Explosionswellen in der Erdatmosphäre." *Beitr. Geophys.*, (Supp.) 1: 236–290 (1931).
27. Dütsch, H.-U., *Photochemische Theorie des atmosphärischen Ozons unter Berücksichtigung von Nichtgleichgewichtszuständen und Luftbewegungen.* Doctoral Dissertation, University of Zürich, 1946.
28. Ehmert, A., "Ueber das troposphärische Ozon." (See [71, pp. 26–28])
29. —— "Ein einfaches Verfahren zur Messung kleinster Jodkonzentrationen, Jod- und Natriumthiosulfatmengen in Lösungen." *Z. Naturforsch.*, 4b: 321–327 (1949).
30. Eucken, A., und Patat, F., "Die Temperaturabhängigkeit der photochemischen Ozonbildung." *Z. phys. Chem.*, (B) 33: 459–474 (1936).
31. Fabry, C., ed., "Rapport de la réunion de l'ozone et de l'absorption atmosphérique, Paris, 15–17 mai 1929." *Beitr. Geophys.*, 24: 1–7 (1929).
32. Fabry, C., et Buisson, H., "Etude de l'extremité ultraviolette du spectre solaire." *J. Phys. Radium*, (6) 2: 197–226 (1921).
33. Flohn, H., "Zur physikalischen Deutung der Witterungsregelfälle." *Ann. Meteor.*, 3: 57–59 (1950).
34. —— und Penndorf, R., "Die Stockwerke der Atmosphäre." *Meteor. Z.*, 59: 1–7 (1942).
35. Glueckauf, E., "The Ozone Content of the Surface Air and Its Relation to Some Meteorological Conditions." *Quart. J. R. meteor. Soc.*, 70: 13–21 (1944).
36. Götz, F. W. P., "Zum Strahlungsklima des Spitzbergensommers." *Beitr. Geophys.*, 31: 119–154 (1931).
37. —— "Das atmosphärische Ozon." *Beitr. Geophys.*, (Supp.) 1: 180–235 (1931).
38. —— "Absorption von Sonnenenergie in hohen Schichten." (See [16])
39. —— "Die vertikale Verteilung des atmosphärischen Ozons." *Beitr. Geophys.*, (Supp.) 3: 253–325 (1938).
40. —— "Der Stand des Ozonproblems." *Vjschr. naturf. Ges. Zürich*, 89: 250–264 (1944).
41. —— "Leuchtvorgänge der hohen Atmosphäre." Tagungsber., "Physik der hohen Atmosphäre." Laboratoire d'Études Ballistiques No. 33/46, St. Louis, 1946. (p. 121)
42. —— "Optics of the Turbid Atmosphere." *P. V. Météor. Un. géod. géophys. int.*, Oslo (1948).
43. —— *Périodicités dans les phénomènes de la haute atmosphère. Les relations entre les phénomènes solaires et géophysiques.* Colloques de Lyon 1947, Paris 1949.
44. —— und Ladenburg, R., "Ozongehalt der unteren Atmosphärenschichten." *Naturwissenschaften*, 19: 373–374 (1931).
45. Götz, F. W. P., Meetham, A. R., and Dobson, G. M. B., "The Vertical Distribution of Ozone in the Atmosphere." *Proc. roy. Soc.*, (A) 145: 416–446 (1934).
46. Gowan, E. H., "Ozonosphere Temperatures under Radiation Equilibrium." *Proc. roy. Soc.*, (A) 190: 219–226 (1947).
47. —— "Night Cooling of the Ozonosphere." *Proc. roy. Soc.*, (A) 190: 227–231 (1947).
48. Gutenberg, B., "Schallgeschwindigkeit und Temperatur in der Stratosphäre." *Beitr. Geophys.*, 27: 217–225 (1930).
49. Harteck, P., "Die Schwankungen des Ozongehaltes der Atmosphäre." *Naturwissenschaften*, 19: 858–860 (1931).
50. Haurwitz, B., "Atmospheric Ozone as a Constituent of the Atmosphere." *Bull. Amer. meteor. Soc.*, 19: 417–424 (1938).

51. Heilpern, W., und Meyer, E., "Die Absorption des Lichtes durch Sauerstoff und Sauerstoff-Stickstoffgemische im Wellenlängengebiet von $\lambda = 2100 - \lambda = 2400$ ÅE." (To appear in *Helv. phys. Acta*, (1950).)
52. Johansen, H., "Eine aerologische Untersuchung mittels Radiosondierungen in Tromsö während der Zeit 31. März—30. April 1939." *Meteor. Ann.*, 1: 45 (1942).
53. Karandikar, R. V., "Studies in Atmospheric Ozone" Part II: "Daily Measurements of Atmospheric Ozone and of $(\delta' - \delta'')$, a Measure of Atmospheric Turbidity, at Delhi." *Proc. Ind. Acad. Sci.*, 28(A):63–82 (1948).
54. —— and Ramanathan, K. R., "Vertical Distribution of Atmospheric Ozone in Low Latitudes." *Proc. Ind. Acad. Sci.*, 29: 330–348 (1949).
55. Kiepenheuer, K. O., "Ueber die Sonnenstrahluug zwischen 2000 und 3000 Å." *Veröff. UnivSternw. Göttingen*, Nr. 57 (1938).
56. Ladenburg, R., and van Voorhis, C. C., "The Continuous Absorption of Oxygen Between 1750 and 1300 Å and Its Bearing Upon the Dispersion." *Phys. Rev.*, 43: 315–321 (1933).
57. Latarjet, R., "Influence des variations de l'ozone atmosphérique sur l'activité biologique du rayonnement solaire." *Rev. Opt. (théor. instrum.)*, 14: 398–414 (1935).
58. Lejay, P., "Mesures de la quantité d'ozone contenue dans l'atmosphère à l'observatoire de Zô-sè, 1934–1935–1936; les variations de l'ozone et les situations météorologiques." *Notes Météor. phys., Zi-Ka-Wei*, Fasc. 7 (1937).
59. Lettau, H., "Zur Theorie der partiellen Gasentmischung in der Atmosphäre." *Meteor. Rdsch.*, 1: 65–74 (1947).
60. Mecke, R., "Zur Deutung des Ozongehalts der Atmosphäre." *Z. phys. Chem.*, Bodenstein-Festband, SS. 392–404 (1931).
61. Meetham, A. R., "The Correlation of the Amount of Ozone with Other Characteristics of the Atmosphere." *Quart. J. R. meteor. Soc.*, 63: 289–307 (1937).
62. —— and Dobson, G. M. B., "The Vertical Distribution of Atmospheric Ozone in High Latitudes." *Proc. roy. Soc.*, (A) 148: 598–603 (1935).
63. Meyer, E., "Ueber die Durchlässigkeit der Erdatmosphäre für Sonnenstrahlung der Wellenlänge $\lambda = 2144$ ÅE." *Helv. phys. Acta*, 14: 625–632 (1941).
64. Möller, F., "Zur Erklärung der Stratosphärentemperatur." *Naturwissenschaften*, 31: 148 (1943).
65. Moser, H., "Ozon und Wetterlage." (See [71])
66. Newell, H. E., Jr., *Upper Atmosphere Research with V-2 Rockets*. Naval Res. Lab. Rep. R-3294, Washington, D. C., 1948.
67. Nicolet, M., "L'ozone et ses relations avec la situation atmosphérique." *Inst. R. météor. Belg., Misc.*, Fasc. 19 (1945).
68. Ny Tsi-zé et Choong Shin-piaw, "Sur l'absorption ultraviolette de l'ozone." *Chin. J. Phys.*, Vol. 1, No. 1, pp. 38–50 (1933).
69. O'Brien, B., Mohler, F. L., and Stewart, H. S., Jr., "Vertical Distribution of Ozone in the Atmosphere." *Nat. Geogr. Soc. Contrib. Tech. Papers, Stratosphere Series*, 2: 49–93 (1936).
70. *Operator's Manual for the Dobson Spectrophotometer*. Beck, Ltd., London O. M. No. 2527, 1950.
71. "Ozon." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 11, Bad Kissingen (1949).
71a. Paetzold, H. K., "Eine Bestimmung der vertikalen Verteilung des atmosphärischen Ozons mit Hilfe von Moudfinsternissen." *Z. Naturforsch.*, 5(a):661–666 (1950).
72. Palmén, E., "Über die dreidimensionale Luftströmung in einer Zyklone und die Ozonverteilung." *Un. géod. géophys. int.*, Assoc. Météor. Réunion, p. 31, Washington, 1939. Bergen, 1939.
73. Paneth, F. A., and Glueckauf, E., "Measurement of Atmospheric Ozone by a Quick Electrochemical Method." *Nature*, 147: 614–615 (1941).
74. Pekeris, C. L., "Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 158: 650–671 (1937).
75. Penndorf, R., "Der Ozongehalt der mittleren Stratosphäre." *Z. Meteor.*, 1: 345–357 (1946–47). "Effects of the Ozone Shadow." *J. Meteor.*, 5: 152–160 (1948).
76. —— "Beiträge zum Ozonproblem II." (Manuscript)
77. —— "The Vertical Distribution of Atomic Oxygen in the Upper Atmosphere." *J. geophys. Res.*, 54: 7–38 (1949).
78. —— "Proposals for Unification of Principal Terms in Research on Atmospheric Ozone." *J. geophys. Res.*, 54: 169–172 (1949).
79. Ramanathan, K. R., and Karandikar, R V., "Effect of Dust and Haze on Measurements of Atmospheric Ozone Made with Dobson's Spectrophotometer." *Quart. J. R. meteor. Soc.*, 75: 527 (1949).
80. Reed, R. J., *The Effects of Atmospheric Motion on Ozone Distribution and Variations*. Thesis, Mass. Inst. Tech., 1949.
81. —— Personal communication.
82. Regener, E., "Ozonschicht und atmosphärische Turbulenz." *Meteor. Z.*, 60: 253–269 (1943).
83. —— "Das atmosphärische Ozon" in *FIAT Rev. German Sci., 1939–1946, Geophysics*, II, J. Bartels, senior ed., pp. 297–307. Off. Milit. Govt. Germany, Field Inform. Agencies, Tech. Wiesbaden, 1948.
84. —— "Über das 'photochemische' Klima der Erde." *Naturwissenschaften*, 33: 163–166 (1946).
85. —— und Regener, V. H., "Aufnahme des ultravioletten Sonnenspektrums in der Stratosphäre und die vertikale Ozonverteilung." *Phys. Z.*, 35: 788–793 (1934).
86. Regener, V. H., "Neue Messungen der vertikalen Verteilung des Ozons in der Atmosphäre." *Z. Phys.*, 109: 642–670 (1938).
87. —— "Messungen der Ozongehaltes der Luft in Bodennähe." *Meteor. Z.*, 55: 459–462 (1938).
88. Rowan, W., *Canada's Premier Problem of Animal Conservation*. Penguin Books, in press.
89. Schröer, E., "Theorie der Entstehung, Zersetzung und Verteilung des atmosphärischen Ozons." (See [71, pp. 13–23])
90. Störmer, C., "Measurements of Luminous Night Clouds in Norway 1933 and 1934." *Astrophys. norveg.*, 1: 87–114 (1935).
91. Stranz, J., *Ozonradiosonde*. Trans. Chalmers Univ. Tech. No. 72, Rep. Res. Lab. Electronics No. 6, 1948.
92. Strong, J., "On a New Method of Measuring the Mean Height of the Ozone in the Atmosphere." *J. Franklin Inst.*, 231: 121–155 (1941).
93. Tien Kiu, "Etude de l'absorption atmosphérique d'après les observations faites à Montezuma de 1920 à 1930, par la Smithsonian Institution." *Publ. Obs. Lyon*, Tome II, Sér. 1, Fasc. 8, pp. 241–251 (1938).
94. Tønsberg, E., and Olsen, K. L., "Investigations on Atmospheric Ozone at Nordlysobservatoriet, Tromsö." *Geofys. Publ.*, Vol. 13, No. 12 (1944).
95. *Upper Atmosphere Research IV*. Naval Res. Lab. Rep. R-3171, 170 pp., Washington, D. C., 1947.
96. Vassy, A., "Sur l'absorption atmosphérique dans l'ultraviolet." *Ann. Phys., Paris*, (11) 16: 145–203 (1941).

97. —— et Vassy, E., "Variations journalières de la température moyenne de l'ozone atmosphérique." *C. R. Acad. Sci., Paris*, 207: 1232–1234 (1938).
98. —— "Rôle de la température dans la distribution de l'ozone atmosphérique." *J. Phys. Radium*, (8) 2: 81–91 (1941).
99. Vegard, L., and Tønsberg, E., "Investigations on the Auroral and Twilight Luminescence Including Temperatures in the Ionosphere." *Geofys. Publ.*, Vol. 13, No. 1 (1940).
100. Wegener, A., "Die Temperatur der obersten Atmosphärenschichten." *Meteor. Z.*, 42: 402–405 (1925).
101. Whipple, F. J. W., "The High Temperature of the Upper Atmosphere as an Explanation of Zones of Audibility." *Nature*, 111: 187 (1923).
102. Wildt, R., "Ozon und Sauerstoff in den Planetenatmosphären." *Veröff. UnivSternw. Göttingen*, Nr. 38 (1934).
103. Wulf, O. R., and Deming, L. S., "The Theoretical Calculation of the Distribution of Photochemically-Formed Ozone in the Atmosphere." *Terr. Magn. atmos. Elect.*, 41: 299–310 (1936).
104. —— "The Distribution of Atmospheric Ozone in Equilibrium with Solar Radiation and the Rate of Maintenance of the Distribution." *Terr. Magn. atmos. Elect.*, 42: 195–202 (1937).

RADIATIVE TEMPERATURE CHANGES IN THE OZONE LAYER

By RICHARD A. CRAIG

Harvard College Observatory

THE OZONE LAYER

The term "ozone layer" as used here refers to that part of the atmosphere that lies above the tropopause and below about 60 km. The bulk of the atmospheric ozone lies in this region. The physical characteristics of the ozone layer have been studied, principally from the surface of the earth, by various indirect methods. Some of the conclusions reached by these methods, however, have been verified by a few direct measurements obtained by means of manned and unmanned sounding balloons and rockets.

The present state of knowledge with regard to the ozone layer is outlined elsewhere in this volume by Götz.[1] Götz has also presented a very complete summary of ozone work as of the year 1938 [20] and in a more recent paper has brought this summary up to the year 1944 [21]. Even more recently Craig [10] has given a somewhat less comprehensive summary. Nevertheless, in order that this contribution may be more or less self-contained, certain basic facts about the ozone layer, particularly those that are necessary to an understanding of the remainder of this paper, are outlined briefly in this section.

Ozone Distribution. The total amount of ozone in a vertical column above the earth's surface is small compared to that of other atmospheric constituents, namely, about 0.3 cm at NTP.[2] Nevertheless, because of its absorbing qualities, ozone is an extremely important constituent of the atmosphere.

A large number of measurements at the earth's surface have determined the variations with season and geographical position of this total amount of ozone.[3] At a given place on the earth, the ozone amount is a maximum in the early spring and a minimum in the late fall. The total ozone amount is a minimum at the equator and increases toward higher latitudes. At least in the Northern Hemisphere, it apparently reaches a maximum near 60°N and decreases poleward from there. A further interesting feature of ozone distribution in middle and high latitudes is that the total amount of ozone above a given place may vary markedly from day to day. The amplitude of this variation of daily values in any given month may be as large as the amplitude of the annual variation of the monthly means. Moreover, the day-to-day ozone variations reveal marked correspondence to weather conditions at the surface, as shown by Dobson [12], Lejay [37], and Tønsberg and Olsen [55].

The vertical distribution of ozone has been most generally studied by means of the Umkehr effect. This effect, first noted by Götz at Arosa and applied by Götz, Meetham, and Dobson [22], refers to the observations of scattered zenith light when the sun is near the horizon. The ratio of the intensities of two wave lengths, both absorbed by ozone but to different degrees, decreases as the sun nears the horizon, reaches a minimum, and then increases. This phenomenon results from the fact that the effective height of scattering increases as the sun nears the horizon and finally lies above the ozone layer. The shape of the Umkehr curve may be used to deduce the vertical distribution of ozone, as has been shown by Götz, Meetham, and Dobson [22] for Arosa observations. Meetham and Dobson at Tromsö [40], Tønsberg and Olsen at Tromsö [55], and Karandikar and Ramanathan at Delhi and Poona [33] have also applied the method. The Umkehr method is only approximate, but its results have been verified by various direct measurements [8, 9, 43, 47, 51]. The measurements show that the maximum density of ozone occurs between 20 and 30 km and that it decreases rapidly above the level of maximum. The density of ozone below the maximum level is quite variable. In fact, the observations show that most of the variability in the total amount of ozone (latitudinal, seasonal, and day-to-day) reflects variability between the tropopause and the level of maximum ozone.

Temperature Distribution. The temperature distribution in the ozone layer, at least above 30 km, is less well known. In the vertical, the temperature between the tropopause and about 30 km is nearly isothermal. This region has been studied directly by radiosonde. Above 30 km, mainly indirect evidence points to a rapid increase of temperature with height, reaching a maximum at 50–60 km. This evidence is based on observations of the anomalous propagation of sound [58] and of the characteristics of meteor trails [38, 59, 60]. Recently, measurements from a V-2 rocket have verified this qualitative picture [31].

In the lower, isothermal region, where direct measurements are available, seasonal and latitudinal variations of temperature are known. In the summer, the temperature is lowest over the equator and increases toward the poles. In the winter, the poleward increase of temperature also occurs in lower latitudes, but a maximum of temperature occurs in middle latitudes and the temperature is again lower at higher latitudes [29]. At a given point in middle latitudes the maximum temperature in the lower part of the ozone layer occurs just before the summer solstice, the minimum just before the winter solstice. This contrasts with the behavior of the upper troposphere, where the maximum and

1. See "Ozone in the Atmosphere" pp. 275–291 of this Compendium.

2. Ozone amounts are generally expressed in terms of the height of the resulting volume of ozone if all the ozone in the column were brought to normal temperature and pressure at the earth's surface.

3. For references to published series of ozone observations, see [10].

minimum follow the solstices. Dobson, Brewer, and Cwilong [11] and Goody [19] explain this behavior as a balance between two conflicting factors. They point out that the temperature of the lower stratosphere is primarily controlled by radiative processes, and is heated at least in part by ozone absorption of infrared radiation from the troposphere. The flux of this radiation reaches a maximum (or minimum) after the solstices. On the other hand, the amount of ozone reaches a maximum (or minimum) much earlier, just after the equinoxes. Therefore the temperature extremes occur at intermediate times.

The temperature variations in the upper, warm, region of the ozone layer are not known. Direct measurements on a routine basis are not as yet obtainable there.[4] Indirect evidence, however, indicates that the seasonal variations, at least, may be very large. Whipple, Jacchia, and Kopal [60] have shown from their meteor studies that the density of the atmosphere at 70 km is approximately 2–3 times greater in summer than in winter in middle latitudes. This astonishing fact can be explained only in terms of a summertime expansion of the ozone layer above 30 km, corresponding to a mean temperature some 50C higher than in winter. Furthermore, Gowan's calculations of radiative equilibrium temperature at 50°N [26], discussed more fully below, indicate a summer-winter temperature difference of the same magnitude.

Ozone Absorption. The explanation for the warming of the upper part of the ozone layer lies in the absorbing qualities of ozone. Laboratory measurements of ozone absorption reveal intense absorption bands in the ultraviolet and less intense bands in the visible. The Hartley bands of ozone, the most intense of all, lie between 2000 and 3200 A, with a strong maximum near 2500 A. The Huggins bands occur in the 3200–3600 A region. In the visible are the Chappuis bands at 4800–7800 A. The most detailed and homogeneous set of laboratory-derived absorption coefficients stem from the work of Ny Tsi-zé and Choong Shin-piaw [45, 46] and of Vassy [56]. Many other investigators [16, 34, 36, 41] have obtained results in agreement with theirs.

Photochemistry of Ozone. The existence of ozone in the upper atmosphere may be explained on the basis of photochemical principles. The photochemistry of atmospheric oxygen has been discussed by many investigators, for example, Chapman [7], Bamford [1], and Wulf [61].

The primary reaction leading to the formation of ozone is the dissociation of the oxygen molecule by solar energy at wave lengths less than 2400 A:[5]

$$O_2 + h\nu\ (\lambda < 2400\ A) \rightarrow O + O, \qquad (1)$$

where ν is the frequency of the incident light and h is Planck's constant. Ozone is then formed by the collision of an oxygen atom, an oxygen molecule, and any third body:

$$O_2 + O + M \rightarrow O_3 + M. \qquad (2)$$

Ozone in turn can be dissociated by absorption of solar energy or by collision with an oxygen atom:

$$O_3 + h\nu\ (\lambda < 11{,}000\ A) \rightarrow O_2 + O, \qquad (3)$$

$$O_3 + O \rightarrow 2O_2. \qquad (4)$$

From these four reactions, the rates of change of the amounts of ozone and atomic oxygen per unit volume are

$$dn_1/dt = 2Q_2 + Q_3 - k_{12}n_1n_2n_m - k_{13}n_1n_3, \qquad (5)$$

$$dn_3/dt = \quad - Q_3 + k_{12}n_1n_2n_m - k_{13}n_1n_3. \qquad (6)$$

The symbols n_1, n_2, and n_3 represent the numbers of molecules per unit volume of atomic oxygen, molecular oxygen, and ozone, respectively; n_m represents the total number of molecules per unit volume in the air. The numbers of quanta absorbed from the solar beam per unit volume and unit time are given by Q_2 (for O_2) and Q_3 (for O_3). The rate of production or destruction of molecules from collisions is proportional to the numbers of colliding molecules in the unit volume, the photochemical reaction factors k_{12} and k_{13} being the constants of proportionality for the collisions represented by (2) and (4), respectively.

Under equilibrium conditions, $dn_1/dt = dn_3/dt = 0$. From (5) and (6), then,

$$n_3 = \left(\frac{k_{12}}{k_{13}}\right) n_2 n_m \left(\frac{Q_2}{Q_2 + Q_3}\right). \qquad (7)$$

This equation was derived and numerically integrated by Wulf and Deming [61, 62, 63]. More recently Dütsch [13], Nicolet [44], and Craig [10] have repeated the calculations, making use of more recent and detailed information about the parameters entering into the calculation. All these computations give the equilibrium amounts of ozone to be expected at various levels in the atmosphere. Despite some uncertainties in the computations, the results are generally compatible with observation.

Among the most interesting information derived from computations of equilibrium ozone amounts is that concerning the degree to which actual ozone amounts may be expected to correspond to equilibrium amounts. At and above the level of maximum ozone, any deviations from equilibrium conditions could last only a few hours. However, below the level of maximum ozone density this time interval rapidly increases until, in the lower part of the layer, it is extremely large. Thus, below perhaps 25 km, the ozone is never necessarily in equilibrium with the sun. It is in just this region that the large variations in ozone amount are observed to occur.

Possible Solar Effects on the Ozone Layer. One interesting aspect of the question of radiative changes in the ozone layer is the suggestion, made most recently

4. Recent work at the Evans Signal Laboratory [3] gives hope that observations from improved radiosondes may soon become available up to 40–50 km.

5. Not all absorption below this wave length produces direct dissociation, but it all at least excites the molecule to the point where it is easily dissociated.

by Haurwitz [28], that solar variability may affect the ozone layer directly and our weather indirectly. Solar variability in the visible is at best very small, while in the ultraviolet it appears, from various phenomena in the upper atmosphere, to be large. The ozone layer is a strong absorber of solar ultraviolet radiation and is situated at a level in the atmosphere where there is still appreciable mass. This interesting question, then, lends added importance to the whole problem of the size and distribution of radiative changes in the ozone layer.

HEATING OF THE OZONE LAYER

Several different radiative processes serve to heat the ozone layer. Most of these processes involve the absorption, by some constituent of the atmosphere, of energy from the direct solar beam. The lower part of the ozone layer, however, is also heated by absorption of infrared radiation from the earth's surface and from the troposphere and by ozone absorption of solar energy reflected and scattered from the troposphere.

The next section, on cooling of the ozone layer, deals with the absorption characteristics of atmospheric gases in the infrared. The general information pertinent to the question of heating due to infrared absorption is available there. In this section, only general considerations relative to the question of absorption in the ultraviolet and visible are included. In a later section some numerical results from computations will be presented.

Ozone Absorption. Ozone is the most important constituent of the ozone layer from the point of view of absorption of solar energy. Particularly, the Hartley bands (2000–3200 A) absorb strongly and are responsible for the sharp cut-off of the solar spectrum near 3000 A.

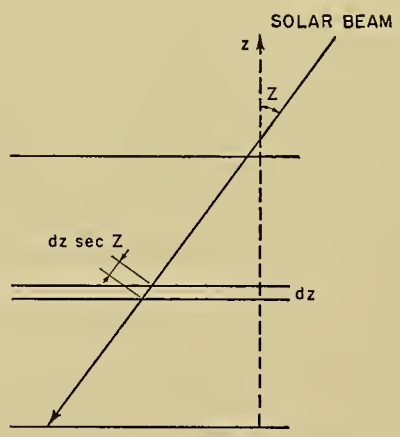


FIG. 1.—Solar beam passing through the ozone layer when the sun is at the zenith angle Z.

In Fig. 1, let a parallel beam of radiation from the sun be incident at the top of the ozone layer at an angle Z from the vertical. In the spectral interval $d\lambda$ let the intensity of the incident solar radiation be $I_{0\lambda}$, and call the ozone absorption coefficient α_λ. Consider a column of air that is parallel to the solar beam and that has unit cross section. While passing through a vertical distance dz, the solar intensity in $d\lambda$ is decreased by the amount

$$dI_\lambda = I_\lambda \alpha_\lambda n \sec Z \, dz, \tag{8}$$

where n is the amount of ozone per unit volume and I_λ is the intensity at the level z. The sign in (8) is positive because z is taken positive upward. Equation (8) can be integrated to give

$$I_\lambda = I_{0\lambda} \exp\left(-\int_z^\infty \alpha_\lambda n \sec Z \, dz\right). \tag{9}$$

The energy absorbed per unit volume in the spectral interval $d\lambda$ is

$$\frac{dI_\lambda}{\sec Z \, dz} d\lambda = I_{0\lambda} \alpha_\lambda n \exp(-\alpha_\lambda N) \, d\lambda, \tag{10}$$

where

$$N \equiv \int_z^\infty n \sec Z \, dz \tag{11}$$

is the total amount of ozone the solar energy has traversed in its oblique path above the level z. The total energy absorbed per unit volume at the level z is

$$E_z = \int_0^\infty \frac{dI_\lambda}{\sec Z \, dz} d\lambda = n \int_0^\infty I_{0\lambda} \alpha_\lambda \exp(-\alpha_\lambda N) \, d\lambda. \tag{12}$$

In practice, the integration in (12) needs to be taken only over the spectral interval of appreciable ozone absorption; namely, the Hartley bands (2000–3200 A), the Huggins bands (3200–3600 A), and the Chappuis bands (4500–6500 A). The spectrum can be divided into several finite intervals characterized by appropriate mean values of $I_{0\lambda}$ and α_λ. Then (12) is a function only of N. Later in this section (Fig. 2) E_z/n is shown graphically as a function of N for specific spectral distributions of $I_{0\lambda}$ and α_λ.

The rate of heating of the unit volume is then

$$\frac{\partial T}{\partial t} = \frac{E_z}{c_p \rho}, \tag{13}$$

where c_p is the specific heat of air at constant pressure, which has the value 0.239 cal g^{-1} deg^{-1}, and ρ is the air density, which of course varies with elevation.

In the Huggins and Chappuis bands, a large part of the extraterrestrial solar radiation penetrates the ozone layer and is later scattered and reflected back from the troposphere to be absorbed by the ozone. The treatment of this phase of the problem is quite similar to the discussion above, except that the radiation is diffuse rather than parallel. The equation corresponding to (12) is

$$E_z' = \frac{2\pi n \int I_{t\lambda} \alpha_\lambda Ei_2(\alpha_\lambda N) \, d\lambda}{3 \sec Z}, \tag{14}$$

where $I_{t\lambda}$ is the intensity reaching the troposphere in the spectral interval $d\lambda$. The exponential transmission

function in (12) gives way to the function $2\pi Ei_2(\alpha_\lambda N)$, where $Ei_2(\alpha_\lambda N)$ is an exponential integral, because the returning radiation is diffuse. The path length N is, of course, now measured vertically upward from the tropopause. The factor 3 in the denominator indicates that only about one-third of the incident radiation is reflected upward from the troposphere, while the factor sec Z takes care of the fact that the radiation is incident at the angle Z from the vertical.

Oxygen Absorption. The absorption spectrum of oxygen in the ultraviolet includes the weak Herzberg bands which converge near 2400 A; the much stronger Schumann band system which begins near 2000 A, converges at about 1750 A, and reaches its maximum intensity at about 1450 A; and, finally, the Hopfield bands between 1000 and 600 A. Absorption by oxygen in the infrared and visible regions of the spectrum is very weak.

The absorption coefficients of air in the Hopfield bands have been estimated by Schneider [53]. These results are only approximate, but leave no doubt that the solar energy in this spectral interval is already absorbed far above the ozone layer. Between 1000 and 1300 A the atmosphere has some strong bands and some transparent regions. Particularly, near the wave length of the Lyman-α emission line of hydrogen (1216 A), the atmosphere seems to be relatively transparent. Here, however, Preston [49] has made careful measurements of absorption coefficients of air and its constituents and finds that solar radiation could hardly penetrate below 60 km in important amounts. Similarly Ladenburg and Van Voorhis [35] have measured oxygen absorption in the Schumann region (1300–1750 A) and their results show that the solar energy here also is depleted above the ozone layer.

The only solar radiation, then, that can heat the ozone layer lies above 1800 A. Above about 2200 A, on the other hand, oxygen is comparatively unimportant as a heating agent because ozone absorbs the bulk of the available energy. For the intermediate region, 1800–2200 A, Granath [27] has measured the oxygen absorption in the 1900–2100 A region. The absorption of air has been determined by Buisson, Jausseran, and Rouard [6] in the interval 1855–2653 A. Because the measurements were made in surface air relatively free from ozone, one can determine the absorption of oxygen from these measurements with some degree of precision. The absorption coefficients so derived agree well with Granath's measurements in the region where they overlap.

The oxygen absorption spectrum is complicated by a pressure dependence of the absorption in the Herzberg region. Heilpern [30] discussed oxygen absorption at 2144 A for pressures varying between 148 and 663 mm Hg at a temperature of 18C and found a rather strong pressure dependence.

In any case, heating caused by oxygen absorption is generally negligible compared to that caused by ozone absorption in the ozone layer. Only in the upper part of the layer (above 40 km), where it may approach 10 per cent of the ozone absorption, does oxygen absorption need to be considered by the careful investigator. At the present stage of this type of study, the figure of 10 per cent is considerably less than the uncertainties introduced by other doubtful factors.

Solar Energy in the Ultraviolet. Until recent V-2 rocket flights were consummated, the spectral distribution of solar energy to the violet of 3000 A was unknown. This factor is, of course, necessary for the calculation of heating in the ozone layer. Nearly all calculations of this heating have been based on the assumption that the sun radiates as a black body at a temperature of 6000K. The rocket measurement has shown that the radiation is actually considerably less intense.

The first source of information about the solar spectrum in the region under consideration (1800–3500 A) is that of measurements made at the earth's surface and corrected for absorption in the atmosphere. Such information can, of course, extend down to only about 3000 A, but even this shows that the radiated energy is considerably less than would be expected from the black-body assumption mentioned above. Several series of such measurements have been summarized by Moon [42]. Below 4000 A, the measurements indicate a steady decrease of emission relative to black-body emission at 6000K until at 3000 A the ratio is only about 40 per cent.

The rocket flight of October 10, 1946 obtained a spectrum at 55 km (already above all but about one per cent of the ozone) extending down to 2200 A [31]. The ratio of observed intensity to black-body intensity for a temperature of 6000K was observed to decrease irregularly from about 70 per cent at 3300–3400 A to about 6 per cent at 2200 A. The agreement with surface observations in the overlapping region (3000–3400 A) is satisfactory. An approximate spectrum based on these two types of information has been given by Craig [10, Fig. 10]. The graph of E_z/n against N in Fig. 2 is based on this distribution.

Ozone Distribution. The vertical distribution of ozone is, of course, an important parameter in the determination of rate of heating of the ozone layer at various levels. Unfortunately, Umkehr measurements give little more than the order of magnitude of the amount of ozone present above 30 km. Neither do these measurements give any reliable information about the seasonal and latitudinal variations of ozone amount above 30 km.

In this connection, the calculations of equilibrium ozone amounts may be useful. There are many uncertainties in the calculations that make them only approximate. However, particularly above 30 km, they agree as to order of magnitude with the Umkehr results. The calculations have the advantage over the observations that they are capable of showing, at least approximately, the variations of ozone amount with zenith angle. The photochemical theory also indicates a substantial variation of equilibrium ozone amount with temperature, because the photochemical factor k_{12}/k_{13} in (7) varies markedly with temperature [15]. Thus calculations of the photochemical-equilibrium amounts of ozone above 35 km, applied to the problem of computing the rates of radiative heating of the ozone layer, give

the same order of accuracy as the Umkehr observations and can show, at least qualitatively, the variations of rate of heating with latitude and season.

Calculation of Heating of the Ozone Layer. To calculate the heating of the ozone layer resulting from ozone absorption of direct solar radiation, one needs to know (1) the absorption spectrum of ozone, (2) the spectral distribution of solar energy, 2000–3500 A and 4500–6500 A, and (3) the vertical ozone distribution.

With regard to the absorption spectrum of ozone and the solar spectrum, Craig [10] has given estimates based on the most recent and reliable data. Figure 2 gives

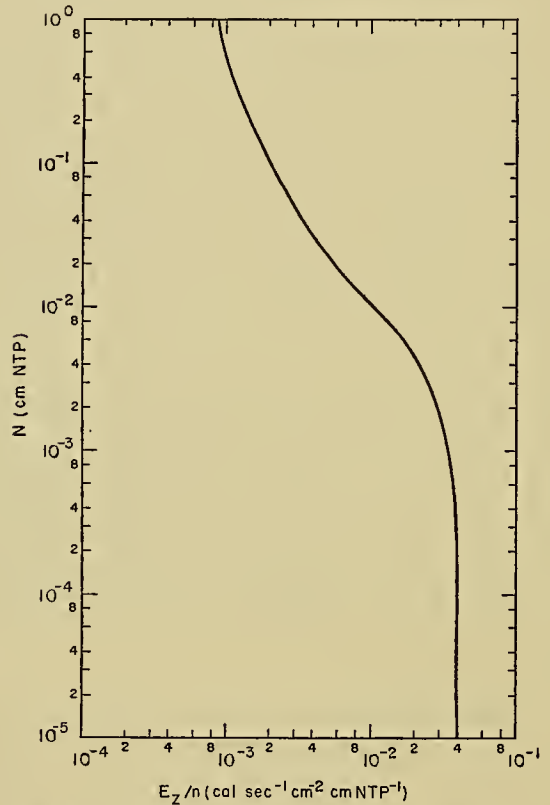


FIG. 2.—Variation of E_z/n as a function of N.

E_z/n as a function of N, from (12), for these estimates. For very small values of N the exponent in (12) is small, so the exponential term is close to unity; hence E_z/n is nearly independent of N. The curve then shows a strong variation of E_z/n with N in the range of path lengths that includes most of the Hartley absorption. When the solar energy in the Hartley region is exhausted, the integration in (12) effectively extends only over the Huggins and Chappuis bands. Here the absorption coefficients are small and, for values of N encountered in the atmosphere, the exponential term in (12) again approaches unity. From this graph the reader can easily find E_z, and hence the heating from (13) at any desired level in the ozone layer and for any assumed solar zenith angle and vertical ozone distribution.

COOLING OF THE OZONE LAYER

Cooling of the ozone layer results from radiative transfer in the infrared. The atmospheric gases responsible for this cooling are water vapor, carbon dioxide, and ozone. Several factors make calculations of the rate of cooling inherently more complex than calculations of the rate of heating. In the first place, the infrared radiation that affects a given level originates at all other levels of the atmosphere and is *diffuse* radiation. In the second place, the absorption bands in the infrared consist of sharp lines with little continuous background absorption, so that the absorption coefficient varies rapidly with wave length. In the third place, the absolute amounts of the gases in the ozone layer are very small, smaller than those used heretofore in most laboratory experiments. Finally, the range of variation of pressure in the region under consideration is two to three orders of magnitude, and pressure effects on the infrared absorption are marked and not completely understood.

Infrared Spectra of Water Vapor, Carbon Dioxide, and Ozone. Water vapor contains two principal absorption bands in the infrared. The most intense, the *rotational* band, is located at the long-wave end of the spectrum, beyond about 20 μ, and has been studied spectroscopically by Randall, Dennison, Ginsburg, and Weber [50]. The band at 6 μ has not been studied as exhaustively, but Fowle [17, 18] has made absorption measurements.

Carbon dioxide has three bands in the infrared, intense ones at about 4 μ and 15 μ and a weak band near 10 μ. The band at 4 μ, while intense, is located in a region of comparatively small radiation for black bodies at atmospheric temperatures. However, the 15-μ band is exceedingly important in the radiative processes of the atmosphere, lying as it does near the peak of the black-body radiation at atmospheric temperatures. Martin and Barker [39] have studied this band spectroscopically.

Ozone has two bands, one near 10 μ and a second that nearly overlaps the 15-μ carbon dioxide band. Strong [54] has studied the absorption of the former band.

Methods of Calculating Cooling. To compute the flux of infrared radiation arriving at a given level in the atmosphere, one needs to consider the radiation originating at all other levels and also the absorption of radiation during its passage from its origin to the reference level. In Fig. 3, let the unit area P, through which the downward flux is to be computed, lie on the horizontal plane $u = 0$, where the symbol u represents the mass of the radiating substance in a vertical column of unit area. Consider an infinitesimal volume element in the plane $u = u$, with vertical thickness du. The line from this element to P makes an angle θ with the vertical and an angle φ with an arbitrary reference direction in the horizontal. The emission of monochromatic radiation from this element is

$$dI\, d\lambda = kI_b \sec\theta\, du \sin\theta\, d\theta\, d\varphi\, d\lambda, \qquad (15)$$

where k is the absorption coefficient and I_b is the black-body radiation of the element at the wave length λ in question. The vertical flux reaching P from this element is $\cos\theta\, dI\, d\lambda \exp(-ku \sec\theta)$. The total monochromatic flux reaching P from the plane $u = u$ is obtained by

integrating this expression over the plane,

$$f(u,\lambda) = du\,d\lambda \int_0^{2\pi} d\varphi \int_0^{\pi/2} kI_b e^{-ku\sec\theta} \sin\theta\, d\theta \qquad (16)$$
$$= 2\pi k I_b Ei_2(ku)\, du\, d\lambda,$$

where $Ei_2(ku)$ is an exponential integral.

The total vertical flux reaching P from all points above is thus

$$F = \int_0^{\infty} \int_0^{u} 2\pi k I_b Ei_2(ku)\, du\, d\lambda. \qquad (17)$$

The integration of (17) solves the computational problem. In practice, however, this integration is very difficult. In the first place, the black-body radiation I_b depends on the temperature, which in turn depends on the path length u in an irregular manner that varies from time to time in the atmosphere. Secondly, the absorption coefficient k depends on the wave length λ in a rapid and irregular manner. Indeed, this variation in the infrared bands is so rapid as to prevent the direct application of methods of numerical integration.

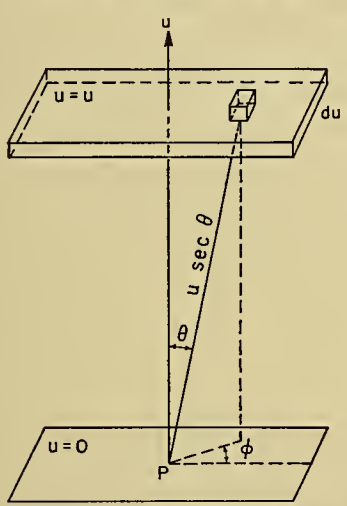


FIG. 3.—Radiation to P from a volume element above P.

Elsasser [14] and Schnaidt [52] have introduced a simplification that makes the integration over wave length practicable. They divide the spectrum into small intervals, each interval containing a number of absorption lines. Under the assumptions that these lines are broadened by pressure effects and that the lines in any one spectral interval are all equal in intensity and equidistant, they derive expressions for the average absorption in the spectral interval. The absorption coefficient, which may vary greatly in the interval, can then be replaced by a "generalized" absorption coefficient which is constant in the interval.

The *transmission function* $\tau_I(u)$ is defined as the ratio I/I_0 of the radiation penetrating a layer of thickness u to the radiation incident at the top of the layer. For diffuse radiation, the transmission function $\tau_f(u)$ is similarly defined in terms of fluxes rather than intensities. The relation between τ_f and τ_I is

$$\tau_f(u) = \int_0^1 \tau_I(u \sec\theta)\, d \sin^2\theta. \qquad (18)$$

Thus for exponential absorption of a beam of monochromatic radiation, $\tau_I = e^{-ku}$ and $\tau_f = 2Ei_3(ku)$. In these more general terms, (17) may be written in the form

$$F = -\int_0^{\infty} d\lambda \int_0^{u} f_b \frac{d}{du} \tau_f[lu]\, du, \qquad (19)$$

where f_b is the black-body flux, given by πI_b and l is the generalized absorption coefficient.

The transmission functions given by Elsasser [14] and Schnaidt [52] apply strictly only to spectral intervals that contain equal and equidistant lines, each line broadened by pressure effects only. Experiment, however, has shown that the formulas apply to a good approximation to the actual behavior of the infrared absorption bands of the atmospheric gases. With the aid of these developments one can integrate (19) over wave length. The integration over path length must be accomplished numerically or graphically in a separate calculation for each atmosphere that has a distinctive relationship between path length and temperature. The cooling at any level in the atmosphere is then proportional to the vertical divergence of the net flux at that level.

A more direct method of computing cooling in the atmosphere has been given by Bruinenberg [5] and Brooks [4]. It gives the divergence of the flux, and hence the cooling, directly. This method is to be preferred for the ozone layer where the fluxes are small in any case.

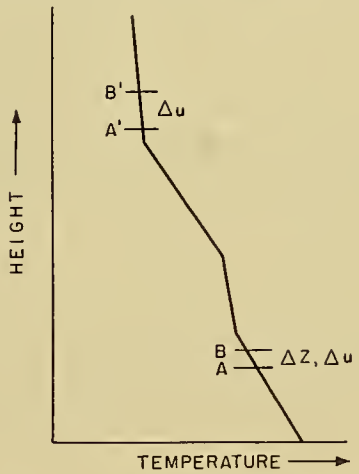


FIG. 4.—Schematic representation of temperature-height sounding of the atmosphere. The path length of absorbent, u, between A and A' is the same as that between B and B'.

Figure 4 gives a schematic representation of the variation of temperature with height. Let the total amount of absorbent above A be u_A, and that above B be u_B, where $u_A - u_B = \Delta u$. Define the flux from an *isothermal* column as

$$\mathcal{F}(u, T) = \epsilon_f(u, T)\, \sigma T^4. \qquad (20)$$

The total emissivity of the isothermal column, ϵ_f, is defined as the ratio of the flux emitted from the column to the black-body flux at the temperature of the column. Thus,

$$\epsilon_f = \frac{1}{\sigma T^4} \int_0^{\infty} f_b(1 - \tau_f)\, d\lambda. \qquad (21)$$

The downward fluxes at A and B can be written as

$$F_A = \int_0^{u_A} \frac{\partial \mathcal{F}_A}{\partial u}\, du, \qquad F_B = \int_0^{u_B} \frac{\partial \mathcal{F}_B}{\partial u}\, du. \tag{22}$$

Equations (20) and (22) are consistent with (19). Therefore,

$$F_A - F_B = \int_0^{u_B} \left(\frac{\partial \mathcal{F}_A}{\partial u} - \frac{\partial \mathcal{F}_B}{\partial u}\right) du + \int_{u_B}^{u_A} \frac{\partial \mathcal{F}_A}{\partial u}\, du. \tag{23}$$

Consider a level A' at a path length u above A and a level B' at the same path length u above B. (The linear distances $A - A'$ and $B - B'$ are not necessarily the same.) The flux from B' through B and the flux from A' through A would be identical except that A' and B' may be at different temperatures.
One can write

$$\left(\frac{\partial \mathcal{F}_A}{\partial u} - \frac{\partial \mathcal{F}_B}{\partial u}\right)_{\text{from } u} = \frac{\partial^2 \mathcal{F}}{\partial u\, \partial T}\,(dT)_u = \frac{\partial^2 \mathcal{F}}{\partial u\, \partial T}\frac{dT}{du}\,\Delta u. \tag{24}$$

The second integral in (23) represents the additional flux at A because $u_A > u_B$ so that

$$\int_{u_A}^{u_B} \left(\frac{\partial \mathcal{F}_A}{\partial u}\right) du = \left(\frac{\partial \mathcal{F}}{\partial u}\right)_t \Delta u, \tag{25}$$

where $(\partial \mathcal{F}/\partial u)_t$ is the value of $(\partial \mathcal{F}/\partial u)$ at the top of the atmosphere. With these values inserted (23) becomes

$$\Delta F = \int_0^{u_B} \left(\frac{\partial^2 \mathcal{F}}{\partial u\, \partial T}\right)_u \left(\frac{dT}{du}\right)_u \Delta u\, du + \left(\frac{\partial \mathcal{F}}{\partial u}\right)_t \Delta u. \tag{26}$$

The cooling is proportional to the divergence of the flux. Divide both sides of (26) by Δz and let both sides approach the limit $\Delta z = 0$. Then

$$\frac{\partial F}{\partial z} = \frac{\partial u}{\partial z}\left[\int_0^{u_B} \left(\frac{\partial^2 \mathcal{F}}{\partial u\, \partial T}\right) dT + \left(\frac{\partial \mathcal{F}}{\partial u}\right)_t\right]. \tag{27}$$

The divergence of the upward flux can be derived in the same manner. However, because the ground acts as a black body, the expression for the upward flux has no term comparable to the second term in (27). Thus, finally,

$$\begin{aligned}\frac{\partial T}{\partial t} &= +\frac{1}{\rho c_p}\left(\frac{\partial F^{\downarrow}}{\partial z} - \frac{\partial F^{\uparrow}}{\partial z}\right) \\ &= +\frac{1}{\rho c_p}\frac{\partial u}{\partial z}\left[\int_0^{u_B} \left(\frac{\partial^2 \mathcal{F}^{\downarrow}}{\partial u\, \partial T}\right)_{u,T} dT \right. \\ &\qquad \left. + \left(\frac{\partial \mathcal{F}^{\downarrow}}{\partial u}\right)_t - \int_0^{\infty} \left(\frac{\partial^2 \mathcal{F}^{\uparrow}}{\partial u\, \partial T}\right)_{u,T} dT\right].\end{aligned} \tag{28}$$

Data for Calculating Cooling. According to (28), the cooling at any level in the ozone layer can be determined by numerical or graphical integration if $\mathcal{F}$ is known as a function of u and T. From (20)

$$\frac{\partial^2 \mathcal{F}}{\partial u\, \partial T} = 4\sigma T^3 \frac{\partial \epsilon_f}{\partial u} + \sigma T^4 \frac{\partial^2 \epsilon_f}{\partial u\, \partial T}. \tag{29}$$

The emissivity ϵ_f can be determined either from laboratory measurements or from the theoretical transmission functions according to (21).

Elsasser [14] has summarized measured values of the emissivities of water vapor and carbon dioxide. Summerfield[6] has given data on the 10-μ band of ozone. Unfortunately, in none of these cases do the measurements extend to values of u as small as those met in the ozone layer. Some radiation computations have made use of extrapolations of these emissivity curves.

An alternative procedure is to make use of the transmission functions for the infrared bands. The functions can first be tested against and fitted to the observation in the range of u where measurements are available. In general, it is possible to get a good agreement between theory and observations. The transmission functions can then be used to extrapolate the available information to small values of u. Even this type of extrapolation, however, is risky. The transmission functions, as stated above, were derived on the basis of idealized bands. That they describe the behavior of the actual infrared bands in a certain range is no guarantee that they will serve equally well in another range. A further difficulty stems from the overlapping of the ozone and carbon dioxide bands at 15 μ. Very little is known about the former band. How much its presence may affect the carbon dioxide emissivities as measured in the laboratory is not known.

Pressure Effects on Infrared Absorption. Still another difficult and unsolved problem is the question of pressure effects on absorption in the infrared. It is certain that the half-width of the absorption lines in the infrared varies with pressure. Elsasser [14] gives a rather complete discussion of the experimental facts about this dependence.

For water vapor, the half-widths of the absorption lines seem to vary as the square root of the pressure. Strong [54] has shown a similar square-root dependence for the 10-μ ozone band. In the case of carbon dioxide evidence is conflicting, and various investigators have assumed that the half-width varies as the pressure according to laws ranging from square-root dependence to direct dependence. This is a question, particularly in the case of carbon dioxide, that should be cleared up before accurate calculations can be made.

RESULTS OF CALCULATIONS

Despite the difficulties discussed above, some investigators have made calculations based on the available knowledge. These calculations correspond qualitatively to the observational information available about the temperature distribution in the ozone layer. It is the purpose of this section to outline the results so far achieved.

Gowan's Calculations of Equilibrium Temperature. E. H. Gowan has pioneered the work in this field. In a pair of early papers [23, 24], he made some preliminary calculations which showed that the suspected

6. In an unpublished doctoral dissertation at the California Institute of Technology in 1941.

temperature maximum near 55 km was reasonable. In 1936 [25], he revised these calculations on the basis of later determinations of the vertical distribution of ozone. Finally, in 1947 [26] he published revised results based on further information about absorption in the infrared. Only the last results are included here.

Gowan gives his results in the form of radiative-equilibrium temperatures at various levels in the ozone layer. These are the temperatures that would exist according to his calculations if the radiative gains of the layer were just balanced by the radiative losses, with no effects of atmospheric circulations. The assumptions made by Gowan are:

1. *Solar Energy.* For all calculations except one, Gowan assumes that the sun radiates as a black body at 6000K in the ultraviolet. In this one case, the emission is taken as that of a black body at 4000K.

2. *Vertical Distribution of Absorbing Gases. Ozone* is assumed to vary vertically according to the Umkehr measurements of Götz, Meetham, and Dobson at Arosa for a total amount of ozone of either 0.20 cm NTP or 0.28 cm NTP. *Water vapor* is assumed to be either 10 per cent or 40 per cent saturated at the tropopause with constant mixing ratio in the ozone layer. *Carbon dioxide* is assumed to be present throughout the ozone layer in a concentration of 0.03 per cent by volume (corresponding to tropospheric conditions).

Table I. Radiative-Equilibrium Temperatures at 50°N (*After Gowan* [26])*

	Summer								Winter	
Amount O_3 (cm NTP)	.280		.280		.280		.200		.280	.280
Amount H_2O (%)	0		10		40		10		10	10
Solar temperature (°K)	6000		6000		6000		6000		4000	6000
Height (km)	Temperature (°K)									
50–55	452	415	*448*	344	*441*	410	*445*	323	*347*	406 *439*
45–50	429	410	*424*	361	*421*	409	*422*	321	*333*	364 *375*
40–45	399	385	*397*	350	*394*	382	*390*	311	*319*	305 *314*
35–40	335	324	*332*	295	*327*	320	*330*	291	*301*	278 *285*
30–35	296	285	*295*	262	*291*	281	*292*	272	*283*	262 *273*
25–30	275	258	*272*	244	*265*	257	*269*	252	*266*	246 *256*
20–25	254	241	*249*	240	*239*	239	*247*	236	*245*	230 *238*
15–20	239	232	*232*	232	*221*	229	*225*	229	*229*	221 *221*
11–15	228	218	*217*	211	*209*	215	*208*	217	*215*	208 *209*

* The two columns under each group of assumptions represent the alternative results if water vapor absorption is (right column) or is not (left column) corrected for a pressure dependence.

3. *Absorption Spectra* In the ultraviolet, absorption coefficients of oxygen are taken from the measurements of Granath [27], those of ozone from the measurements of various investigators [16, 36]. In the infrared, the emissivity of carbon dioxide is taken from laboratory measurements, as summarized by Elsasser [14], and extrapolated on log-log paper to smaller values of path length. Similarly the absorption coefficients of Fowle and Hettner [17] for water vapor are extrapolated on log-log paper. Summerfield's measurements at the 10-μ hand of ozone are utilized. The water vapor coefficients are corrected according to the square-root pressure law in some cases. The calculations apply to a latitude of 50°N in summer or winter according to various combinations of the assumptions listed above. Table I gives the results.

These results show reasonable agreement with observations. The temperature increases with height to the top of the ozone layer, although the maximum is not usually reached in these calculations. The temperatures are considerably higher than those that actually occur in the ozone layer; this is due in part to the assumption of solar black-body radiation at 6000K. Recent rocket measurements show much less energy.

Penndorf's Calculations of Rate of Heating and Cooling because of Ozone Alone. Penndorf [48] has computed the rates of heating and cooling of the ozone layer that would result from the effects of ozone alone. He takes the solar emission curve in the ultraviolet as that of a black body at 5910K and assumes the vertical distribution of ozone to be that measured by the Umkehr effect at Arosa. He integrates the temperature changes over the period of a day.

Penndorf's results show maximum heating resulting from ozone absorption of solar ultraviolet radiation to lie between 45 and 50 km. The cooling effect of long-wave radiation is a maximum at 50 km. The absolute value of the rate of heating is 10–50 times greater than that of the rate of cooling, at least from 30 to 50 km. This discrepancy is probably at least partially due to two factors: (1) smaller solar emission in the ultraviolet than the assumed amount; and (2) the cooling effects of carbon dioxide and water vapor are not included.

Karandikar's Calculation of Heating of the Ozone Layer. Karandikar [32] has given a rather complete discussion of the rate of heating of the ozone layer. He considers not only heating caused by ozone absorption of ultraviolet solar radiation, but also the heating caused by absorption of solar radiation in the infrared bands of ozone, carbon dioxide, and water vapor. He assumes the sun to radiate as a black body at 6000K. The vertical distribution of ozone is based on available Umkehr measurements, that of carbon dioxide on a constant proportion by volume (0.03 per cent) throughout the ozone layer. Several alternative assumptions are made about the total amount of water vapor present, the vertical distribution of the water vapor being taken to follow Dalton's law.

The results show a maximum of absorbed energy between 40 and 50 km. For a solar zenith angle of 0°, the maximum absorption occurs at 40–45 km. For a solar zenith angle of 75°, the maximum absorption occurs at 45–50 km and is only about half as intense as for the smaller zenith angle. Karandikar's computations show that the heating produced by infrared absorption of solar energy can be safely neglected above 30 km in comparison with the heating produced by ultraviolet absorption of ozone. Below 30 km, on the other hand, the former process becomes predominant, water vapor playing the most important role.

Craig's Calculations of Heating and Cooling. The writer has carried through some unpublished calculations of heating and cooling of the ozone layer. They are mentioned here to show the results that have been

obtained with somewhat different assumptions than the ones on which the published calculations are based. The assumptions that differ markedly from those outlined above are:

1. *Solar Energy.* The solar emission curve in the ultraviolet corresponds to the results of the rocket measurements. As mentioned previously, these show considerably less solar energy than a black body at 6000K would radiate.

2. *Vertical Distribution of Absorbing Gases.* The vertical distribution of ozone above 35 km is computed from photochemical-equilibrium theory [10]. This gives the same magnitude as Umkehr measurements and shows to a first approximation variations with solar zenith angle and temperature.

3. *Absorption Spectra.* The emissivities of carbon dioxide and ozone for small path lengths are obtained from the theoretical transmission functions rather than from straight extrapolation of existing measurements. Cooling is evaluated directly by graphical integration of (28).

For the temperatures generally assumed to occur in the ozone layer [57], the calculated rates of heating and cooling are of the same order of magnitude at all levels. This contrasts with Gowan's results; since his equilibrium temperatures are considerably higher than those that actually occur, his assumptions would give greater heating than cooling at the lower temperatures. It also contrasts with Penndorf's results. In both cases, the difference in assumed solar energy is the principal reason for the discrepancy. The level of maximum heating is at 40–45 km, lower than that found by Karandikar. This difference is probably also a result of the different assumed-energy curves. The ratio of the measured solar intensity to the black-body intensity is relatively smaller near the maximum of the Hartley bands of ozone, which heat the upper part of the ozone layer, than at longer wave lengths.

The infrared cooling due to the carbon dioxide band at 15 μ is more intense than that due to ozone or water vapor, at least above 35 km. This result must be considered somewhat doubtful until further light is shed on the problems of pressure effects on this band, and of the overlapping of the 15-μ ozone band. The absolute values of the rates of heating and cooling are relatively low, of the order of magnitude of 0.1C per three hours below 30 km and of 1C per three hours at 35–50 km. These are no larger than might be expected from normal atmospheric circulation processes.

SUBJECTS FOR FURTHER RESEARCH

It has become evident throughout this discussion that several fruitful avenues of research lie open to the interested investigator. In this concluding section, these are summarized and briefly discussed.

Solar Energy in the Ultraviolet. Even though rocket measurements of the solar spectrum down to 2200 A have thus far been of great help, many more are needed. For the present problem, only a slight further extension into the ultraviolet is necessary, perhaps to 1800 A. However, many more measurements should be made to check the accuracy of the information now available and to show whether there are any significant variations of the spectrum with time.

Vertical Distribution of Absorbing Gases. The vertical distributions of all of the absorbing gases in the ozone layer are in doubt. In the case of ozone, the vertical distributions above 30 km at various latitudes and times of year are urgently needed. For carbon dioxide, some direct measurements should attempt to test the usual assumption that the concentration in the ozone layer is the same as in the troposphere. Particularly above 30 km this is desirable. Water vapor is probably not important to the present problem at levels above 30 km (unless it is present in much greater concentration than now assumed). However, in the lower isothermal part of the layer its concentration should be measured carefully under a variety of conditions. Recent developments in England [11] and the United States [2] give encouragement in this direction.

Absorption Spectra. The absorption spectra of oxygen and ozone in the pertinent part of the ultraviolet seem to be satisfactorily known. For the problem of calculation of photochemical-equilibrium amounts of ozone, however, further study of the pressure dependence of oxygen absorption should be made in the laboratory. The spectral region where information is vitally needed is 1800–2200 A and the pressures range from 10 to 0.1 mb.

Absorption data in the infrared are urgently needed. Perhaps the most practical type of laboratory measurements for the present problem would give the isothermal emissivities of carbon dioxide and ozone as a function of path length, pressure, and temperature, particularly the first two. The range of path lengths of carbon dioxide that needs further study is from 0.01 to 1 cm NTP at pressures ranging from 0.1 to 10 mb. Ozone path lengths from 10^{-4} to 10^{-2} cm NTP at pressures of 1 to 10 mb need further study. Furthermore, emissivity measurements should be made on various mixtures of ozone and carbon dioxide within these limits to determine the effect of the former on the latter at 15 μ.

Calculations with Existing Data. Further calculations with existing data are possible, and indeed desirable, unless some of the above experimental information is forthcoming in the immediate future. The calculations made up to now show that useful results can be obtained. Particularly, calculations for various latitudes and seasons may begin to show the extent of temperature variations in the ozone layer, a matter about which we now have no information.

REFERENCES

1. Bamford, C. H., "Photochemical Processes in an Oxygen-Nitrogen Atmosphere" in *Reports on Progress in Physics*, 9: 75–91. Phys. Soc., London, 1943.
2. Barrett, E. W., Herndon, L. R., Jr., and Carter, H. J., "A Preliminary Note on the Measurement of Water-Vapor Content in the Middle Stratosphere." *J. Meteor.*, 6: 367–368 (1949).
3. Brasefield, C. J., "Exploring the Ozonosphere." *Sci. Mon.*, 68: 395–399 (1949).

4. Brooks, D. L., *Measurements of Atmospheric Radiation Applied to the Heat Transfer by Infrared Radiation in the Free Atmosphere.* Sc.D. Thesis, 74 pp., Mass. Inst. Tech., 1948.
5. Bruinenberg, A., "Een numerieke Methode voor de Bepaling van Temperatuurs-veranderingen door Straling in de vrije Atmosfeer." *Meded. ned. meteor. Inst.*, (B) Deel 1, Nr. 1 (1946).
6. Buisson, H., Jausseran, C., et Rouard, P., "La transparence de la basse atmosphère." *Rev. Opt. (théor. instrum.)*, 12: 70–80 (1933).
7. Chapman, S., "The Photochemistry of Atmospheric Oxygen" in *Reports on Progress in Physics*, 9: 92–100. Phys. Soc., London, 1943.
8. Coblentz, W. W., and Stair, R., "Distribution of Ozone in the Stratosphere." *J. Res. nat. Bur. Stand.*, 22: 573–606 (1939).
9. —— "Distribution of Ozone in the Stratosphere: Measurements of 1939 and 1940." *J. Res. nat. Bur. Stand.*, 26: 161–174 (1941).
10. Craig, R. A., "The Observations and Photochemistry of Atmospheric Ozone and Their Meteorological Significance." *Meteor. Monogr.*, Vol. 1, No. 2 (1950).
11. Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., "Meteorology of the Lower Stratosphere." (Bakerian Lecture) *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
12. Dobson, G. M. B., Harrison, D. N., and Lawrence, J., "Measurements of the Amount of Ozone in the Earth's Atmosphere and Its Relation to other Geophysical Conditions—Part III." *Proc. roy. Soc.*, (A) 122: 456–486 (1928).
13. Dütsch, H.-U., *Photochemische Theorie aes atmosphärischen Ozons unter Berücksichtigung von Nichtgleichgewichtszuständen und Luftbewegungen.* Doctoral Dissertation, University of Zürich, 1946.
14. Elsasser, W. M., "Heat Transfer by Infrared Radiation in the Atmosphere." *Harv. meteor. Studies*, No. 6, 107 pp. (1942).
15. Eucken, A., und Patat, F., "Die Temperaturabhängigkeit der photochemischen Ozonbildung." *Z. phys. Chem.*, (B) 33: 459–474 (1936).
16. Fabry, C., et Buisson, M., "L'absorption de l'ultraviolet par l'ozone et la limite du spectre solaire." *J. Phys. Radium*, 3: 196–206 (1913).
17. Fowle, F. E., "Water Vapor Transparency to Low Temperature Radiation." *Smithson. misc. Coll.*, Vol. 68, No. 8 (1917).
18. —— "The Transparency of Aqueous Vapor." *Astrophys. J.*, 42: 394–411 (1915).
19. Goody, R. M., "The Thermal Equilibrium at the Tropopause and the Temperature of the Lower Stratosphere." *Proc. roy. Soc.*, (A) 197: 487–505 (1949).
20. Götz, F. W. P., "Die vertikale Verteilung des atmosphärischen Ozons." (*Ergebn. kosm. Phys.*) *Beitr. Geophys.*, (Supp.) 3: 253–325 (1938).
21. —— "Der Stand des Ozonproblems." *Vjschr. naturf. Ges. Zürich*, 89: 250–264 (1944).
22. —— Meetham, A. R., and Dobson, G. M. B., "The Vertical Distribution of Ozone in the Atmosphere." *Proc. roy. Soc.*, (A) 145: 416–446 (1934).
23. Gowan, E. H., "The Effect of Ozone on the Temperature of the Upper Atmosphere." *Proc. roy. Soc.*, (A) 120: 655–669 (1928).
24. —— "The Effect of Ozone on the Temperature of the Upper Atmosphere—II." *Proc. roy. Soc.*, (A) 128: 531–550 (1930).
25. —— "The Effect of Ozone on the Temperature of the Upper Atmosphere." (Contrib. to Conf. on Atmos. Ozone, No. 13) *Quart. J. R. meteor. Soc.*, (Supp.) 62: 34–37 (1936).
26. —— "Ozonosphere Temperatures under Radiation Equilibrium." *Proc. roy. Soc.*, (A) 190: 219–226 (1947).
27. Granath, L. P., "The Absorption of Ultra-violet Light by Oxygen, Water Vapor, and Quartz." *Phys. Rev.*, 34: 1045–1048 (1929).
28. Haurwitz, B., "Relations between Solar Activity and the Lower Atmosphere." *Trans. Amer. geophys. Un.*, 27: 161–163 (1946).
29. —— and Austin, J. M., *Climatology.* New York, McGraw, 1944.
30. Heilpern, W., "Die Absorption des Lichtes durch Sauerstoff bei der Wellenlänge $\lambda = 2144$ ÅE. in Abhängigkeit vom Druck." *Helv. phys. Acta*, 14: 329–354 (1941).
31. Hulburt, E. O., "The Upper Atmosphere of the Earth." *J. opt. Soc. Amer.*, 37: 405–415 (1947).
32. Karandikar, R. V., "Radiation Balance of the Lower Stratosphere. Part I—Height Distribution of Solar Energy Absorption in the Atmosphere." *Proc. Ind. Acad. Sci.*, (A) 23: 70–96 (1946).
33. —— and Ramanathan, K. R., "Vertical Distribution of Atmospheric Ozone in Low Latitudes." *Proc. Ind. Acad. Sci.*, (A) 29: 330–348 (1949).
34. Krüger, F., und Moeller, M., "Über die Absorption der Ultravioletten Strahlung im Ozon und ihre Verwendung zur Bestimmung geringer Ozonkonzentration." *Phys. Z.*, 13: 729–732 (1912).
35. Ladenburg, R., and Van Voorhis, C. C., "The Continuous Absorption of Oxygen between 1750 and 1300 A and Its Bearing upon the Dispersion." *Phys. Rev.*, 43: 315–321 (1933).
36. Läuchli, A., "Zur Absorption der ultravioletten Strahlung im Ozon." *Z. Phys.* 53: 92–94 (1929).
37. Lejay, P., "Mesures de la quantité d'ozone contenue dans l'atmosphère à l'observatoire de Zô-sè, 1934–1935–1936; les variations de l'ozone et les situations météorologiques." *Notes Météor. phys., Zi-Ka-Wei*, No. 7 (1937).
38. Lindemann, F. A., and Dobson, G. M. B., "A Theory of Meteors and the Density and Temperature of the Outer Atmosphere to Which It Leads." *Proc. roy. Soc.*, (A) 102: 411–437 (1923).
39. Martin, P. E., and Barker, E. F., "The Infrared Absorption Spectrum of Carbon Dioxide." *Phys. Rev.*, 41: 291–303 (1932).
40. Meetham, A. R., and Dobson, G. M. B., "The Vertical Distribution of Atmospheric Ozone in High Latitudes." *Proc. roy. Soc.*, (A) 148: 598–603 (1935).
41. Meyer, E., "Über die Absorption der ultravioletten Strahlung in Ozon." *Ann. Physik*, 12: 849–859 (1903).
42. Moon, P., "Proposed Standard Solar-Radiation Curves for Engineering Use." *J. Franklin Inst.*, 230: 583–617 (1940).
43. Newell, H. E., Jr., and Siry, J. W., *Upper Atmosphere Rescarch Report No. III.* Naval Res. Lab. Rep. R-3120, 1947.
44. Nicolet, M., "L'ozone et ses relations avec la situation atmosphérique." *Inst. R. météor. Belg., Misc.*, Fasc. 19 (1945).
45. Ny Tsi-zé et Choong Shin-piaw, "L'absorption de la lumière par l'ozone entre 3050 et 3400 Å (Région des bandes de Huggins)." *C. R. Acad. Sci., Paris*, 195: 309–311 (1932).
46. —— "L'absorption de la lumière par l'ozone entre 3050 et 2150 Å." *C. R. Acad. Sci., Paris*, 196: 916–918 (1933).
47. O'Brien, B., Stewart, H. S., and Mohler, F. L., "Vertical Distribution of Ozone in the Atmosphere." *Nat.*

Geogr. Soc. Contrib. Tech. Papers, Stratosphere Series, 2: 49–93 (1936).

48. Penndorf, R., "Beiträge zum Ozonproblem; Die Rolle des Ozons im Wärmehaushalt der Stratosphäre." *Veröff. geophys. Inst. Univ. Lpz.*, 2. Ser., 8: 181–285 (1936).
49. Preston, W. M., "The Origin of Radio Fade-outs and the Absorption Coefficient of Gases for Light of Wave-Length 1215.7 A." *Phys. Rev.*, 57: 887–894 (1940).
50. Randall, H. M., and others, "The Far Infrared Spectrum of Water Vapor." *Phys. Rev.*, 52: 160–174 (1937).
51. Regener, E., und Regener, V. H., "Aufnahmen des ultravioletten Sonnenspektrums in der Stratosphäre und die vertikale Ozonverteilung." *Phys. Z.*, 35: 788–793 (1934).
52. Schnaidt, F., "Über die Absorption von Wasserdampf und Kohlensäure mit besonderer Berücksichtigung der Druck- und Temperaturabhängigkeit." *Beitr. Geophys.*, 54: 203–234 (1939).
53. Schneider, E. G., "An Estimate of the Absorption of Air in the Extreme Ultraviolet." *J. opt. Soc. Amer.*, 30: 128–132 (1940).
54. Strong, J., "On a New Method of Measuring the Mean Height of the Ozone in the Atmosphere." *J. Franklin Inst.*, 231: 121–155 (1941).
55. Tønsberg, E., and Olsen, K. L., "Investigations on Atmospheric Ozone at Nordlysobservatoriet, Tromsö." *Geofys. Publ.*, Vol. 13, No. 12 (1944).
56. Vassy, A., "Sur l'absorption atmosphérique dans l'ultraviolet." *Ann. Phys., Paris*, ser. 11, 16: 145–203 (1941).
57. Warfield, C. N., "Tentative Tables for the Properties of the Upper Atmosphere." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1200 (1947).
58. Whipple, F. J. W., "The Propagation of Sound to Great Distances." *Quart. J. R. meteor. Soc.*, 61: 285–308 (1935).
59. Whipple, F. L., "Meteors and the Earth's Upper Atmosphere." *Rev. mod. Phys.*, 15: 246–264 (1943).
60. —— Jacchia, L., and Kopal, Z., "Seasonal Variations in the Density of the Upper Atmosphere," Chap. V of *The Atmospheres of the Earth and Planets*. Chicago, University of Chicago Press, 1949.
61. Wulf, O. R., and Deming, L. S., "The Theoretical Calculation of the Distribution of Photochemically-formed Ozone in the Atmosphere." *Terr. Magn. atmos. Elect.*, 41: 299–310 (1936).
62. —— "The Effect of Visible Solar Radiation on the Calculated Distribution of Atmospheric Ozone." *Terr. Magn. atmos. Elect.*, 41: 375–378 (1936).
63. —— "The Distribution of Atmospheric Ozone in Equilibrium with Solar Radiation and the Rate of Maintenance of the Distribution." *Terr. Magn. atmos. Elect.*, 42: 195–202 (1937).

TEMPERATURES AND PRESSURES IN THE UPPER ATMOSPHERE

By HOMER E. NEWELL, Jr.

Naval Research Laboratory

Introduction

The stratosphere was discovered in the pioneering days of high-altitude balloon research. Between 1899 and 1902 experiments performed by Assmann and Teisserenc de Bort revealed a cold isothermal layer extending upwards from the top of the troposphere. The initial height of the layer appeared to vary from about 6 km in the polar regions to 18 km above the equator. As a result of such balloon measurements, it was long supposed that beyond the troposphere the temperature ceased to change with altitude. In the presence of the earth's gravitational field such an isothermal region would, of course, be characterized by a diffusive separation of its various constituents, with the heavier gases settling out and the lighter ones predominating at the higher altitudes. The name *stratosphere* aptly describes such an atmosphere.

In the course of time, however, evidence began to accumulate to show that the temperature of the atmosphere is not the same at all heights above the tropopause. On the basis of data now available from sound and meteor studies, from rocket measurements, and from a number of other sources, it is plain that atmospheric temperatures vary markedly with altitude. In higher latitudes and the polar regions an isothermal region does exist above the troposphere; but as one moves southward the purely isothermal stratosphere disappears to be supplanted by a rather flat temperature minimum somewhere between 10 and 20 km above the surface of the earth [16]. Above the (improperly named) stratosphere, in the region now referred to simply as the *upper atmosphere*, the air is definitely not isothermal. On the other hand, the total pressure difference between ground and 110 km is nearly that to be expected in an isothermal atmosphere of about 240K.[1] In this sense one may regard 240K as an average temperature for the entire region below 110 km. Above 110 km there are some indications of very high temperatures.

There are now numerous sources of data on which to base conclusions about temperatures and pressures in the upper atmosphere. Some of these are discussed in the sections below. No attempt is made, however, either to exhaust the literature or to furnish complete details of the temperature and pressure studies which are discussed.

Balloon Studies in the Upper Atmosphere

Until recently the highest-flying balloons could rise not much higher than 30 km. Measurements on such balloon flights indicated a small but definite rise in the temperature near the top of the stratosphere. Such a rise appeared consistent with temperatures deduced from studies of the anomalous propagation of sound, which suggested a sharp rise above 30 km, but which furnished only an indirect determination of temperature. Now, with improved balloons, it is possible to trace out by direct measurement an additional 10 km or more of the temperature curve.

The graph of Fig. 1 shows a curve of temperature variation from the ground up to 43 km. The measurements were made at Belmar, New Jersey, on September 28, 1948 at 1:20 P.M., using a new type of balloon especially developed for the Signal Corps [4]. It will be noted that a positive temperature gradient is observed from 15 km to the peak of the flight. The measured temperatures indicate an abrupt inversion in the neighborhood of 15 km, and show no isothermal region. They do not even appear consistent with any flat temperature minimum in the stratosphere. The measurements are

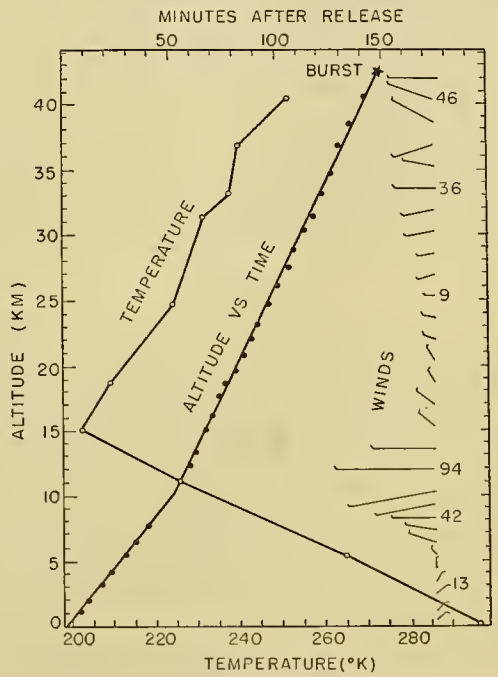


FIG. 1.—Results obtained from balloon flight at Evans Signal Laboratory, Belmar, New Jersey, on September 28, 1948, 1:20 P.M. The length of the wind vector is proportional to the wind speed (mph). The tail of the wind vector indicates the direction from which the wind was blowing. Thus, near the ground, the wind was northeasterly; at 15,000 feet, northerly; at 40,000 feet, westerly; and at 50,000 feet northwesterly (taken from [4] Brasefield: *Sci. Mon.*, 68:398 (1949), by permission of the publishers).

1. This statement is based on the pressure curve of Fig. 7 below.

in keeping with both balloon and rocket measurements at White Sands, New Mexico, which likewise fail to show any genuinely isothermal layer [9].

Propagation of Sound Through the Upper Atmosphere

Occasionally sound generated by an explosion can be heard distinctly at remote points, but not at all at points nearer to the source. The total region of audibility consists of a primary zone containing the source itself, plus one or more surrounding rings separated from each other and from the primary zone by regions of silence. Throughout the primary zone the sound is propagated at normal speeds. Within the secondary zones, however, the sound arrives much later than would be expected of waves propagated along the earth's surface.

Anomalous sound propagation was first noted in connection with gunfire at Queen Victoria's funeral. It was soon concluded that waves reaching the annuli of abnormal audibility had traveled into and through the upper atmosphere before returning to earth. An early explanation was that high-altitude winds caused the waves to bend back to the ground. The observed omnidirectionality of anomalous sound propagation, however, was not compatible with the wind theory.

Von dem Borne proposed an explanation based on refraction of sound waves [3]. If atmospheric properties vary only with height, the refraction law is

$$\frac{v}{\cos e} = V, \tag{1}$$

where for a given path V is constant. The quantities v and e are respectively the local speed of sound and the angle between the ray path and the horizontal plane. The ray paths are straight, curve upward, or bend downward according as v is constant, decreasing, or increasing with altitude. Assuming refraction as the cause of anomalous propagation, one concludes that sound returning to earth from the upper atmosphere passed through a region in which its speed increased with altitude. Moreover, at the apex of its path, the wave front was moving horizontally at the speed V.

The speed of sound through a gas is given by

$$v = \sqrt{\gamma \frac{R}{M} T}, \tag{2}$$

where γ is the ratio of the specific heat at constant pressure to that at constant volume, R is the universal gas constant, M the average molecular weight of the gas, and T its absolute temperature. To explain anomalous sound propagation, von dem Borne postulated a decreasing value of M with height caused by increasing proportions of the lighter gases above the troposphere [3]. Data available at present, however, show far too small a change in composition to vary either M or γ sufficiently to account for the observed refraction, which must accordingly be due to changes in the only remaining variable, namely temperature.

In 1923 F. J. W. Whipple first suggested an explanation of anomalous sound propagation based on a positive temperature gradient in the upper atmosphere [19]. Since then Whipple, Gutenberg, Duckert, and numerous other authors have written extensively on the subject [5, 6, 8]. Whipple's explanation is the one now generally accepted, and leads directly to a method for determining upper-air temperatures.

Referring to Fig. 2, suppose that within a zone of abnormal audibility P_1 and P_2 are neighboring points

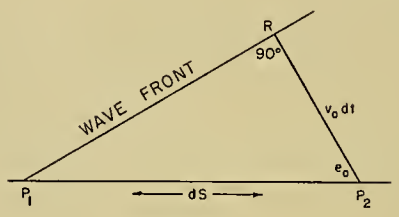


Fig. 2.

in line with the sound source. At the moment the wave front reaches P_1, it still has a distance RP_2 to travel before reaching P_2. If dt is the time interval between the arrival of the sound at P_1 and its arrival at P_2, and if v_0 is the speed of sound in the neighborhood of P_1 and P_2, then RP_2 is equal to $v_0\,dt$. Letting $dS = P_1P_2$, one has:

$$dS \cos e_0 = v_0\,dt;$$

or, using equation (1):

$$\frac{dS}{dt} = \frac{v_0}{\cos e_0} = V. \tag{3}$$

The quantity dS/dt is the apparent speed of sound along the earth between P_1 and P_2, and can be measured quite easily. Moreover, as shown by (3), the measured value of dS/dt is precisely the speed of sound V at the apex of its path through the upper atmosphere. The value of V can also be obtained by measuring the angle of arrival e_0 of the wave front.

Once determined, V can be used in equation (2) to compute the temperature at the apex of the sound path. To associate the temperature so calculated with a specific height, the ray path must be traced out in order to determine the altitude of the apex. This can be done quite accurately at lower altitudes, using data from balloon observations. Assuming a positive temperature gradient, the remainder of the path in the upper atmosphere is then constructed so as to fit the observed transit times for the sound and so as to join smoothly onto the lower segments.

Temperatures determined by Gutenberg, and more recently by Cox, are shown in Figs. 3 and 4 [8, 5]. Between 40 and 60 km the temperatures quoted by Cox are considerably below those obtained by Gutenberg. Cox's results are more in conformity with rocket findings at White Sands. A temperature maximum, such as that shown by Cox near 55 km, has always been observed in rocket flights.

Winds may have a large effect upon temperatures deduced from sound observations. With this in mind

Weekes and Wilkes have analyzed sound data obtained from more accurate controlled experiments carried out by the British Meteorological Office. They found that

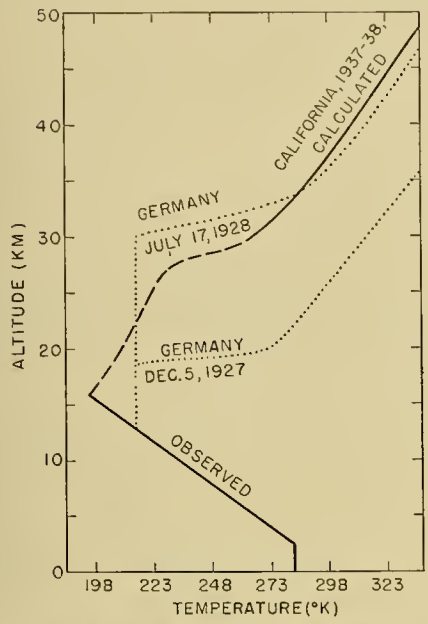


FIG. 3.—Temperatures derived from velocity of sound waves in Southern California, with results for Northern Germany plotted for comparison (taken from [8, p. 200]).

at night "the temperature starts to rise at about 35 km and rises at a rate of about 5C km^{-1} to at least 50 km, at which height the temperature lies between 290 and 350K, with the lower value more likely." [18, p. 87].

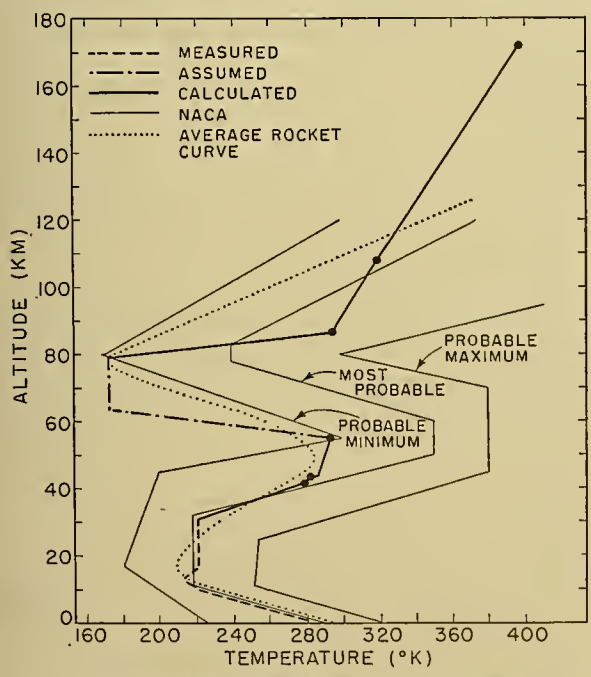


FIG. 4.—Comparison of measured upper-atmosphere temperatures (Helgoland blast), V-2 rocket results, and N.A.C.A. tentative standards (taken from [5] Cox: *Amer. J. Phys.*, 16: 473 (1948), by permission of the publishers). Rough rocket data appearing on Cox's published figure have been replaced by an average curve of temperatures calculated from Naval Research Laboratory rocket flights.

The lower values are more in keeping with Cox's results and those obtained from rockets.

Ozone Heating in the Upper Atmosphere

The high-temperature region deduced from sound-propagation experiments coincides with the upper portion of the ozonosphere, which lies between 15 and 55 km. The total amount of ozone in the atmosphere is at most a few millimeters at standard conditions. Nevertheless ozone absorption in the ultraviolet is strong, and the ozone content appears to be sufficient to account for observed atmospheric heating immediately above the stratosphere.

Gowan [7] and Penndorf [15] have made extensive calculations to show how much heating can be expected from atmospheric ozone. Gowan assumes the ozonosphere to be essentially nonconvective and in radiative

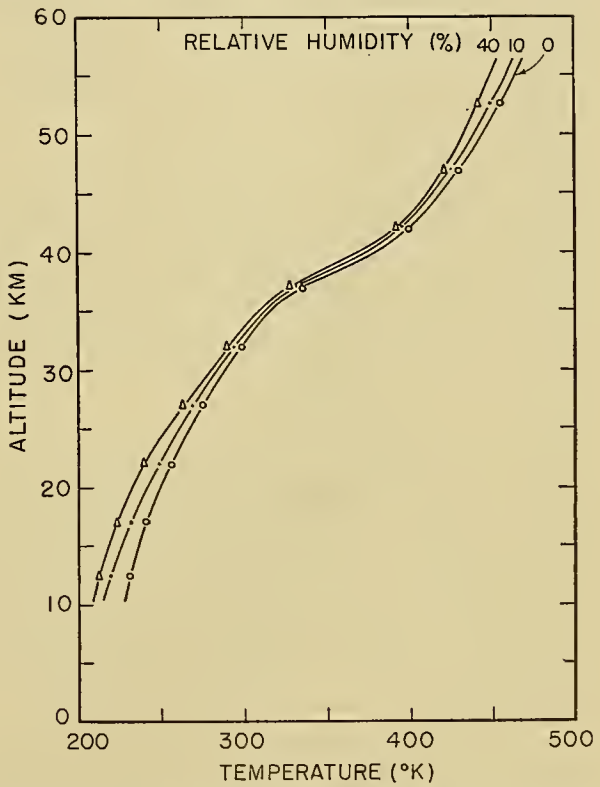


FIG. 5.—Atmospheric temperatures based on calculations of ultraviolet absorption in the ozonosphere (taken from [7] Gowan: *Proc. roy. Soc.*, (A) 190: 223 (1947), by permission of The Royal Society).

equilibrium. The validity of these assumptions, however, is open to question. Water vapor, ozone, and carbon dioxide are taken to be the principal gases involved, and the calculations are made for a number of different humidities. The assumed solar temperature is 6000K. Ozone content is taken from experiments of Götz, Meetham, and Dobson and applies to the atmosphere above Switzerland. A temperature-distribution curve is obtained by dividing the ozonosphere into nine layers and calculating the equilibrium temperatures for the various layers. Some of the results are set forth in Fig. 5. It will be noted that the temperature begins to

rise at somewhat below 30 km, and at 50 km is still rising. Above 42 km, however, the rate of rise falls off. The temperatures are very much higher than those given by Cox and by rocket soundings at White Sands.

Penndorf [15] does not assume radiative equilibrium and considers only ozone heating and cooling. He considers the effect of water vapor in the ozonosphere to be negligible. Above 3000 A the solar radiation is taken from Abbot's table; from 3000 A to 2000 A the radiation is assumed to be that of a black body at 5910K. Rates of heating and cooling in the ozonosphere are computed for different temperature distributions. Penndorf's results on daytime heating and nighttime cooling correspond to rather high equilibrium temperatures during the day, with little change during the night. According to Penndorf the maximum rate of heating and most of the ultraviolet absorption both occur near 50 km.

Gowan's and Penndorf's calculations agree qualitatively with sound propagation and rocket results in showing a temperature rise in the upper ozonosphere. The computed ozone heating, however, is much greater than Cox's measurements and rocket findings would indicate.

At present, calculations of ozone heating involve so many uncertainties that conclusions based upon them must be regarded with considerable caution. It appears highly desirable to perform a number of rocket experiments to measure the total energy absorbed at various atmospheric levels from the solar and terrestrial radiations. Such data would provide a firmer foundation for temperature calculations than is now available.

Free Oscillations in the Atmosphere

Observed amplitudes of solar tides produced in the earth's atmosphere are about one hundred times as great as one normally would expect. Kelvin suggested that this might be a resonance phenomenon, due to a free atmospheric oscillation period of 12 hr. Evidence also is available to show the existence of a second free period in the atmosphere. The speed of large explosion waves, such as those generated by the Krakatau eruption and the Great Siberian Meteor, corresponds to a period of 10.5 hr.

It is to be expected that the natural periods of the atmosphere are directly associated with its temperatures. Pekeris has, in fact, shown that an atmosphere with the temperature distribution set forth in curve 7 of Fig. 8 possesses both the 10.5- and the 12-hr periods [14]. His results are further substantiated and clarified by later work of Weekes and Wilkes [18]. The 10.5-hr period is a property of the lower atmosphere with the temperature distribution revealed by balloon measurements. The 12-hr period, on the other hand, is associated with the upper atmosphere, and requires the indicated temperature drop above the hot ozonosphere. Pekeris' conclusions, based upon analysis of a number of special cases, were that the atmosphere must become cold again at or above 80 km.

Meteors and Atmospheric Temperatures

A meteor striking the earth's atmosphere is heated by impact with the molecules of air. Depending upon the speed, size, and material of the meteor, the heating causes incandescence along some portion of the meteor path. The altitudes at which incandescence begins and ends depend also upon the density of the atmosphere.

In a pioneering work on meteor studies, Lindemann and Dobson [12] developed a theory which enabled them to calculate atmospheric densities at the heights of appearance and disappearance of a meteor from its speed, size, and composition. The theory applies to an isothermal atmosphere, whereas it now appears that the region in which meteors are observed visually is definitely not isothermal. Nevertheless, the theory can provide a curve of densities from which an average temperature can be estimated. The densities obtained by Lindemann and Dobson were not consistent with an average temperature of 220K in the neighborhood of stratospheric values, but were consistent with a much higher temperature, about 300K.

More recently F. L. Whipple [20] presented a theory relating the mass, shape, composition, speed, deceleration, and luminosity of a meteor to atmospheric densities. From a study of available photographic meteor data he obtained a curve of log-density versus altitude, with which a number of temperature distributions can reasonably be associated. In his paper Whipple states:

> The best solution appears to be one in which the height-log ρ curve corresponds to a flat temperature maximum of about 375K near the 60-km level, a rapid drop to 250K near 80 km, and a constant or slowly rising temperature at greater heights to about 110 km.

Rocket Measurements in the Upper Atmosphere

Since the spring of 1946 rockets have been employed at White Sands, New Mexico, for measuring pressures and temperatures in the upper atmosphere [1, 2, 10]. Initially the German V-2 was the only large rocket available, but at present the V-2 and the Navy's Aerobee and Viking are all being used. So far, pressure measurements extend only to 130 km, although some of the rockets have exceeded 100 miles in altitude. Instead of being measured directly, atmospheric temperatures are computed from the pressure data.

On each rocket there is an area, usually just ahead of the tail fins, where ambient pressures exist during flight. Gages are mounte in this area to measure atmospheric pressures directly. As a check on the accuracy of readings from gages so mounted, it is customary to compare measurements at low altitudes with balloon data acquired just before and just after the rocket flight. In the past there has been excellent agreement between rocket and balloon measurements.

Gages are also mounted at various points on the nose of the rocket. Measurements from such gages must be converted to ambient pressures by some method such as application of the Taylor-Maccoll theory. Stagnation pressures, recorded with gages at the nose tip, provide essentially a measure of density.

Temperatures are computed from the pressure versus altitude curve using the relation

$$\frac{1}{p}\frac{dp}{dh} = -\frac{Mg}{RT},$$

where M is the average molecular weight of the air, R is the universal gas constant, g the acceleration of gravity, T the absolute temperature of the air, h the altitude, and p the measured pressure. Temperatures can also be calculated from the stagnation pressure at the nose, using the following simplification of Rayleigh's formula for supersonic speeds:

$$\frac{\text{stagnation pressure}}{\text{ambient pressure}} = \cdots + .46 + 1.29 \left(\frac{\text{speed of rocket}}{\text{speed of sound}}\right)^2,$$

and using the known speed of the rocket to obtain the local speed of sound. Once the speed of sound is known, the temperature is quickly deduced.

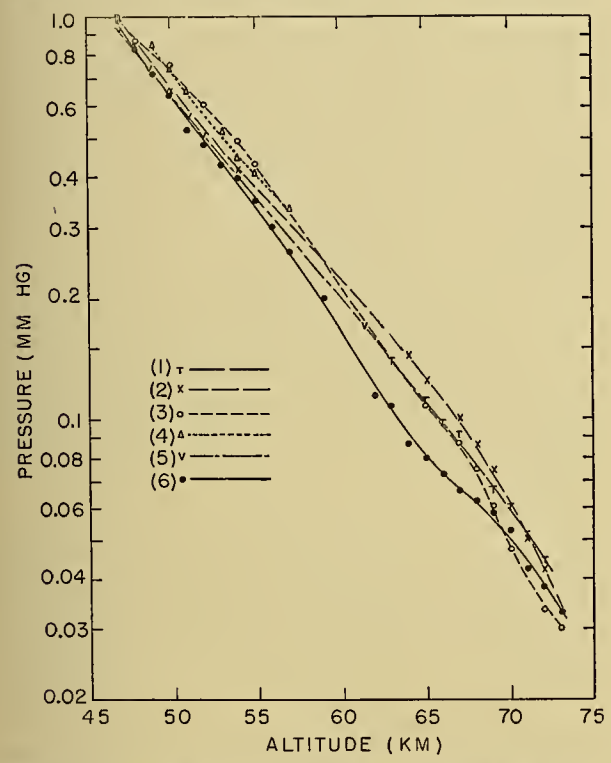


FIG. 6.—Pressures obtained from rocket flights above White Sands Proving Ground, New Mexico. Except for the point at 61.3 km which is 0.17 ± 0.03 mm of *Hg*, the data are probably correct to within 10 per cent of ambient. The flights are: (1) 10 October 1946, 1102 MST; (2) 7 March 1947, 1123 MST; (3) 22 January 1948, 1313 MST; (4) 5 August 1948, 1837 MST; (5) 28 January 1949, 1020 MST; (6) 3 May 1949, 0914 MST (taken from [10]).

Three types of gages were employed in making rocket pressure measurements. For pressures down to 50 mm of *Hg*, bellows gages were used. Pirani-type gages measured pressures in the interval from 2 mm to 7×10^{-3} mm of *Hg*. Philips gages were employed for the range from 10^{-3} down to 10^{-5} mm of *Hg*. Unfortunately there is a gap between the ranges covered by the bellows and Pirani gages, and this gap exists in the data obtained so far. Moreover, the Philips gage measurements were not too accurate in their range, partly because of peculiarities of the gage, and partly because of the behavior of the rocket. At altitudes above 80 km the rocket pressure measurements were influenced to varying extents by motions of the missile and by emission of gas from the interior and surface of the rocket. Because of differences in missile behavior, it was necessary to make separate estimates for each rocket of the errors introduced by pitching and yawing. In many cases emission of gas placed an upper limit on the altitude to which pressures could be measured.

Recently a new type gage invented by R. J. Havens [11] has been introduced into the rocket work. With this gage it should be possible in the future to cover the entire pressure range up to 150 km or more.

Pressure and temperature data accumulated by R. J. Havens and his colleagues at White Sands during the past several years are shown in Figs. 6, 7, and 8.

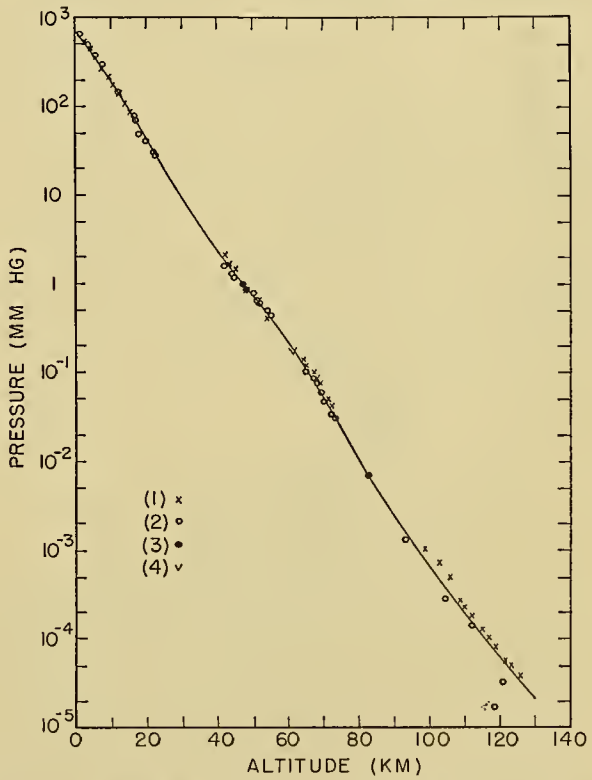


FIG. 7.—Pressures obtained from rocket flights above White Sands Proving Ground, New Mexico. The point at 82.4 km is 0.007 ± 0.002 mm of *Hg* and that at 61.3 km is 0.17 ± 0.03 mm of *Hg*. These were taken at the top of the rocket flight and pressure was therefore ambient within the error of the measurement. All other data up to 80 km are probably within 10 per cent of ambient. Data obtained above 90 km are not ambient due to the yaw of the rocket and the location of the gages. It is estimated that they are within a factor of two from ambient. (1) 7 March 1947, 1123 MST; (2) 22 January 1948, 1313 MST; (3) 3 May 1949, 0914 MST; (4) 28 January 1949, 1020 MST (taken from [10]).

The temperatures are based on the assumption of sea-level composition throughout.

Some of the differences in the pressure curves of Fig. 6 could be due to seasonal and diurnal changes, and possibly to variations in solar conditions. The same is true of the temperature curves of Fig. 8. In the latter case, however, the differences are harder to pin down

because of inaccuracies introduced in differentiating the pressure curves to obtain temperatures. Havens estimates that the pressure measurements permit the determination of an average temperature for a layer 20 km thick correct to within ±10K. The error for a 10-km layer would be on the order of ±20K. The curves of actual temperature variation with altitude must, therefore, be regarded as only indicative.

The double hump on curve (5), Fig. 8, appears to be real. Its presence was suspected in the reduction of data on other flights, and strongly suggests two distinct heating processes. The lower hump is due to ozone absorption in the ultraviolet above 2000 A. The upper hump may be due to oxygen absorption in the ultraviolet below 2000 A.

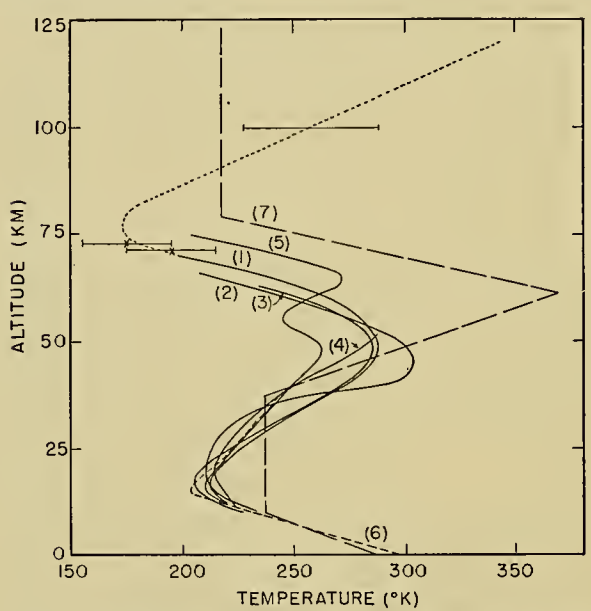


Fig. 8.—Temperatures calculated from pressures measured on rocket flights above White Sands Proving Ground, New Mexico. The points indicated by × with their probable errors are temperatures calculated from the ratio of stagnation pressure at the nose of the rocket to ambient pressure. Sea-level composition is assumed throughout. The curves in the drawing are: (1) 7 March 1947, 1123 MST; (2) 22 January 1948, 1313 MST; (3) 5 August 1948, 1837 MST; (4) 28 January 1949, 1020 MST; and (5) 3 May 1949, 0914 MST (taken from [10]). Also shown are: (6) Balloon flight shown in Fig. 1; and (7) Temperature distribution assumed by Pekeris for calculation of atmospheric oscillations (taken from [14] Pekeris: *Proc. roy. Soc.*,(A) 158: 653 (1937), by permission of The Royal Society).

The increase above 80 km indicated by the dotted extension of curve (1), Fig. 8, was drawn so as to have the average value of about 260K between 80 and 120 km. This average value, indicated at the 100-km level, is the sole calculated point on which the dotted portion of the curve is based. The heating of this atmospheric level is possibly due to absorption of solar X-radiation such as that recently discovered in V-2 flights by T. R. Burnight.

Temperatures and Pressures at Extreme Altitudes

Rocket measurements [1, 2, 10] and Cox's sound propagation studies [5] indicate a continually increasing temperature above the minimum near 80 km. The former measurements extend to about 130 km, and the latter to about 172 km, although Cox states that he places little faith in the highest point on his curve. According to Whipple a positive temperature gradient above 80 km is consistent with data from meteor studies. From the energy distribution in the nitrogen bands of the auroral spectrum Rosseland and Steensholt [17] report a temperature of 347K (at 110 km approximately), correcting a much lower value obtained earlier by Vegard.

It appears reasonable to assume a general temperature increase to great heights, with a corresponding lessening in the rate of decrease in pressure. For one thing, if the temperature were essentially constant above 100 km, the atmosphere at 400 km would be practically nonexistent, as a simple calculation will show, whereas observations of aurorae indicate appreciable densities at these heights. Secondly, helium appears to be escaping from the earth's atmosphere at a rate which requires temperatures on the order of 1000K between 600 and 800 km [13, pp. 21–24]. Thirdly, Babcock's measurements of the width of the night-sky line 5577 A indicate temperatures on the order of 1200K at several hundred km [13, p. 509].

In comparison with information available on the atmosphere below 100 km, data on temperatures and pressures at higher altitudes are meagre indeed. To be sure there is an appreciable amount of spectroscopic data; and ionospheric studies also indicate high temperatures in the ionospheric layers. Nevertheless, at present, conditions in the outer reaches of the earth's atmosphere are largely a matter of speculation.

Conclusion

Even a casual survey of the literature shows wide differences in the temperatures and pressures ascribed to the upper atmosphere. This is apparent in the limited selection of material presented above. Some of the differences can be traced to diurnal and seasonal effects and to geographical differences. Also atmospheric conditions are closely associated with conditions in the sun, which fluctuate continually. On the other hand, most of the methods of determining atmospheric temperatures and pressures cannot be proved free of systematic errors. The rather involved ozone calculations of Gowan and Penndorf, for example, and Whipple's meteor theory, contain many uncertainties which cannot be resolved without additional information. The most direct measurements are those made by balloons and rockets.

It would be worth while to compare critically the various methods of determining pressures and temperatures in the upper atmosphere, at least to point up systematic differences in the results. At present, accomplishment of such a comparison is hindered by lack of any single standard. To provide such a standard, data should be available from a single credible method covering different times and seasons and a wide range of geographical and solar conditions.

Because of their on-the-spot character, rocket meas-

urements appear to hold the greatest promise for determining conditions in the upper atmosphere. Recent improvements in technique and the lessons of past experience indicate that good results should be forthcoming in the next few years. With the mobility afforded by the Navy's *Norton Sound*, the measurements now being made at White Sands can be extended to widely separated parts of the world. The three branches of the Armed Forces are devoting considerable attention to this phase of rocket research in the upper atmosphere. Their efforts should aid materially in solving the temperature-pressure problem, at least for altitudes below 150 km. A critically comparative survey of the field is probably best left in abeyance until more work has been done both with rockets and with other methods.

A summary of current knowledge about atmospheric temperatures may be found from the N.A.C.A. Tentative Standard Atmosphere, drawn up in 1947. This curve was shown in Fig. 4 for comparison with Cox's sound propagation results. At the time it was constructed, the tentative standard was intended to be a rough average of data then available. The wide variation in temperature data was reflected in the upper and lower limits provided with the standard curve. Rocket results and Cox's sound propagation measurements, obtained since issuance of the standard, indicate that the temperatures should be lowered in the neighborhood of 55 km to less than 300K. The rocket measurements also reveal quite low temperatures near 80 km, perhaps as low as 180K. This is in keeping with temperatures required for formation of noctilucent clouds seen near 80 km in the higher latitudes. At present, however, the quantity of rocket data is too small, and the question of temporal, solar, and geographic influences too unsettled, to warrant changing the standard curve now.

Evidences of high temperatures, on the order of 1000K or more, at several hundred kilometers and above, were presented in the preceding section.

The best curves of pressure variation with altitude are probably those of Figs. 6 and 7, obtained from rocket soundings.

The rough sketch in the preceding paragraphs oversimplifies the true picture, as plainly it must, considering the large number of variables involved. None of the measurements to date are adequate to pin down small local phenomena which may exist. Mother-of-pearl clouds, for example, appearing at 22 to 30 km, indicate a region of cooling at the top of the stratosphere. It is not unlikely that such local temperature variations come and go with time. The double hump shown in the Viking temperature curve of Fig. 8 appears real. Hints of such a phenomenon were seen in the reduction of other rocket data. As pointed out above, the presence of the two maxima suggests the possibility of two types of atmospheric heating. Possibly the lower altitude maximum was due to ozone heating primarily, and the upper one to oxygen absorption of ultraviolet radiation below 2000 A. Then again, the phenomenon may be due simply to mass motions in the air. Certainly wind structure and temperature of the atmosphere are closely associated. In particular, the effect of mixing or lack of mixing upon composition of the air should influence atmospheric temperature-pressure relationships appreciably, although at present it appears that mixing below 80 km is adequate to prevent any significant diffusive separation of atmospheric constituents. Heating in the E-layer of the ionosphere may be caused by an influx of X-radiation from the sun.

Even a brief consideration of the complexity of the atmosphere serves to emphasize the need for an abundance of accurate data before a truly complete understanding of the atmosphere can be had. Specifically:

1.) Further flights with balloons to maximum attainable altitudes are highly desirable. Such flights should be made over as wide a range of time and geographical position as is possible.
2.) Extensive sound propagation studies in various parts of the world, with especial care to eliminate errors due to winds, should be made to obtain further temperature data.
3.) Rocket pressure-temperature measurements should be extended to reach different latitudes and to cover the various times of day and year.
4.) Whenever possible, the experiments listed above should be conducted at the same time and place, and the results from the three methods carefully compared.
5.) Absorption of terrestrial and solar radiations at the various altitudes should be measured in rocket flights to provide a basis for calculations of atmospheric heating. These measurements also should cover a wide temporal and geographical range.
6.) Rocket and balloon measurements of wind structure in the upper atmosphere are essential.
7.) Before the pressure-temperature problem in the upper atmosphere can be completely solved, atmospheric composition at the various altitudes must be determined. It is to be hoped that rocket soundings can aid in this.
8.) Eventually a careful, critical comparison of the various methods of determining atmospheric pressures, temperatures, and densities should be made to point up systematic differences, and to "calibrate" the different methods, so to speak. With the *Norton Sound* the Armed Forces may be able within the next several years to provide enough rocket data to form a standard for comparison. Needless to say, not only the methods discussed above, but also others, such as ionospheric and spectroscopic studies, should be included in the comparison.

REFERENCES

1. Best, N. R., Durand, E., Gale, D. I., and Havens, R. J., "Pressure and Temperature Measurements in the Upper Atmosphere." *Phys. Rev.*, 70: 985 (1946).
2. Best, N. R., Havens, R., and LaGow, H., "Pressure and Temperature of the Atmosphere to 120 km." *Phys. Rev.*, 71: 915–916 (1947).
3. Borne, G. v. d., "Über die Schallverbreitung bei Explosionskatastrophen." *Phys. Z.*, 11: 483–488 (1910).

4. Brasefield, C. J., "Exploring the Ozonosphere." *Sci. Mon.*, 68: 395–399 (1949).
5. Cox, E. F., "Upper Atmosphere Temperatures from Remote Sound Measurements." *Amer. J. Phys.*, 16: 465–474 (1948).
6. Duckert, P., "Über die Ausbreitung von Explosionswellen in der Erdatmosphäre." *Beitr. Geophys.*, (Snpp.) 1: 236–290 (1931).
7. Gowan, E. H., "Ozonosphere Temperatures Under Radiation Equilibrium." *Proc. roy. Soc.*, (A) 190: 219–226 (1947).
8. Gutenberg, B., "The Velocity of Sound Waves and the Temperature in the Stratosphere in Southern California." *Bull. Amer. meteor. Soc.*, 20: 192–201 (1939).
9. —— "New Data on the Lower Stratosphere." *Bull. Amer. meteor. Soc.*, 30: 62–64 (1949).
10. Havens, R. J., Koll, R. T., and LaGow, H., *Pressures and Temperatures in the Earth's Upper Atmosphere.* Nav. Res. Lab. Rep., Washington, D. C., March, 1950.
11. —— "A New Vacuum Gage." *Rev. sci. Instrum.*, 21: 596–598 (1950).
12. Lindemann, F. A., and Dobson, G. M. B., "A Theory of Meteors, and the Density and Temperature of the Outer Atmosphere to Which It Leads." *Proc. roy. Soc.*, (A) 102: 411–437 (1923).
13. Mitra, S. K., *The Upper Atmosphere.* Calcutta, Royal Asiatic Society of Bengal, 1948. (See pp. 21–24, 509)
14. Pekeris, C. L., "Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 158: 650–671 (1937).
15. Penndorf, R., "Beiträge zum Ozonproblem, Die Rolle des Ozons im Wärmehaushalt der Stratosphäre." *Veröff. geophys. Inst. Univ. Lpz.* (2nd ser.), 8: 181–285 (1936).
16. Ratner, B., *Temperature, Pressure, and Relative Humidity over the United States and Alaska.* Washington, U. S. Dept. of Commerce, 42 pp., 1945.
17. Rosseland, S., and Steensholt, G., "On the Relative Intensity of Bands in a Sequence and the Temperature of the Upper Atmosphere." *Univ. Obs., Oslo,* Pub. No. 7, 17 pp. (1933).
18. Weekes, K., and Wilkes, M. V., "Atmospheric Oscillations and the Resonance Theory." *Proc. roy. Soc.*, (A) 192: 80–99 (1947).
19. Whipple, F. J. W., "The High Temperature of the Upper Atmosphere as an Explanation of Zones of Audibility." *Nature, Lond.*, 111: 187 (1923).
20. Whipple, F. L., "Meteors and the Earth's Upper Atmosphere." *Rev. mod. Phys.*, 15: 246–264 (1943).

WATER VAPOUR IN THE UPPER AIR

By G. M. B. DOBSON and A. W. BREWER

The Clarendon Laboratory, Oxford

INTRODUCTION

Water vapour in the upper air plays an important part in the dynamics of the atmosphere because (1) it is a source of atmospheric energy, (2) its presence affects the release of energy, (3) it is the origin of all hydrometeors, and (4) it is the main constituent of the lower atmosphere to absorb infrared radiation. In addition the water-vapour content of the air can be used in meteorological studies as a natural "tracer" element.

The importance of water vapour as a source of atmospheric energy has been discussed by Normand [12], but generally when the temperature is below about −20C the water vapour is not thermodynamically significant. All hydrometeors originate from the condensation of water vapour in the upper air, but the production of large amounts of rain or snow requires the larger vapour pressures, which occur only at temperatures above about −20C. Condensation at lower temperatures than this may be of importance in inaugurating the colloidal instability of a cloud whereby the liquid cloud droplets can fall as snow, but the quantities of water contributed by the regions of low temperatures will not usually be significant.

Measurements of the humidity of the free air have been made for many years using, for example, hair or goldbeater's skin hygrometers, or wet- and dry-bulb psychrometers, but these are useful only when the amount of water vapour in the air is large. We do not propose to discuss these phenomena, which are already well known.

For the remaining processes listed in the first paragraph, water vapour is important even if present in very small amounts, but since there has been no hygrometer capable of measuring very small water-vapour pressures, this field of study has received little attention. Since 1943 it has been possible, by a simple development of the well-known dew-point hygrometer, to obtain reliable measurements of the water content of the upper air at all levels which multiseated aircraft can reach, and a large number of measurements up to 12 or 13 km have been made by the Meteorological Research Flight of the British Meteorological Office.

No entirely reliable measurements of the distribution of water vapour at levels above about 13 km have yet been made, but the values obtained by Regener [14] appear to be very high.[1] The rare mother-of-pearl clouds which have been described by Störmer [17] are usually formed at about 25 km and appear to be very tenuous water clouds, which would indicate that at times the air may be saturated at these levels.

1. This was written before the results of Barrett, Herndon, and Carter [1] were available. Their results are discussed below.

THE MEASUREMENT OF WATER VAPOUR IN THE UPPER ATMOSPHERE

The difficulty of measuring the water-vapour content of the upper air lies in the extremely small amount of water vapour present, and sometimes even at moderate levels the air is too dry to permit measurement of its water content with ordinary hygrometers (see Fig. 9). One cubic metre of saturated air contains only 26 mg of water at a temperature of 220K, 1.6 mg at 200K, and 0.3 mg at 190K. Water-vapour densities of the order of 1 mg m^{-3} are found in the stratosphere over southern England. Since it is possible to deal with only a small fraction of a cubic metre of air, it will be realized how very small an amount of water or ice has to to be measured, and it is for this reason that most of the usual methods of measuring humidity fail when used in the upper air.

Glueckauf [8] showed that a very thin film of polyvinyl acetate containing a hygroscopic salt could be used as a very rapid hygrometer. The thickness of his films was about 0.1 μ; the films changed their thickness rapidly with changing relative humidity (with respect to *water*). If illuminated by monochromatic light, the thickness of the film could be estimated from the reflection coefficient. Unfortunately these films are very delicate, and the method has not come into general use Glueckauf [9] has also obtained humidity measurements at high levels by correcting the hair hygrometer of a Dines meteorograph for its lag.

Another possible hygrometer involves the use of a Wilson cloud chamber. Air with suitable nuclei is supplied to the chamber and expansions of increasing ratio are made until a fog is just formed. The first formation of a fog can be detected by the scattering from a strong beam of light passing through the cloud chamber. This method has not yet been fully developed, but tests made by Palmer [13] show that it should be effective down to a frost point of about 210K, below which the fog formed is too thin for convenient observation.

A method of measuring the water content of the high atmosphere which has some unique advantages is to measure the absorption of the sun's infrared energy by water vapour. Because of the very small amount of water vapour in the upper atmosphere it is necessary to use the 6.3-μ absorption band of water vapour since, except for the absorption band at 2.7 μ, the bands of shorter wave lengths have absorptions which are too small. The 2.7-μ band is confused with a CO_2 absorption band and cannot be used.

The absorption gives information about the total amount of water vapour above the height of observation, and the vapour pressure at various levels must be obtained by differences. If these differences can be com-

pared with frost-point hygrometer measurements made at the same time, they should provide a useful check on the infrared absorption data. The method is being developed by G. B. B. M. Sutherland and R. M. Goody at Cambridge, England.

Simon [16] has suggested that enough air to allow accurate analysis could be condensed in a small container cooled in liquid hydrogen and carried by a balloon to the height at which the sample is to be collected. This method is now being developed by F. E. Simon and A. J. Peckover at Oxford.

The dew- or frost-point hygrometer has been found satisfactory even at low temperatures. It has the great practical advantage that it needs only a very simple calibration, since all that is required is the temperature of the thimble when a thin deposit of hoarfrost is neither increasing nor decreasing. At −80C it is necessary to detect deposit changes of the order of 10^{-8} grams of ice, corresponding to a uniform deposit on the thimble surface of the order of thickness of one atomic diameter. This is not impossible because the ice is deposited as a large number of small crystals which scatter light very efficiently, so that changes in the deposit can be detected by changes in the light which it scatters. This will be understood, since only 10^{-10} grams is equivalent to 1000 crystals each 0.4 μ across. With proper instrument design the change in the deposit can be detected by eye, though the observations are difficult when the frost points to be measured are below about 200 to 210K. A photoelectric method of estimating the amount of light scattered by the deposit, though more complicated, is usually preferable.

DESIGN OF A FROST-POINT HYGROMETER

Three types of instruments have been employed in Great Britain: (1) an instrument in which the deposit is observed visually, and the cooling of the thimble performed manually, (2) an instrument in which the amount of deposit is indicated photoelectrically, and the cooling operated by hand, and (3) an automatic instrument in which the thimble is held constant at the frost point.

If the instrument is to measure the frost point of very dry air accurately (*e.g.*, frost points down to 190K), there are certain conditions which must be observed:

1. *Air Supply.* A constant source of trouble, when measuring very dry air, is the possibility that the air might pick up water vapour from the walls of piping leading to the hygrometer. In aircraft observations this error is guarded against by passing a large and rapid current of air through the tubing leading to the hygrometer. Most of this air passes out of the aircraft again and only a small fraction is used in the hygrometer. The branch pipe actually leading to the thimble is kept very short.

2. *Determination of the Amount of Deposit.* As stated above, the only way in which it has been found possible to observe the very small deposit which is obtained with very dry air is to illuminate the surface of the thimble strongly and measure the light scattered by the deposit. The surface of the thimble must be highly polished so that little light is scattered by it. If the deposit is detected by a photoelectric cell, the light scattered into the cell by the clean thimble must be kept as small as possible, so that the additional light scattered by a faint deposit is as large a fraction as possible of the total light entering the photocell.

3. *Temperature Control of the Thimble.* For the hand-operated hygrometer the thimble has been cooled by cold petrol (cooled in a surrounding bath of petrol and solid CO_2), or by liquid air, forced as a jet against the hollow underside of the thimble. With suitable design it is easy to control the temperature of the thimble by varying the rate of pumping. An alternative—which is always used in the automatic hygrometer—is to cool the thimble by conduction to liquid air and to warm it by an adjustable heating current.

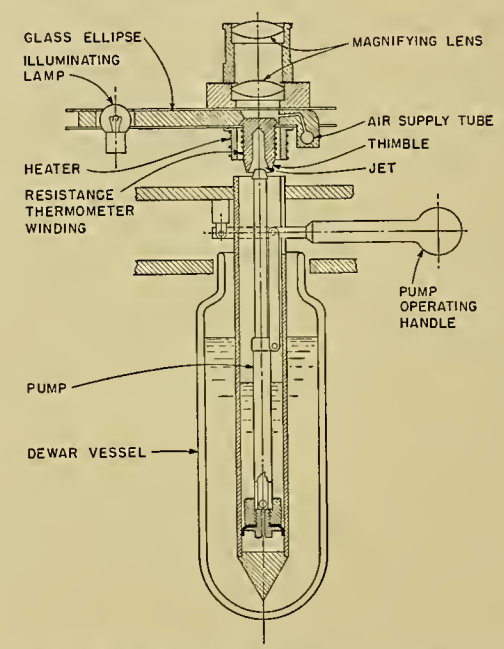


FIG. 1.—The frost-point hygrometer. A deposit of dew or hoarfrost on the top surface of the thimble is illuminated obliquely and observed by eye through a ×7 magnification. The thimble temperature is controlled by operation of the pump and measured by a platinum resistance wire wound on the thimble skirt. (*Reproduced by courtesy of the Royal Society, London.*)

In the automatic hygrometer the scattered light from the thimble is received by a photocell or photomultiplier. A second photocell or multiplier receives a constant small light from the same source which illuminates the deposit, and the difference of the photocurrents for the two multipliers is amplified. To avoid hunting, a differentiating circuit is usually necessary so that the final control of the heating current depends both on the amount of the deposit and on its rate of growth (or evaporation).

One of the difficulties with which an automatic hygrometer must contend is the very extended range of the absolute humidity of the air which it must measure. Also, the rate of growth of ice crystals decreases progressively at temperatures below about 218K, and for practical purposes crystals do not grow at temperatures

lower than 180K. Thus if the thimble should be cooled below this temperature, the ice will form as a glassy layer which scatters no light.

In using the eye-observation hygrometer it is essential that two thimble temperatures be found, one at which the deposit is slowly evaporating and the other at which it is growing equally slowly. The mean of the temperatures will be the frost point. Since a deposit will usually not form on a clean thimble until it is cooled to near the dew point, large errors will be made if the temperature at which the deposit is first seen is used. The temperature at which the deposit finally evaporates is also of no significance. With photoelectric indication of the amount of deposit, one can adjust the thimble temperature until the deposit indicator is steady and then read the thimble temperature, which gives the frost point directly.

Figures 1 and 2 show the general design of the types of hygrometer used in Great Britain [4, 6].

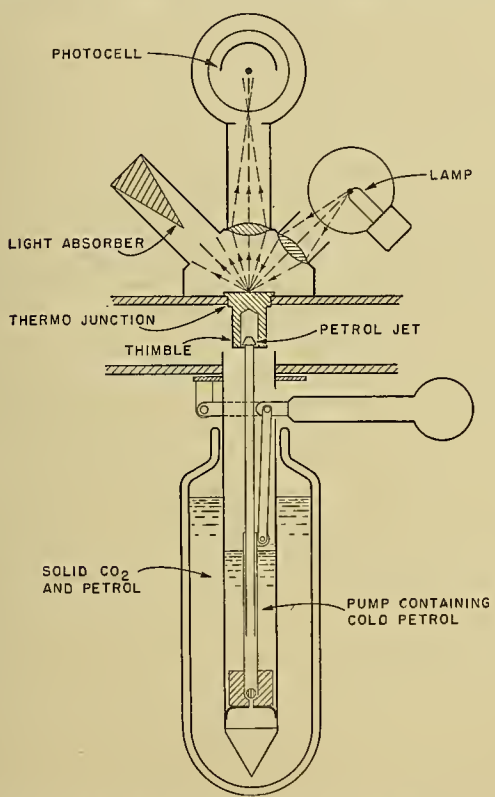


Fig. 2.—The frost-point hygrometer with photoelectric deposit indication. A deposit on the top surface of the thimble scatters light which is received by the phototube. The thimble is at frost point when a small deposit is neither growing nor evaporating. (*Reproduced by courtesy of the Royal Society, London.*)

THE RESULTS OF AIRCRAFT ASCENTS

The measurements which have so far been made show that wide variations of relative humidity occur in the atmosphere, and any relative humidity between 30 and 90 per cent, with respect to ice, is about equally common. Relative humidities of about 5 per cent are relatively frequent, and the lowest which has yet been observed is 0.65 per cent. This humidity was observed at an air temperature of 266K and a frost point of 219K. It is characteristic of the frost-point hygrometer that even low humidities of this kind can be measured precisely. If the relative humidity were, say, 0.75 per cent, this would correspond to an increase in the frost point by one degree centigrade to 220K, which could be measured quite readily. A similar relative humidity, 0.65 per cent, has also been observed in the stratosphere with an air temperature of 225K and a frost point of 189K, but this could not be measured so precisely. Supersaturation with respect to ice is a relatively frequent occurrence, but supersaturation with respect to supercooled water has not been observed.

Comparison with Radiosonde Measurements. The comparison of radiosonde humidity measurements with humidity measurements made by a dew-point hygrometer is difficult because the air is often patchy and any differences may not be instrumental unless the ascents are made at the same time and place. Two ascents have been made by Brewer and Harrison [5] in which the aircraft from which the dew-point measurements were made circled the ascending radiosonde which had a goldbeater's-skin hygrometer. The two ascents are shown plotted as relative humidity against height in Fig. 3.

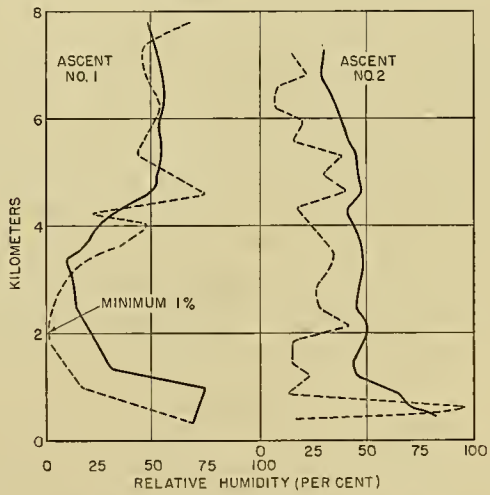


Fig. 3.—Comparison of humidities as measured by frost-point hygrometer (broken curve) and Kew pattern radiosonde (solid curve). The aircraft from which the frost-point hygrometer measurements were made circled the balloon, which was operated at reduced lift. (*Reproduced by courtesy of the Physical Society, London.*)

It will be seen that the radiosonde very greatly smooths the curves and in ascent No. 1 does not show the dry layer at 2 km where the lowest humidity is 1.0 per cent. Below 1 km the difference between the radiosonde and the aircraft hygrometer is due to the presence of broken cloud layers. The aircraft was flown through the dry air between the clouds, but a balloon tends to become entrained in clouds.

The Lower Stratosphere. In the ascents which have so far been made into the stratosphere (about seventy, all in southern England), the frost point generally falls with height in the region of the tropopause, but immediately above the tropopause there is an increase in the rate of fall, and one or two kilometres above the tropo-

pause the air normally has a frost point of about 190–200K and a humidity with respect to ice of 2 or 3 per cent. The effect is normally striking and consistent. An ascent is shown in Fig. 5, and a mean curve (due to Shellard [15]) of the results so far available (southern England) is shown in Fig. 4.

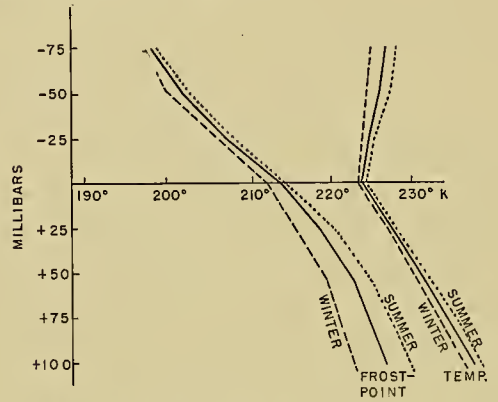


Fig. 4.—Mean variation of temperature and frost point with height above and below the tropopause, for southern England. All data so far available. (*Curve due to H. C. Shellard* [15].)

This effect had been predicted by Jaumotte [11], but other workers on the radiation balance of the lower stratosphere had assumed that the stratospheric air would be at or near saturation. The relative importance of water vapour in the radiation balance of the stratosphere is therefore much less than has usually been supposed. Dobson [6] has, for example, suggested that as a result of the small amount of water vapour present, the carbon dioxide and the ozone, which also take part in the radiation processes, may be equally important and that the radiative equilibrium may be determined by the relative amounts of these gases present.

If we consider the water vapour as a "tracer," it is difficult to give a satisfactory explanation of the presence and origin of the extremely dry air of the stratosphere. Since the dry air is consistently found with upper winds from all directions, it seems certain that the effect occurs in all temperate latitudes, and probably in polar regions also. Photolysis of the water vapour at these relatively low levels seems unlikely, for if it occurred, the reaction would be well known. If the photolysis occurs at high levels, it is difficult to see why the transition to dry air should be as low as 10 or 12 km.

The temperature at the equatorial tropopause, which is the lowest temperature found in the atmosphere, is about 190K and all the results so far obtained might be explained by assuming that the air is dried there, the excess water being condensed out at the low temperatures. Cwilong [6] has stated that at very low temperatures condensation occurs in the form of large ice crystals which would fall out easily.

Air which has been dried to a frost point of about 190K at the equatorial tropopause can then be brought to the polar or temperate stratosphere either (1) by ordinary advection parallel to the tropopause, or (2) by a zonal circulation in which air rises at the equator, enters the stratosphere, and moves to higher latitudes where it sinks into the troposphere. These processes may, of course, occur together in any proportion. Whatever the process, it must be considered in conjunction with the work of Glueckauf and Paneth [10], who found that up to 20 km the proportion of helium in the atmosphere was constant to within ½ per cent (the accuracy of their measurements). This suggests that turbulent mixing in the stratosphere is enough to maintain the composition of stratospheric air constant, at least up to 20 km, as otherwise the helium concentration would increase rapidly at high levels. The water-vapour results, on the other hand, suggest intense stratification.

It seems unlikely that the water-vapour distribution is maintained by the advection of dry air from the equator without the assistance of large-scale but slow zonal overturning, for it should be noted that the flow would have to be parallel to the tropopause, and there seems no obvious reason for this. Also, in the absence of large-scale vertical motions, turbulence is required to maintain the constancy of the helium content, and on these, and other grounds, the value of K (the diffusion coefficient) in the lower stratosphere would be expected to be at least 10^3. In this case, to maintain the shape of the observed vapour pressure versus height profile, it would be necessary for the air to be redried at the equator every twenty or thirty days. Since the potential temperature of the air at the equatorial tropopause is at least 20C higher than at the tropopause over England, it would be necessary for the air to be cooled by radiation at a rate of one or two degrees centigrade per day during its journey from the equator; this rate seems improbably high.

A mean zonal overturning in which air entered the stratosphere via the equatorial tropopause in a slow but roughly continuous motion would simultaneously maintain the dryness of the stratosphere and would prevent gravitational settling out of the helium. An extremely slow circulation would suffice to account for the observed distribution of helium, and the "water-vapour–height" curves would be maintained against the effects of diffusion ($K = 10^3$), with a circulation rate such that the mean time taken for the journey from the equator is about six months, and with a mean subsidence rate in the stratosphere over England of about 30 m per day. This would require radiative cooling of the air in temperate and polar latitudes of 0.3C per day. On the other hand, zonal overturning does not enjoy current favour and on dynamical grounds it is generally believed to be impossible, except in regions of zero or negative absolute vorticity. Thus the water-content observations raise the whole problem of the origin of the stratosphere, first because radiation conditions are drastically changed by reducing the available radiating water vapour, and second because it seems unlikely that the dryness of the air can be maintained without significant dynamic movements. If these dynamic movements occur, they will affect the temperature by adiabatic compression or expansion and the basic assumption of a stratosphere in radiative equilibrium may have to be abandoned. These points have been discussed by Brewer [2].

Results of a balloon ascent to 30 km, which included dew- or frost-point measurements, have been given by Barrett and others [1] and other successful ascents have since been made. They have observed an increase of humidity mixing ratio with height in the stratosphere in the summer and a decrease in the winter. A moist layer has also been observed, but only a single layer and not a double layer as reported in their note. This work will be watched with great interest.

On thermodynamic grounds, kinetic energy cannot be released in the atmosphere, except where low temperature is associated with low pressure. In the stratosphere, therefore, the complex dynamic movements which occur in the troposphere would not be expected as the energy for them is not available. The explanation of the shallow moist layer which Barrett found at 16 km will therefore be of special interest. The difference between summer and winter rather suggests that the low temperatures now known to occur in the stratosphere over the poles in winter may be a factor in maintaining the dryness of the stratosphere.

The Tropopause. It was first suggested by Gold that the air at and immediately below the tropopause would be substantially saturated. Observations now show that the actual humidity is highly variable. Its average value is about 50 per cent. The lowest relative humidity observed at the tropopause is 3 per cent. When high humidities, including supersaturation with respect to ice, occur, the temperature is usually low enough for dens- or persistent condensation trails to be formed by aire craft. A simple account of the relation between aircraft condensation trails and the humidity and temperature of the air has been given by Brewer [3]. When humidity observations are being made, the condensation trails formed by the aircraft should always be noted, as they are a most useful check on the observations.

Low relative humidities at or just below the tropopause could be due to a recent sinking movement of the tropopause, in which the air at the tropopause, and possibly in the stratosphere as well, moved downward. This process has often been suggested to explain the low tropopause heights which are found to the rear of depressions. The low humidities could also be caused by the incorporation of a substantial volume of stratospheric air into the upper troposphere.

It may be mentioned that the reverse process, the incorporation of moist tropospheric air into the stratosphere, must be quite rare over southern England. The stratosphere air is so very dry that a small admixture of air from the troposphere would raise its frost point markedly.

The Troposphere. In the troposphere the water-vapour distribution is extraordinarily complex. The atmosphere can be very moist, very dry, or very variable with intense stratification. The structure of the water-vapour distribution presumably arises from the complex dynamic movements which occur in the atmosphere, and these are too poorly understood at present to permit a clear explanation of any particular observed distribution. The purpose of this section will be to present a small selection of the ascents which have so far been made, to indicate the usefulness of dew-point observations as a technique for dynamic investigation, and to show how the dew-point curve shows features which are otherwise difficult to identify. At present only isolated ascents have been made, and no analysis on a synoptic basis is possible.

Subsiding Polar Air, May 3, 1944 (Figs. 5*a* and 5*b*). This ascent shows strongly subsiding polar air behind a vigorous depression. The surface synoptic situation is shown in Fig. 5*b*. The convective friction layer of moist air extends to about 820 mb, but at 800 mb the relative humidity is 5 per cent and cloud development is suppressed by the dryness of the air. The upper air is so dry that there is only about 0.2 mm of precipitable water in the whole atmosphere above 800 mb.

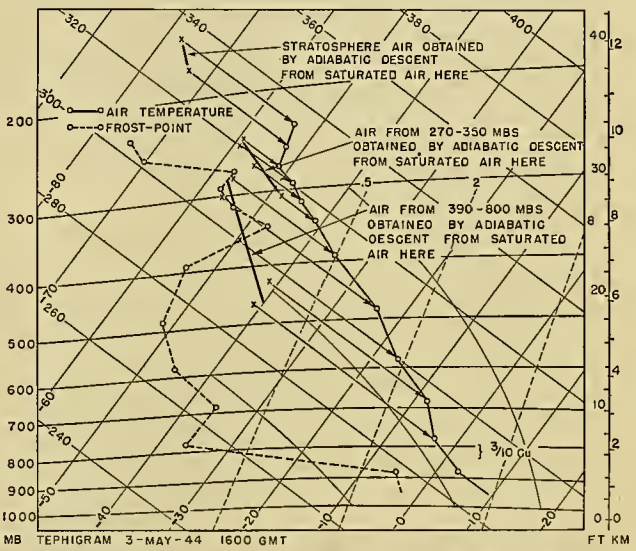


Fig. 5*a*.—Typical frost-point hygrometer ascent in strongly subsiding polar air (1600 GMT, May 3, 1944). Ascent made over northeast England, at approximately 54°N 0°W.

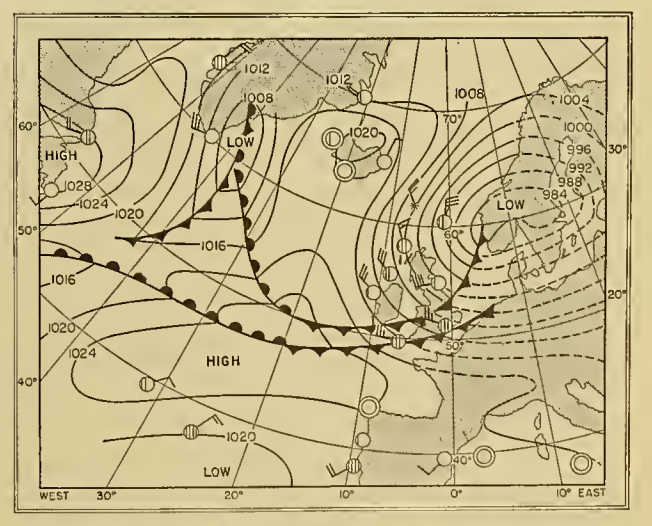


Fig. 5*b*.—Sea-level synoptic chart at about 0600 GMT, May 3, 1944.

Except for the shallow layer at 345 mb all the air above 600 mb has a lower humidity mixing ratio than would correspond to saturation at the tropopause, while the observations at 345, 700, and 800 mb show that at these levels the air is only slightly more humid.

If we assume that all the air was saturated and has moved continuously downward during any recent dynamic movements, during which entropy and humidity mixing ratio have been conserved, we may compute, for each observation, the level from which the air has subsided. This is shown in Fig. 5*a* in which the temperature-pressure levels from which the air has subsided are shown joined to the corresponding observed temperatures by isentropic lines. It will be seen that there is a sharp discontinuity in the nature of the source air mass between 345 and 400 mb, the air above 345 mb presumably being warm air and the air below 345 mb polar air, most of which has subsided strongly. It should be noted that the demonstration of this feature depends entirely upon the water-vapour measurements. The temperature-height curve does not indicate any significant change in the layer between 345 and 400 mb or in any other layer.

The thermal stability of the polar air mass from which the present air mass is derived is to be noted. This stability may be due to the cooling by radiation of the lower parts of the subsided polar air, and radiative adjustments of temperature may be the cause of the smoothness of the temperature-height curve. Alternatively, the air now found at the level of 600 to 800 mb may have originated from relatively farther north, so that the coldest air has been selected for the greatest subsidence. On thermodynamic grounds it is to be expected that the coldest air would subside most strongly. There is little doubt that all the processes of the atmosphere are not adiabatic, as may be seen from the observations in the stratosphere. The apparent source from which the stratospheric air could have been obtained (by adiabatic compression from saturated air) is shown in Fig. 5*a*, but this is a temperature-pressure relation which is not known to occur in the atmosphere.

The Water-Vapour Structure of an Anticyclone. During March 1945 an anticyclone persisted over Europe, its central position varying between the limits of southern England and southern France. Several ascents were made in the northern sector of this anticyclone. All the ascents showed a very sharp temperature inversion, the level of which varied from 900 to 800 mb, though it was most frequently found near 800 mb, higher than in most anticyclones. In the region of the inversion, and immediately above it, there was always a relatively dry layer, but this was usually only about 500 to 1000 m deep, with moister air above. The depth and dryness of this shallow layer varied considerably from day to day. One ascent (March 21, 1945) which showed this dry layer very strongly has already been published [6]. As further examples, two additional ascents (March 18 and March 20, 1945) are shown in Figs. 6 and 7 together with the corresponding sea-level and 500-mb charts. On March 18th the dry layer at the inversion

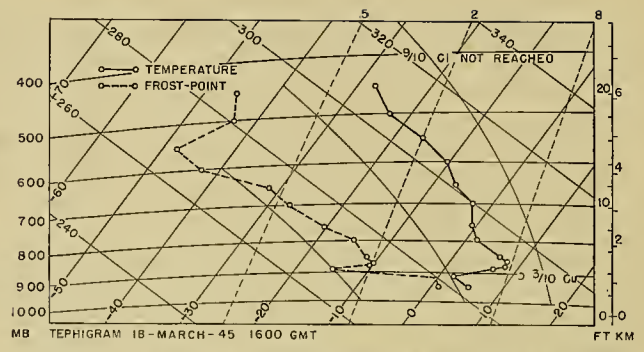


Fig. 6*a*.—Ascent in an anticyclone, March 18, 1945. Ascent made over southeastern England at approximately 51°N 2°W.

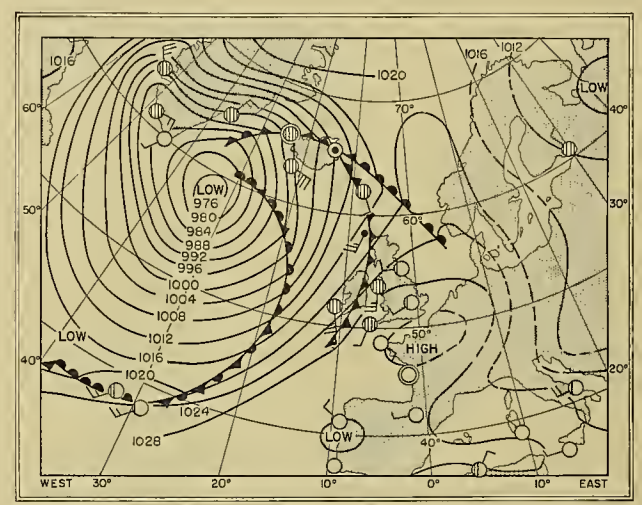


Fig. 6*b*.—Sea-level chart at about 0600 GMT, March 18, 1945.

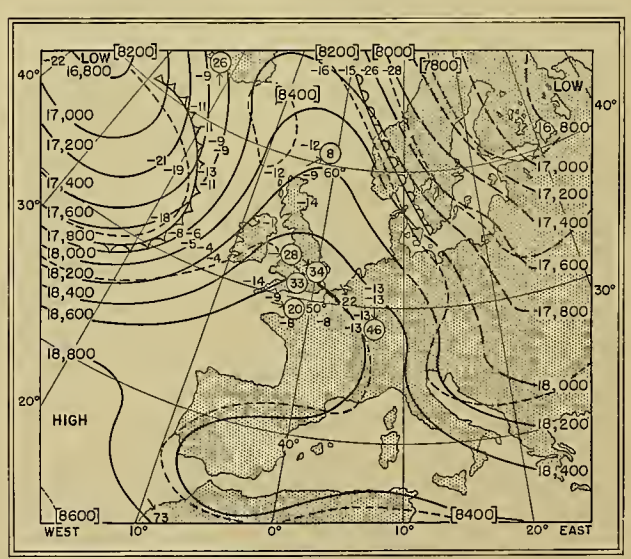


Fig. 6*c*.—Contour chart from 500 mb at 0600 GMT, March 18, 1945. Contour heights are given in feet and temperatures in degrees Fahrenheit. Figures in circles are wind speed in knots. Broken curves show isopleths of thickness of the layer 750–500 mb.

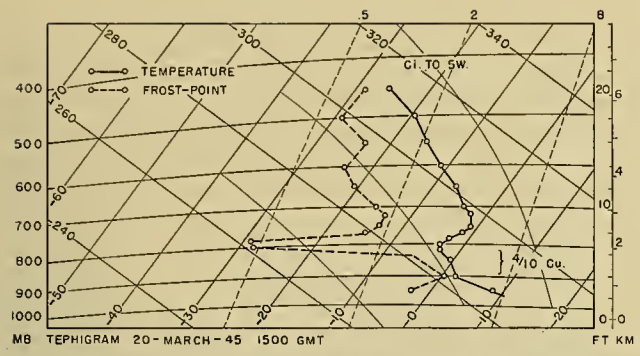


FIG. 7*a*.—Ascent in an anticyclone, March 20, 1945. Ascent made over southern England at approximately 51°N 2°W.

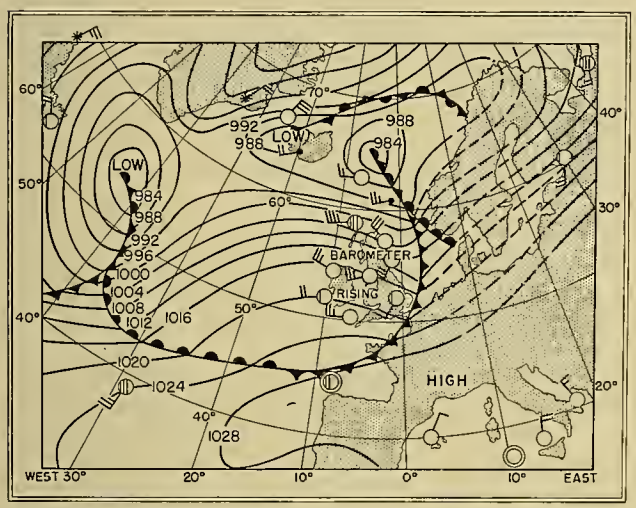


FIG. 7*b*.—Sea-level chart at about 0600 GMT, March 20, 1945.

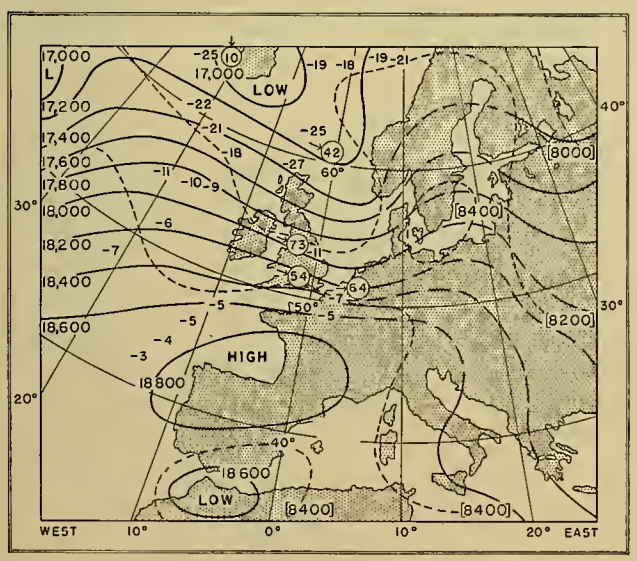


FIG. 7*c*.—Contour chart for 500 mb at about 0600 GMT, March 20, 1945. Contour heights are given in feet and temperature in degrees Fahrenheit. Figures in circles are wind speed in knots. Broken curves show isopleths of thickness of the layer 750–500 mb.

was shallow and weak and the air above was quite dry, particularly at 550 to 600 mb, where the frost-point depression was 32C. On March 20th, the dry layer at the inversion was well marked and it was accentuated by the moistness of the air above. If the surface synoptic charts are compared, the ascent of the 18th is seen to have been made in a weak warm sector with falling barometer, while the ascent of the 20th was made in a ridge. This would seem to contradict the relative dryness of the upper air on the 18th. On the 500-mb charts, however, the position is reversed. The dry air observed on the 18th between 700 and 500 mb is seen to be in an upper-air ridge, while the moister air of the 20th is in an upper-air trough.

Figures 10*a* and 10*b* give another example of an ascent in an anticyclone.

Intrusion of Moist Air Tongues into Dry Air. The ascent of June 7, 1943 (Fig. 8*a* with the surface synoptic chart Fig. 8*b*) was made in a ridge in advance of a weak frontal system which is shown on the charts as having a double structure. The nature of the frost-point

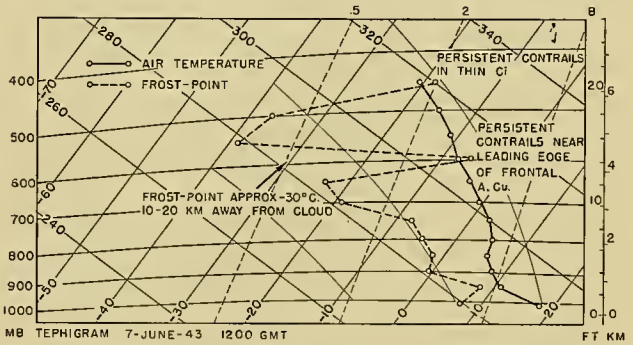


FIG. 8*a*.—Ascent showing moist air tongue intruding into dry air, 1200 GMT, June 7, 1943. Ascent made over southern England at approximately 51°N 2°W.

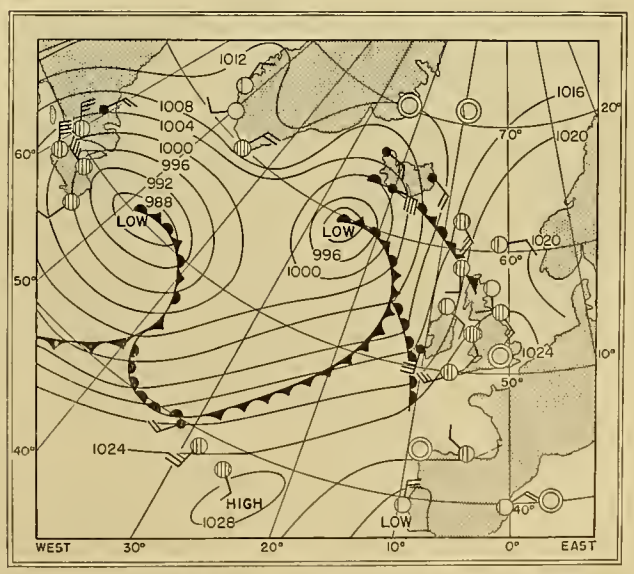


FIG. 8*b*.—Sea-level synoptic chart for approximately 0600 GMT, June 7, 1943.

curve is to be noted. At 600 mb, within one or two kilometres of the leading edge of a sheet of frontal altocumulus, the air was supersaturated with respect to ice, the observed supersaturation being confirmed by the formation of persistent condensation trails by the aircraft. At the same level, but 10 or 20 km from the cloud sheet, the air was dry with a frost-point depression of about 22C and a relative humidity of about 13 per cent, substantially on the smooth curve of the environment. The supersaturation with respect to ice at 450 mb is also to be noted.

Very Low Relative Humidities. Air of relative humidity as low as 1 per cent is occasionally found over southern England in ordinary subsiding polar maritime air, but the lowest relative humidity we have observed, 0.65 per cent, was at 800 mb on January 17, 1946, in air which was an outflow from a strong continental winter anticyclone (Figs. 9*a* and 9*b*). On that day the general structure of the overrunning moist air was shown by all the radiosonde ascents in southern England, though the details vary. The moist air is of Mediterranean origin.

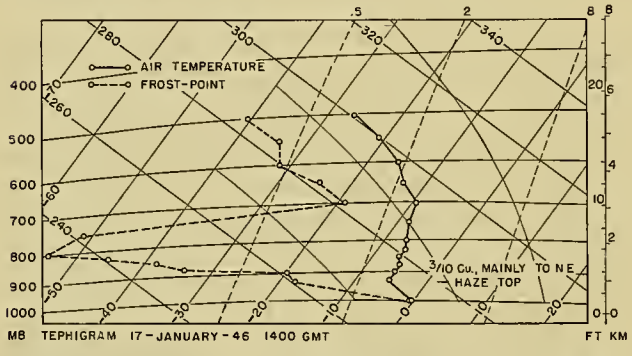


Fig. 9*a*.—Very dry air in a continental winter anticyclone, 1400 GMT, January 17, 1946. Ascent made over southern England at approximately 51°N 2°W.

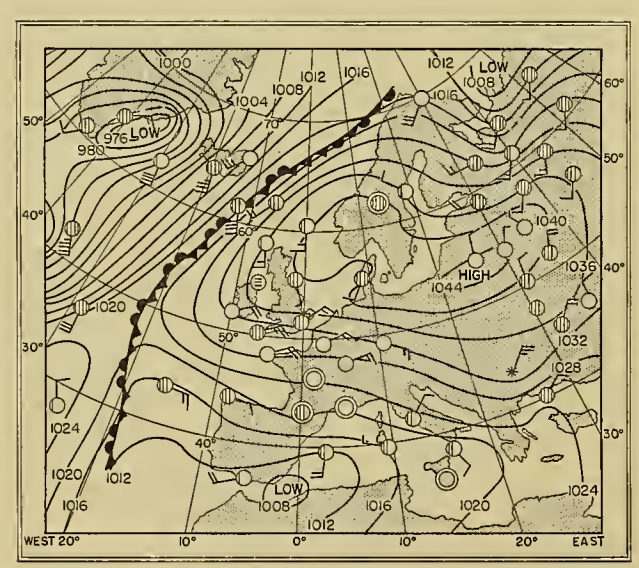


Fig. 9*b*.—Sea-level synoptic chart for approximately 0600 GMT, January 17, 1946.

The dry air could have originated from saturated air by adiabatic compression from a level of 350 mb at 212K. If water vapour had diffused into the layer then its origin would be higher and colder. It is possible that the air had originated from within the stratosphere and that the additional water vapour required, 18 mg m^{-3}, had diffused into it. It would then be necessary for the air to have cooled by radiation to its present potential temperature.

Small-Scale Temperature and Humidity Patterns in Free Air. Frith [7], using a frost-point hygrometer, has made careful investigations of the detailed temperature and humidity structure in the free air both in a horizontal and in a vertical plane. In a horizontal plane he finds evidence for the existence of closed temperature and humidity patterns on a scale of tens of miles and some similarity between the temperature and humidity isopleths; low frost points tend to be associated with high air temperature and vice versa. While the temperature variations are normally only of the order of one degree centigrade, the frost-point variations may be up to 15C.

Similarly, in the vertical plane, when strong strati-

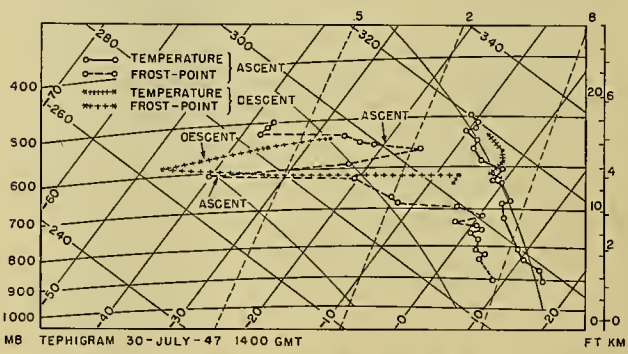


Fig. 10*a*.—Variation of observed conditions with position. Observations during ascent and descent showing changes, 1400 GMT, July 30, 1947. Ascent over southern England at approximately 51°N 1°W. (*By R. Frith.*)

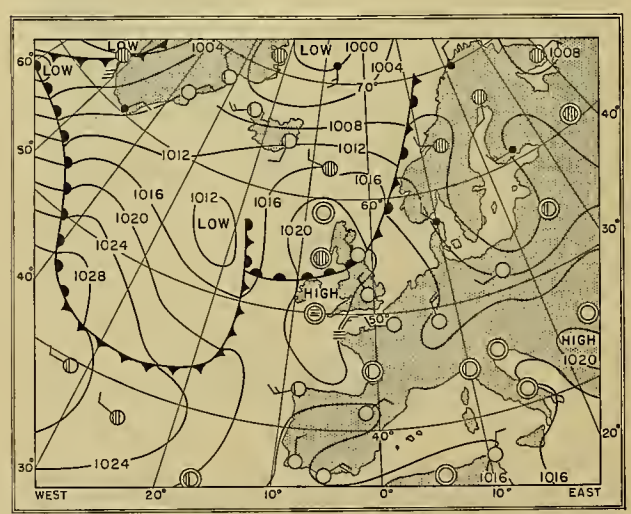


Fig. 10*b*.—Sea-level synoptic chart for approximately 0600 GMT, July 30, 1947.

fication of the water content is observed, the temperature variations from a smooth temperature-height curve are only about $\frac{1}{10}$ or $\frac{1}{20}$ of the variations shown in the frost-point–height curve. As in the horizontal plane, low frost points tend to be associated with relatively high air temperatures.

It is the writers' experience that the atmosphere is extremely variable in this respect; the structure is sometimes relatively simple but sometimes very complex. Figure 10, for July 30, 1947, shows measurements on both ascent and descent through a dry layer. As might be expected, the height and detailed structures change between ascent and descent, there being a difference in both the time and place.

FUTURE WORK

Only data collected by the Meteorological Research Flight and by Barrett and collaborators are at present available, and only a very few ascents with accurate humidity measurements have yet been published.

Experience in England has shown that almost every ascent is of great interest, and features are frequently observed which are difficult to explain. There is an urgent need for more observations in every part of the world in all types of weather and particularly on a synoptic basis, mainly to permit a systematic study of the widespread and fundamental phenomena of subsidence. The structure of anticyclones, both warm anticyclones and also winter continental anticyclones such as the Siberian High, will prove of interest. The higher tropopauses which occur in warm anticyclones have not yet been investigated in England because they are beyond the ceiling of the Mosquito aircraft used. The scope for the useful collection of observational data is believed to be world wide. Meteorologists will also have to learn to use and appreciate the data.

On the instrumental side the eye instrument is very satisfactory as a reliable instrument which will get results, but the observational skill required will inevitably restrict its use. A fully automatic hygrometer for widespread use is urgently needed, but the development of such an instrument is a difficult problem and will not easily be solved. It is also hoped that alternative methods, such as the spectroscopic method, or the method of Simon and Peckover, will be pursued.

We are greatly indebted to the Director of the British Meteorological Office for permission to reproduce unpublished data obtained by the Meteorological Research Flight of the Office. Figures 1 and 2 are reproduced from the *Proceedings of the Royal Society* by permission of the Council. Figure 3 is reproduced by permission of the Physical Society, London. The synoptic charts are reproduced from the appropriate British Daily Weather Reports by permission of the Controller of His Britannic Majesty's Stationery Office and the Director of the Meteorological Office. British Crown Copyright is reserved.

REFERENCES

1. Barrett, E. W., Herndon, L. R., Jr., and Carter, H. J., "A Preliminary Note on the Measurement of Water-Vapor Content in the Middle Stratosphere." *J. Meteor.*, 6: 367–368 (1949).
2. Brewer, A. W., "Evidence for a World Circulation Provided by the Measurements of Helium and Water Vapour Distribution in the Stratosphere." *Quart. J. R. meteor. Soc.*, 75: 351–363 (1949).
3. —— "Condensation Trails." *Weather*, 1:34–40 (1946).
4. —— Cwilong, B., and Dobson, G. M. B., "Measurement of Absolute Humidity in Extremely Dry Air." *Proc. phys. Soc. Lond.*, 60: 52–70 (1948).
5. Brewer, A. W., and Harrison, D. N., *Direct Comparison of Aircraft and Radiosonde Temperature and Humidity Measurements*. Meteor. Res. Com. Paper No. 205, Air Ministry, London, 1944.
6. Dobson, G. M. B., Brewer, A. W., and Cwilong, B., "Meteorology of the Lower Stratosphere." *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
7. Frith, R., *Small Scale Temperature and Humidity Patterns in the Free Air*. Meteor. Res. Com. Paper No. 402, Air Ministry, London, 1948.
8. Glueckauf, E., "Investigations on Absorption Hygrometers at Low Temperatures." *Proc. phys. Soc. Lond.*, 59: 344–365 (1947).
9. —— "Notes on Upper Air Hygrometry. Part II: On the Humidity in the Stratosphere." *Quart. J. R. meteor. Soc.*, 71: 110–114 (1945).
10. —— and Paneth, F. A., "The Helium Content of Atmospheric Air." *Proc. roy. Soc.*, (A) 185: 89–98 (1946).
11. Jaumotte, J., "Transformations thermodynamiques de la stratosphère et nuages nacrés." *Inst. R. météor. Belg., Mém.*, Vol. 5 (1936).
12. Normand, C., "Energy in the Atmosphere." *Quart. J. R. meteor. Soc.*, 72: 145–167 (1946). (See articles by same author in earlier papers)
13. Palmer, H. P., *The Wilson Cloud Chamber as a Hygrometer*. Meteor. Res. Com. Paper No. 499, Air Ministry, London, 1949.
14. Regener, E., "Aufbau und Zusammensetzung der Stratosphäre." *Dtsch. Akad. Luftfahrtforsch.*, SS. 7–48 (1940).
15. Shellard, H. C., *Humidity of the Lower Stratosphere—The Results of High Level Ascents by Aircraft of the Meteorological Research Flight*. Meteor. Res. Com. Paper No. 486, Air Ministry, London, 1949.
16. Simon, F. E., "A New Method of Sampling the Upper Atmosphere." *Nature*, 164: 179 (1949).
17. Störmer, C., "Mother-of-Pearl Clouds." *Weather*, 3: 13–18 (1948).

DIFFUSION IN THE UPPER ATMOSPHERE

By HEINZ LETTAU

Geophysics Research Division of the Air Force Cambridge Research Center

FUNDAMENTALS OF THE THEORY OF DIFFUSION

1. General Remarks. The kinetic theory of gases defines diffusion as the average motion of selected molecules relative to other molecules. The physical units of diffusion are number per square centimeter per second, that is, diffusion velocity times number of selected molecules per cubic centimeter. Diffusion depends on the composition of the gaseous mixture. A classical model considers two sets of molecules of different mass, effective diameter, and velocity distribution. The theory of diffusion results in complicated equations, even in the simple case of binary mixtures in a closed system when the physical state is well defined.

Pure, dry air is more complex than a binary mixture. Water (vapor, liquid, solid) and particulate matter (nuclei, smoke, dust) complicate the composition of the atmosphere. The atmosphere is not a closed system and the physical state beyond the scope of soundings is uncertain in many respects [35]. Sources[1] and sinks[2] of constituents must be considered. Large- and small-scale atmospheric motions produce eddy diffusion.

Diffusion equilibrium results when molecular and eddy diffusion balance the effects of forces acting on the molecules or the effects of sources and sinks so that the composition is a steady function of height or of height and the horizontal coordinates. Time variations of composition result from variations of the physical state—especially of motion and turbulence—and from intensity variations of sources and sinks of certain constituents.

Because of the scarcity of direct observations, the problems of diffusion in the upper atmosphere were developed theoretically and on the basis of conjecture. In this article, the author has tried to point out the inadequacies of the field by thoroughly outlining the assumptions necessary for a mathematical analysis of atmospheric diffusion. In general, applications of the theory must be confined to the lowest 200 km. Future work will be more promising when theoretical research can be supported by more and better observations from the stratosphere and ionosphere.

2. Molecular Diffusion. The study of nonuniform gases by Chapman and Cowling [4] resulted in the following general equation of diffusion velocity in a binary gas mixture:

$$\mathbf{c}_1 - \mathbf{c}_2 = -d_{12}\left\{\frac{\nabla\nu_1}{\nu_1\nu_2} - (\mu_1 - \mu_2)\nabla \ln p + \frac{m_1 m_2}{mkT}(\mathbf{F}_1 - \mathbf{F}_2)\right\} - d_T \frac{\nabla \ln T}{\nu_1 \nu_2}. \quad (1)$$

1. Gas and particle production of the lithosphere and hydrosphere, volcanic activity, photochemical reactions, industrial processes, etc.

2. Outflow of light gases into space, condensation, sedimentation, coagulation, recombination, etc.

The symbols in equation (1) are explained in Table I.

Table I. List of Symbols

The subscript s denotes one constituent of the mixture; for a binary mixture, $s = 1$ and 2. The rectangular coordinates x, y, and z are oriented so that z is positive towards the zenith. The unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are in the directions x, y, and z.

Constituent gases

$\mathbf{c}_s = c_{sx}\mathbf{i} + c_{sy}\mathbf{j} + c_{sz}\mathbf{k}$ = vector of mean molecular motion
n_s = number of molecules per cm³
m_s = mass of the molecule
p_s = hydrostatic partial pressure
$\mathbf{F}_s$ = external acceleration acting on the molecules
d_T = coefficient of thermal diffusion
d_{ss} = coefficient of self-diffusion

Gaseous mixture

$d_{12} = d_{21}$ = coefficient of mutual diffusion
$n = \Sigma n_s$ = total number of molecules per cm³ (Loschmidt's number) $= 2.705 \times 10^{19}$ at NTP
NTP = normal temperature and pressure
p = hydrostatic pressure; normal pressure = 1013 mb
T = absolute temperature; normal temperature = 273K
$\nu_s = n_s/n$ = number concentration
$m = \sum_s \nu_s m_s$ = mass of a fictitious average molecule
$\rho = nm$ = density of the mixture
$\mu_s = (m_s - m)/m$ = relative weight factor
$k = 1.372 \times 10^{-16}$ ergs per degree = Boltzmann's constant

Fundamental equations are Dalton's law:

$$p = \sum_s p_s , \quad (2)$$

and the equation of state:

$$p_s = n_s kT \quad \text{or} \quad p = nkT. \quad (3)$$

Equation (1) shows that the vector of diffusion velocity $\mathbf{c}_1 - \mathbf{c}_2$ has four constituents,

$$\mathbf{c}_1 - \mathbf{c}_2 = \mathbf{c}_d + \mathbf{c}_p + \mathbf{c}_F + \mathbf{c}_T ,$$

where $\mathbf{c}_d$ is ordinary diffusion velocity,
$\mathbf{c}_p$ is pressure diffusion velocity,
$\mathbf{c}_F$ is forced diffusion velocity,
$\mathbf{c}_T$ is thermal diffusion velocity.

Experiments are usually based on ordinary diffusion due to initial nonuniform composition in a closed system, $\nabla\nu_s \neq 0$, when $\mathbf{c}_p$, $\mathbf{c}_F$, and $\mathbf{c}_T$ are neglected. In the atmosphere, pressure diffusion (an indirect effect of gravity since the direct effect of gravitational acceleration on all molecules is the same) must be considered; $\mathbf{c}_p$ is due to the gradient of $\ln p$ set up by gravity and/or rotation in compressible media. The most important example of forced diffusion is the effect of electric fields on the motion of ionized gases. Thermal diffusion results when different parts of the mixture are maintained at different temperatures; however, the degree of separation produced by $\mathbf{c}_T$ is small; experience has proven that $d_T/d_{12} \approx 0.1$ and decreases when n_2/n_1

decreases. Thus, $\mathbf{c}_T$ can be disregarded in the atmosphere in view of the smallness of $\nabla \ln T$.

Table II shows experimental d_{12}-values when one gas diffuses into pure air. Observations of d_{12} are difficult and are liable to experimental errors; this fact explains the discrepancies between the findings by different authors.

TABLE II. OBSERVED DIFFUSION COEFFICIENTS
(According to Chapman and Cowling [4])

Gaseous Mixture	d_{12} at NTP (cm² sec⁻¹)
H_2—air	0.616
O_2—air	0.178
CO_2—air	0.138

Theoretically, d_{12} does not vary with the proportions of a binary mixture. There exist only approximate expressions for d_{12}, the degree of approximation being determined by the assumptions of the nature of molecular encounters. For atmospheric diffusion, it is sufficient to study the model of rigid, elastic, spherical molecules of effective diameter σ_s. Then, the kinetic theory results in

$$d_{12} = \frac{3}{4n(\sigma_1 + \sigma_2)^2}\sqrt{\frac{2kT(m_1 + m_2)}{\pi m_1 m_2}}\ \text{cm}^2\ \text{sec}^{-1}, \quad (4)$$

where $\pi = 3.14159\ldots$. Equation (4) fails when air is one constituent of the binary mixture under consideration, since air itself is a mixture. It is a favorable circumstance that nitrogen dominates in the atmosphere, and that molecular nitrogen, N_2, abounds approximately up to 200 km above sea level. For the sake of clarity, atmospheric diffusion is defined as the motion of relatively small numbers of different gas molecules in the medium of an N_2-atmosphere. Only then can the laws of mutual diffusion be applied to the atmosphere, that is, the diffusion of different gases can be studied separately.

Consequently, $\nu_s \approx n_s/n_{N_2}$ = concentration by volume, $\mu_s \approx (m_s - m_{N_2})/m_{N_2}$, and $d_{12} = d_{sN_2}$. The quantity ν_s is a small variable number, $0 < \nu_s \ll 1$; μ_s is a constant for each gas, and d_{sN_2} is defined by equation (4). Table III lists the molecular properties of N_2 and other gases of the atmosphere. The relatively small differences between corresponding values of Tables II and III justify the foregoing assumptions at NTP. However, they must become dubious at levels above approximately 200 km.

TABLE III. MOLECULAR PROPERTIES AND DIFFUSION COEFFICIENTS OF ATMOSPHERIC CONSTITUENTS*

Gas	m_s (relative)	μ_s	$\sigma_s \times 10^8$ cm	d_{sN_2} at NTP (cm² sec⁻¹)		$\delta_s = d_{sN_2}/d$
				Equation (4)	Observed	
H_2	2.015	−0.93	2.5	0.67	0.67	3.7
He	4.00	−0.86	2.0	0.59	—	3.3
N	14.0	−0.50	1.4	0.45	—	2.5
O	16.0	−0.43	1.2	0.47	—	2.6
H_2O	18.0	−0.36	3.5	0.21	0.20	1.2
Ne	20.2	−0.28	2.6	0.26	—	1.4
N_2	28.02	0.000	3.5	0.18	—	1
O_2	32.0	0.14	3.3	0.19	0.18	1.1
A	39.9	0.42	3.2	0.18	—	1.0
CO_2	44.0	0.57	4.0	0.14	0.14	0.8
O_3	48.0	0.71	4.0	0.14	—	0.8
Kr	82.9	1.96	3.4	0.15	—	0.8
Xe	130.2	3.61	3.8	0.13	—	0.7

* Mass of the molecules $= m_s \times 1.66 \times 10^{-24}$ g; σ_s and observed d_{sN_2} values were averaged from tables in [4] and [15]. It is assumed that $\sigma_{O_3} \approx \sigma_{CO_2}$. The coefficient of self-diffusion $d_{N_2N_2} = 0.18$ is taken as the atmospheric standard coefficient d in the ratio $\delta_s = d_{sN_2}/d$.

3. Eddy Diffusion. In consequence of the assumptions referred to in Section 2, the diffusion velocity of a gas in the atmosphere equals $\mathbf{c}_s - \mathbf{c}_{N_2}$ when diffusion itself $= n_s(\mathbf{c}_s - \mathbf{c}_{N_2}) = n_s\mathbf{c}_s - n_s\mathbf{c}_{N_2}$. In order to measure diffusion on an absolute scale, the term $n_s\mathbf{c}_{N_2} = \nu_s n_{N_2}\mathbf{c}_{N_2}$ or the mean motion of the medium of diffusion must be investigated. For convenience, we express $\mathbf{c}_{N_2}$ by the instantaneous wind vector $\mathbf{V}_t$ which is a function of space and time.

The motion of the atmosphere is turbulent; irregular, random movements are superposed on the regular or representative flow $\mathbf{V}$ such that

$$\mathbf{V}_t = \mathbf{V} + \dot{\mathbf{V}}. \quad (5)$$

The vector $\mathbf{V}$ is an average with respect to time as denoted by the bar:

$$\mathbf{V} = \overline{\mathbf{V}_t}, \qquad \overline{\dot{\mathbf{V}}} = 0. \quad (6)$$

When $\nabla\nu_s \neq 0$ in an eddying medium, turbulent fluctuations of ν_s are created such that

$$(\nu_s)_t = \nu_s + \dot{\nu}_s\ ; \quad \overline{(\nu_s)_t} = \nu_s\ ; \quad \overline{\dot{\nu}_s} = 0. \quad (7)$$

If turbulent fluctuations of n_{N_2} are neglected,

$$\overline{n_s\mathbf{c}_{N_2}} = n_{N_2}\overline{\nu_s\mathbf{c}_{N_2}} = n_{N_2}(\nu_s\mathbf{V} + \overline{\dot{\nu}_s\dot{\mathbf{V}}}) = n_s\mathbf{V} + \frac{n_s}{\nu_s}\overline{\dot{\nu}_s\dot{\mathbf{V}}}. \quad (8)$$

Let the eddy diffusion-velocity be defined by

$$\mathbf{C}_s = \frac{1}{\nu_s}\overline{\dot{\nu}_s\dot{\mathbf{V}}}. \quad (9)$$

In order to separate the properties of the wind from the properties of the ν_s-distribution, a radius vector $\dot{\mathbf{l}}$ is introduced which is assumed to be independent of the special ν_s, that is,

$$\dot{\nu}_s = -\dot{\mathbf{l}}\cdot\nabla\nu_s\ , \quad (10)$$

Then,

$$\overline{\dot{\nu}_s\dot{\mathbf{V}}} = -\overline{\dot{\mathbf{V}}\cdot\dot{\mathbf{l}}}\nabla\nu_s, \text{ or } \mathbf{C}_s = -\frac{1}{\nu_s}\overline{\dot{\mathbf{V}}\cdot\dot{\mathbf{l}}}\nabla\nu_s, \quad (11)$$

where $\overline{\dot{\mathbf{V}}\cdot\dot{\mathbf{l}}}$ is the coefficient of eddy diffusion. Its components are

$$\overline{(\dot{\mathbf{V}}\cdot\mathbf{i})\,(\dot{\mathbf{l}}\cdot\mathbf{i})} \equiv D_x\ , \quad (12)$$

$$\overline{(\dot{\mathbf{V}}\cdot\mathbf{j})\,(\dot{\mathbf{l}}\cdot\mathbf{j})} \equiv D_y\ , \quad (13)$$

$$\overline{(\dot{\mathbf{V}}\cdot\mathbf{k})\,(\dot{\mathbf{l}}\cdot\mathbf{k})} \equiv D_z \equiv D. \quad (14)$$

The units of D_x, D_y, and D are cm² sec⁻¹ = length times velocity, the same as d_{sN_2} and d.

Let λ and ζ be lengths and velocities such that the product $\lambda\zeta$ equals a given diffusion coefficient. In molec-

ular diffusion, λ_d and ζ_d correspond to the mean free path and the average velocity of molecular motions. In eddy diffusion, λ_D is a length fixing the scale of turbulent displacements (mixing length), and ζ_D is a velocity fixing the speed of turbulent motions.

The differences between molecular and eddy diffusion are more obvious in terms of λ and ζ than in terms of d and D. The quantity ζ_d^2 corresponds to the internal energy or heat of a gas, while the energy for maintaining ζ_D^2 will be taken from external sources, such as the potential energy of the horizontal pressure distribution. The term λ_d is a unique function of the density of the atmosphere when λ_D depends on geometrical conditions like distances from bounding surfaces; the vertical mixing length is also affected by the vertical thermal stratification.

The value of D can equal that of d when $\lambda_d/\lambda_D = \zeta_D/\zeta_d$; however, this ratio is never unity. Differences between the mechanisms of diffusion are emphasized when time terms (λ/ζ) and acceleration terms (ζ^2/λ) are studied. Reasonable conditions—such as $\lambda_{d,D} < z$, $\lambda_D/\zeta_D < 1$ day, and $\zeta_D^2/\lambda_D < g$ when g = gravity—narrow a priori the variability of D-values. The diffusion diagram, Fig. 1, summarizes our present knowledge of

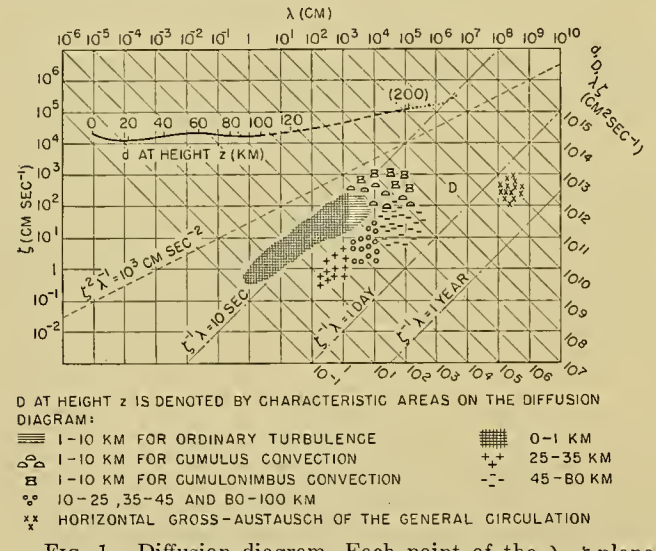


FIG. 1.—Diffusion diagram. Each point of the λ, ζ plane determines a diffusion coefficient (cm² sec⁻¹). In molecular diffusion, $\lambda \approx$ free path and $\zeta \approx$ mean molecular speed; $d = \lambda\zeta$ is fixed by the density and temperature of the atmosphere; consequently, the height variation of d is marked by a curve. In eddy diffusion, $\lambda \approx$ mixing length and $\zeta \approx$ mixing velocity; owing to the variability of these elements, $D = \lambda\zeta$ and its variation with height are denoted by characteristic areas when the possible variability of D is narrowed by the consideration of limiting values of eddy accelerations (ζ^2/λ) and time terms (λ/ζ).

molecular and vertical eddy-diffusion coefficients in the atmosphere. Values of D above 15 km and values of d above 100 km are not very reliable. Inasmuch as temperature and density are known, d_{sN_2} is a unique function of height. However, D will vary with height, weather, season, and latitude. Inasmuch as D is a statistical parameter, it will depend on the averaging process when λ_D/ζ_D is a measure of the minimum time interval required for the definition of representative values of wind and concentration ν_s .

4. The Equation of Atmospheric Diffusion. Let us define an atmospheric gas as "permanent" when there are no sources or sinks, as "neutral" when gravity is the only force acting on the molecules, and "at rest" when the vertical pressure distribution follows from the hydrostatic equation, $\partial p_s/\partial z = -n_s m_s g$. Let molecular nitrogen, as the medium of atmospheric diffusion, be a permanent, neutral gas at rest. Then, in equation (1), $\mathbf{F}_{N_2} = -g\mathbf{k}$ and $\mathbf{c}_{N_2}\cdot\mathbf{k} \equiv \mathbf{k}\cdot\mathbf{V} = 0$. However, $\mathbf{k}\cdot\tilde{\mathbf{V}} \neq 0$.

Other gases may be nonpermanent and ionized. The most important assumption is $\nu_s \ll 1$. Then, the density distribution in the N_2-atmosphere is not affected by the processes of diffusion and turbulence: $\nu_{N_2} \approx 1$, $n_{N_2} \approx n$, $m_{N_2} \approx m$, $\mu_{N_2} \approx 0$. Since diffusion equals diffusion-velocity times the number of diffusing molecules per cubic centimeter, the equation of atmospheric diffusion follows from (1), (8), and (9),

$$n_s\mathbf{c}_s = n_s(\mathbf{c}_d + \mathbf{c}_p + \mathbf{c}_F + \mathbf{C}_s + \mathbf{V}) \equiv \alpha\mathbf{i} + \beta\mathbf{j} + \gamma\mathbf{k}, \tag{15}$$

when $\mathbf{c}_T$ is neglected and

$$\mathbf{c}_d = -d_{sN_2}(\nabla\nu_s)/\nu_s\,, \tag{16}$$

$$\mathbf{c}_p = d_{sN_2}\mu_s\nabla \ln p, \tag{17}$$

$$\mathbf{c}_F = -d_{sN_2}m_s(\mathbf{F}_s - g\mathbf{k})/kT \equiv -d_{sN_2}m_s\mathbf{f}_s/kT, \tag{18}$$

$$\mathbf{C}_s = -\tilde{\mathbf{V}}\cdot\tilde{\mathbf{l}}(\nabla\nu_s)/\nu_s\,. \tag{19}$$

Let us define as an auxiliary parameter

$$Q_s = \frac{d_{sN_2}}{(d_{sN_2} + \tilde{\mathbf{V}}\cdot\tilde{\mathbf{l}})}. \tag{20}$$

With the aid of (15)–(20), the fundamental equation of atmospheric diffusion becomes

$$n_s\mathbf{c}_s = -nd_{sN_2}\left\{\frac{\nabla\nu_s}{Q_s} - \nu_s\left[\mu_s\nabla(\ln p) - \frac{m\mathbf{f}_s}{kT}\right]\right\} + n\nu_s\mathbf{V}. \tag{21}$$

The equation of continuity is

$$\frac{\partial n_s}{\partial t} = -\nabla\cdot(n_s\mathbf{C}_s) \equiv -\left(\frac{\partial\alpha}{\partial x} + \frac{\partial\beta}{\partial y} + \frac{\partial\gamma}{\partial z}\right). \tag{22}$$

Steady states or diffusion equilibriums are defined by $\partial n_s/\partial t = 0$, unsteady states by $\partial n_s/\partial t \neq 0$.

In the upper atmosphere, diffusion acts mainly in the vertical. Under the assumption that $n_s\mathbf{c}_s = \gamma\mathbf{k}$, $\partial\nu_s/\partial x = \partial\nu_s/\partial y = 0$, $\mathbf{f}_s = f_s\mathbf{k}$, $\partial \ln p/\partial z = -mg/kT$, and $\mathbf{k}\cdot\mathbf{V} = 0$, when $\tilde{\mathbf{V}}\cdot\mathbf{k} \neq 0$, equation (21) yields a differential equation for ν_s where height is the independent variable and the coefficients of the equation are fixed by molecu-

lar and eddy diffusion and the forces acting on the s-molecules:

$$\frac{\partial v_s}{\partial z} - \frac{v_s Q_s(\mu_s g m + m_s f_s)}{kT} + \frac{Q_s \gamma}{n d_{sN_2}} = 0, \quad (23)$$

when

$$Q_s = \frac{d_{sN_2}}{d_{sN_2} + D}, \quad (24)$$

and equation (22) is reduced to

$$\frac{\partial n_s}{\partial t} = -\frac{\partial (n_s \mathbf{c}_s \cdot \mathbf{k})}{\partial z} \equiv -\frac{\partial \gamma}{\partial z}. \quad (25)$$

Permanent gases are characterized by $\gamma_0 = 0$ in both steady and unsteady states; unsteady states are caused by time variations of T, Q_s, and/or f_s. Sources and sinks of nonpermanent gases can be *external* when matter enters one and leaves the other boundary of the atmosphere (or of the layer under consideration), and *internal* when production and destruction of matter (as measured by $q^{(+)}$ and $q^{(-)}$) are height functions in the layer under consideration. In steady states of nonpermanent gases with external sources and sinks,

$$\gamma = \gamma_0 = \text{const} \neq 0, \quad (26)$$

where γ_0 measures the strength of the continuous external source which equals the negative strength of the corresponding sink. In steady states of nonpermanent gases with internal sources and sinks, which are independent of x and y,

$$-\frac{\partial \gamma}{\partial z} + q^{(+)} + q^{(-)} = 0, \quad (27)$$

or diffusion balances the effects of internal production and destruction. Unsteady states of nonpermanent gases result, mainly, from time variations of γ_0, $q^{(+)}$, and $q^{(-)}$. Complications arise when the production of matter influences the physical state of the atmosphere (temperature, motion, turbulence), and when the ejection from sources causes an initial movement of matter, as in case of eruptions or explosions. Further details will be discussed in Section 12.

Atmospheric diffusion can also be studied in connection with the motion and distribution of particles (dust, nuclei, droplets, etc.). Let a_i be the diameter and n_i the number of particles per cubic centimeter. The kinetic theory represented by equation (4) fails in defining a coefficient of particle diffusion when a_i is large in comparison with $\sigma_{N_2} \approx 10^{-8}$ cm. There will be a Brownian movement of such particles, which becomes less intense as the size of the particles increases. When a_i is large in comparison with the free path of nitrogen molecules (10^{-5} cm at NTP), the integrated effect of the continuous action of numerous collisions with the molecules of the atmosphere is called the viscosity. If gravity is the only force acting on the particles, the motion reduces to a simple fall in a hypothetical atmosphere without wind and turbulence. For instance, spherical particles with $10^{-5} < a_i < 10^{-2}$ cm attain a rate of descent

$$c_i = -g a_i^2 (\rho_i - \rho)/9\mu, \quad (28)$$

according to Stokes' law, where ρ_i = density of the particle and μ = dynamic viscosity of the atmosphere = 0.00017 g cm^{-1} sec^{-1} at T = 273K.

The equation of particle diffusion in a turbulent atmosphere becomes

$$n_i \mathbf{c}_i = -n_i c_i \mathbf{k} + n_i \mathbf{C}_i + n_i \mathbf{V}, \quad (29)$$

where $\mathbf{C}_i$ is defined by (11) when $v_s = v_i = n_i/n_{N_2} \equiv n_i/n$.

DIFFUSION IN THE VERTICAL—STEADY CONDITIONS

5. The General Solution. One of the possible causes of forced diffusion is electromagnetic fields. The heat transfer in gases—which is very similar to the process of diffusion—is perceptibly affected at NTP by electric fields greater than 10^4 v cm^{-1}, as was shown experimentally by Senftleben and Gladisch [34]; the effect becomes less intense when pressure decreases. The magnitude of electric fields in the upper atmosphere normally is smaller than the value given above when ionospheric layers are electrically neutral. The possible influence of the earth's magnetic field on the diffusion of ions was found to be dubious by Ferraro [10]. Thus, we assume $f_s = 0$ in the following computations, based on equation (23).

The linear differential equation

$$\frac{dy}{dx} + y\phi(x) + \psi(x) = 0 \quad (30)$$

has the solution

$$y = \left[y_0 - \int \psi \exp\left(\int \phi \, dx\right) dx\right] \exp\left(-\int \phi \, dx\right), \quad (31)$$

when y_0 = const.

We define the scale height (height of the homogeneous atmosphere) as

$$H = kT_0/mg_0. \quad (32)$$

When N_2 is the medium of atmospheric diffusion, $m = m_{N_2} = 28.02 \times 1.66 \times 10^{-24}$ g. Normally, the scale height is expressed in terms of m_0, the average molecular weight of pure dry air ($m_0/m_{N_2} = 1.033$). In the following computations, the value of H = 8 km is used.

When $\gamma = \gamma_0$, that is, when internal sources and sinks do not exist between 0 and z, the solution of (23) with the aid of (31) and (32) is

$$v_s = v_{s0} \exp\left(-\frac{\mu_s}{H}\int_0^z \frac{gT_0}{g_0 T} Q_s \, dz\right) \cdot \left[1 - \frac{\gamma_0}{v_{s0}} \int_0^z \frac{Q_s}{n \, d_{sN_2}} \exp\left(\frac{\mu_s}{H}\int_0^z \frac{gT_0}{g_0 T} Q_s \, dz\right) dz\right]. \quad (33)$$

If $T = T_0$ and $g = g_0$ throughout the layer under consideration,[3] then ϕ/ψ = const in equation (30), and

3. Some investigators may prefer to omit the assumption $g = g_0$ by measuring the vertical scale in dynamic meters rather than in geometric meters.

equation (33) becomes

$$\nu_s = \nu_{s0}\left\{\exp\left(-\frac{\mu_s}{H}\int_0^z Q_s\,dz\right) - \frac{H\gamma_0}{\nu_{s0}\mu_s n\, d_{sN_2}} \cdot\left[1-\exp\left(-\frac{\mu_s}{H}\int_0^z Q_s\,dz\right)\right]\right\}. \tag{34}$$

In steady states, the height variation of the number concentration of a gas without internal sources in a nitrogen atmosphere depends on the ground concentration ν_{s0}, the relative weight factor μ_s, the scale height H, the molecular diffusion coefficient d_{sN_2}, the eddy diffusion (in the ratio Q_s), and the intensity of continuous external sources and sinks γ_0. Height variations of temperature and gravity are of secondary importance. The general solutions given above can readily be specialized, as was shown by Lettau [21].

6. The Border Cases of Maximum Mixture and Maximum Separation. A permanent gas constituent is defined by $\gamma_0 = 0$. Then, two extreme cases of equation (33) exist:

(*i*) The state of maximum mixture,

$$Q_s = 0 \quad \text{or} \quad D = \infty; \quad \nu_s = \nu_{s0} = \text{const.} \tag{35}$$

(*ii*) The state of maximum separation or the Dalton atmosphere,

$$Q_s = 1 \quad \text{or} \quad D = 0;$$
$$\nu_s = \nu_{s0}\exp\left(-\frac{\mu_s}{H}\int_0^z \frac{gT_0}{g_0T}\,dz\right), \tag{36}$$

$$\nu_s = \nu_{s0}\exp\left(-\mu_s z/H\right), \quad \text{if} \quad T = T_0 \quad \text{and} \quad g = g_0. \tag{37}$$

Different investigators have described these cases by different terms. For instance, case (*i*) was called "turbulent mixture" by Bartels [2], "state of convection" by Götz [12], and "adiabatic equilibrium" by Mitra [26]; case (*ii*) was called "gravity equilibrium" by Maris [24], "separation" by Penndorf [28], "atmosphere at rest" by Regener [31], and "isothermal equilibrium" by Mitra [26].

Normally, case (*i*) was attributed to the lower atmosphere. Chapman and Milne [5] introduced the hypothetical "datum level" as the height below which (*i*) and above which (*ii*) should be verified. Götz [12] and Haurwitz [17] remarked that eddy mixing can hardly stop completely at a certain level; there should be a gradual change from pronounced to slight mixing in the vertical direction. Lettau [21] pointed out that molecular diffusion acts everywhere when the effects of eddy diffusion vary with elevation, being nowhere infinite; in reality $0 < D < \infty$ or $0 < Q_s < 1$; (*cf.* equation (24)).

Consequently, Q_s is termed the separation factor. In reality, Q_s and $\partial Q_s/\partial z$ will vary with elevation. In the border cases (*i*) and (*ii*) $\partial Q_s/\partial z = 0$; then, the vertical pressure distribution can readily be found. In case (*i*)

$$P = P_0\exp\left(-\frac{1}{H}\int_0^z \frac{gT_0}{g_0T}\,dz\right) \tag{38}$$

or, if $T = T_0$ and $g = g_0$

$$P = P_0\exp\left(-z/H\right). \tag{39}$$

If $\partial\nu_s/\partial z = 0$, the vertical diffusion velocity as defined by equations (15)–(19) becomes

$$c_s = \mathbf{c}_s\cdot\mathbf{k} = \mu_s\, d_{sN_2}\frac{\partial \ln P}{\partial z} \tag{40}$$

or, if $T = T_0$ and $g = g_0$,

$$c_s = -\frac{\mu_s}{H}\frac{n\, d_{sN_2}}{n_0}e^{z/H}, \tag{41}$$

when C_s is undetermined.

Since $nd_{sN_2}/n_0 = (d_{sN_2})_0 = \text{const}$, c_s is inversely proportional to P/P_0. Values are given in Table IV in which the height variation of c_s is expressed by the varying velocity units. These motions must be thought of as compensated by an unlimited degree of vertical turbulence such that $\nu_s = \nu_{s0}$.

TABLE IV. DIFFUSION VELOCITY* IN THE HYPOTHETICAL STATE OF MAXIMUM MIXTURE (H = 8 km, $g = g_0$, $T = T_0$)

z (km)	H_2	He	O	O_2	A	Velocity units
0	25	20	8	−1	−3	cm yr^{-1}
40	99	81	32	−4	−12	cm day^{-1}
80	10	8	3	−0.4	−1.2	cm min^{-1}
120	25	20	8	−1	−3	cm sec^{-1}
200	6	5	2	−0.2	−0.7	km sec^{-1}

* Positive when directed upward

The magnitude of c_s in Table IV shows that, at and above 200 km, the state of maximum mixture will not be realized.

In a Dalton atmosphere, that is, case (*ii*), the pressure distribution is readily expressed by P as defined by (40) and (41):

$$p_s = p_{s0}\,(P/P_0)^{1+\mu_s}, \tag{42}$$

$$p = \sum_s p_s = P\sum_s \nu_{s0}\,(P/P_0)^{\mu_s}. \tag{43}$$

The number concentration of the light constituents ($\mu_s < 0$) must increase, the heavy constituents ($\mu_s > 0$) must decrease with elevation; inasmuch as $m = m_{N_2}$,

$$\nu_s = \frac{\nu_{s0}(P/P_0)^{\mu_s}}{\sum_s \nu_{s0}(P/P_0)^{\mu_s}}. \tag{44}$$

Hann [16] first verified this concept in 1903. Since then, computations of the composition of the upper atmosphere have been carried out repeatedly. Obviously, (44) is deficient with regard to eddy diffusion and the case where γ_0 or $\partial\gamma/\partial z$ differs from zero.

7. Partial Separation. If $\gamma_0 = 0$, $0 < D < \infty$,

equations (33) and (34) yield the equations of "partial separation" for permanent constituents:

$$\nu_s = \nu_{s0} \exp\left(-\frac{\mu_s}{H}\int_0^z \frac{gT_0}{g_0 T} Q_s\, dz\right) \tag{45}$$

or, if $T = T_0$ and $g = g_0$,

$$\nu_s = \nu_{s0} \exp\left(-\frac{\mu_s}{H}\int_0^z Q_s\, dz\right). \tag{46}$$

For convenience, a general separation factor Q, which is independent of the special gas, is defined as

$$Q = \frac{d}{d + D}, \tag{47}$$

in which $d = 0.18(p_0/p)(T/T_0)^{\frac{3}{2}}$ is the standard coefficient of the N_2-atmosphere (see Table III). We may also write (*cf.* equation (24))

$$Q = \frac{Q_s}{\delta_s + Q_s(1 - \delta_s)}, \quad \text{and} \quad Q_s = \frac{Q\delta_s}{1 - Q(1 - \delta_s)}. \tag{48}$$

The factor Q depends on pressure, temperature, and turbulence as functions of height. With reference to [38], $0.78 < T/T_0 < 1.25$ for $0 \leqq z \leqq 100$ km; the average T/T_0 is rather close to unity. If $T/T_0 = 1$, the logarithm of d increases proportionally to z. Crude values of D were estimated from the intensity of the atmospheric circulation and vertical thermal stratification for $0 \leqq z \leqq 100$ km (see Fig. 2). Equation (46) was applied for molecular oxygen, argon, and helium. The theoretical ν_s-variation, as shown in Fig. 2, reflects the Q-variation with height.

Inasmuch as $Q \neq 0$, the tendency towards separation begins at sea level. However, when $Q \leqq 10^{-4}$ and $(z_2 - z_1) \leqq 10$ km, all permanent constituent gases show that $|\nu_{s1} - \nu_{s2}|/\nu_{s1} \leqq 0.01$ per cent. Therefore, the concentration of permanent gases is practically constant throughout the entire troposphere where $Q \approx 10^{-5}$, as follows from Fig. 1.

Regener [30] showed that ν_{O_2} decreases slightly in the stratosphere from 15 to 29 km. Lettau [21] and Dütsch [8] used this fact for computing $A \equiv \rho D$ between 15 and 30 km. Because of the term $\delta_s\mu_s/[1 - Q(1 - \delta_s)]$, the percentage separation of helium should be approximately 20 times that of molecular oxygen. However, helium is not a permanent constituent. The observed helium distribution will be explained in Section 8. When no helium source and sink would exist, the height variation of ν_{He} should follow the dotted line in Fig. 2.

The coefficient d increases monotonically with height; increasing temperature can only slightly modify this trend. Regions above 100 km are subject to turbulence caused by diurnal heating and cooling and by tidal effects. However, in contrast to d, the value of D is limited. An estimated maximum D-value is approximately 10^8 cm^2 sec^{-1} since ζ_D and λ_D are unlikely to exceed 10^3–10^4 cm sec^{-1} and 10^5–10^4 cm, respectively (see Fig. 1). Thus

$$\lim_{z\to\infty} Q = 1. \tag{49}$$

When $D \leqq 10^8$ cm^2 sec^{-1} and (39) is used for computing d,

$$1 > Q > 0.5 \quad \text{if} \quad z \geqq 160 \text{ km}. \tag{50}$$

$$1 > Q > 0.99 \quad \text{if} \quad z \geqq 200 \text{ km} \tag{51}$$

If $Q \approx 0.99$, there should result, practically, a Dalton atmosphere. From this point of view, the datum levels

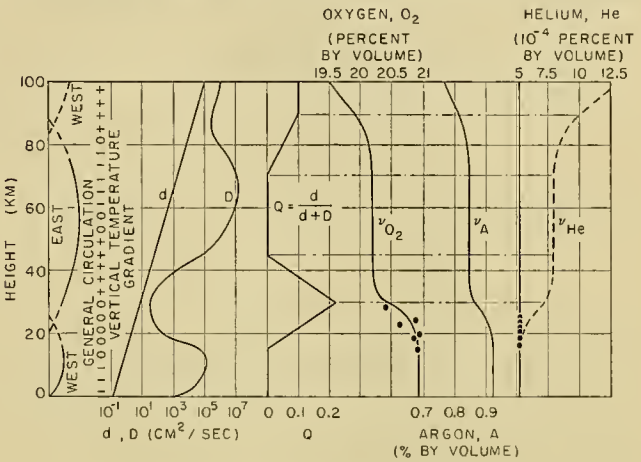


FIG. 2.—The influence of the separation factor Q on the height variation of concentration of heavy and light permanent gases. Regener's and Paneth's observations of ν_{O_2} and ν_{He} (highest levels of analyses are 29 and 25 km, respectively) are marked by dots. Helium is not a permanent gas and therefore ν_{He} = const when condition (52) is satisfied. The absolute values and the height variations of the eddy diffusion coefficient D will be valid for temperate latitudes since they were estimated from assumptions with regard to the representative zonal circulation in temperate latitudes and the sign of $\partial T/\partial z$ (−, 0, and + denote lapse, isothermal, and inversional stratification, respectively) such that strong zonal motion and/or lapse conditions cause large D-values when weak zonal motion and/or inversional stratification cause small D-values. Other facts giving support to the existence of layers of $\partial D/\partial z < 0$ above approximately 15 and 70 km are discussed in Sections 9 and 14.

of 12–50, 100, and 150 km as assumed by Chapman and Milne [5], Mitra [26], and Maris [24], respectively, appear to be too low, except possibly for the last one. Internal sources and sinks due to photochemical processes—which also affect nitrogen above 200 km—impair the computation of diffusion effects in the ionosphere, even though D might be neglected in comparison with d.

8. Effects of Sources and Sinks. If $\gamma = \gamma_0 \neq 0$ and $0 \leqq Q \leqq 1$, equation (34) describes steady height variations of nonpermanent gases with external sources and sinks. A basic flow enters one boundary of the layer under consideration and leaves the other. A highly interesting special case exists when

$$\gamma_0 = \gamma_{\text{crit}} \equiv -\nu_{s0}\mu_s n d_{sN_2}/H \equiv -n_{s0}\mu_s d_{sN_2}/H \equiv -n_{s0}(c_p)_0\,. \tag{52}$$

Then (34) reduces to

$$\nu_s = \nu_{s0} = \text{const.} \tag{53}$$

The concentration is constant when γ_0/n_{s0} equals the negative velocity of pressure diffusion at sea level (see equations (17) and (41)). Petersen [29] discussed this

result for a Dalton atmosphere; Lettau [21], for the general case $0 \leqq Q \leqq 1$.

Paneth and Glueckauf [27] showed that ν_{He} between 15 and 25 km is slightly larger than at the ground. In contrast to the oxygen deficit measured by Regener, the helium surplus is not a systematic height function. It is relatively small, so that on the scales of Fig. 2, ν_{He}-averages for layers of 2.5 km are practically constant. The ground concentration is $(\nu_{He})_0 = 5.2 \times 10^{-6}$. The assumption that (52) is satisfied yields

$$(\gamma_0/n)_{He} = -(\nu_{He})_0 \mu_{He}\, d_{HeN_2}/H = 3.3 \times 10^{-12}\ \text{cm}^3 He\ \text{cm}^{-2}\ \text{sec}^{-1}, \tag{54}$$

or, integrated over the entire earth's surface (5.1×10^{18} cm^2),

$$\overline{(\gamma_0/n)_{He}} = 17\ \text{m}^3\ He\ \text{sec}^{-1}. \tag{55}$$

By multiplying (54) with Loschmidt's number (Table I), the units of γ_0 are number of molecules per square centimeter per second. Helium is discharged from rocks which contain uranium or thorium. Gutenberg [14] discussed the total lithospheric helium production and estimated its order of magnitude as 10 m^3 sec^{-1}, which is sufficiently close to (55) in order to justify the assumption that $\gamma_0 \approx \gamma_{crit}$. Consequently, helium cannot increase with height, even in the Dalton atmosphere above 200 km. This would correspond to the absence of helium lines in the spectra of the aurora and of the night sky. Similar conditions will exist for hydrogen. The departure into space at the same steady rate with which these lightest gases leave the ground appears consistent with Jeans' theory of escape.

When $\gamma_0 \neq \gamma_{crit}$, equations (33) and (34) must be studied. Lettau [21] considered a small $\partial \nu_{He}/\partial z$ between 15 and 25 km and corrected the result (55) slightly. When γ_0 differs considerably from γ_{crit}, the concentration will vary, even in layers where Q is very small. Examples of such nonpermanent gases are carbon dioxide (produced at the surface over land), ozone (produced in the lower stratosphere), and atomic oxygen (produced in the upper stratosphere), when the regions of production are left outside the layer under consideration. The concentration decreases with increasing distance from the region of production. The intensity of continuous sources can be computed when $\partial \nu_s/\partial z$ is steady and Q is known. As an example, the CO_2-observations of Glueckauf [11] in Great Britain, $\nu_{s0} = 3.4 \times 10^{-4}$ and $\nu_s \approx 2.5 \times 10^{-4}$ at $z \approx 7$ km, yield

$$(\gamma_0/n)_{CO_2} \approx -(\nu_s - \nu_{s0})D/z \approx 5 \times 10^{-5}\ \text{cm}^3\ CO_2\ \text{cm}^{-2}\ \text{sec}^{-1}, \tag{56}$$

for $D = 5 \times 10^4$ cm^2 sec^{-1} and $d_{CO_2N_2} \ll D$. In all probability, this rate of CO_2-diffusion holds for industrialized areas only. For central Europe, Lettau [21] derived 10^{-6} cm^3 CO_2 cm^{-2} sec^{-1} from Wigand's observations. A compensating negative flux of carbon dioxide will exist in the troposphere over the oceans.

In a similar manner, the average downward-directed diffusion of ozone through the troposphere was found to be approximately 10^{-9} cm^3 O_3 cm^{-2} sec^{-1} $= 4 \times 10^9$ molecules cm^{-2} sec^{-1}; Dütsch [8] estimated $(\gamma_0)_{O_3}$ as 7.1×10^{10}, 2.6×10^9 and 0.4×10^9 molecules cm^{-2} sec^{-1} at latitudes 0°, 45°, and 80°N, respectively.

The problem of diffusion equilibrium becomes more complex when $\partial\gamma/\partial z \neq 0$ at certain regions of the layer under consideration (*cf.* equation (27)). Then the main difficulties arise from the mathematical analysis of the rate of production $q^{(+)}$ and extinction $q^{(-)}$ as functions of height. These processes will depend on the physical state of the atmosphere and on solar radiation.

With regard to equations (23) and (27), the classical model of such investigations will be the distribution of ozone. The processes of production and extinction are fairly well known and localized in the lower stratosphere and at ground level. The requirements that N_2 can be considered the medium of diffusion and that ν_s is small are satisfied. Dütsch [8] studied the subject comprehensively and found the effects of molecular diffusion small in comparison with those of eddy diffusion. However, the vertical ozone distribution is not in equilibrium, that is, the main problem is the explanation of time variations at different heights and latitudes due to meridional and seasonal differences of $q^{(+)}$ and $q^{(-)}$, eddy diffusion, and general circulation. Consequently, Dütsch's results will be dealt with in a discussion of three-dimensional diffusion.

Layers above the stratosphere are characterized by the production and destruction of ions due to radiation and recombination. According to the theory of ionization, $q^{(+)}$ is a complicated function of height and the sun's zenith distance when $q^{(-)}$ is proportional to the square of the number density of ions. It must be mentioned that the assumptions made in Section 4 do not hold, especially when N_2 is also ionized; the number density of ions can become of the same order as that of the molecules. Therefore, equation (1) rather than equation (15) must be considered in connection with $q^{(+)}$ and $q^{(-)}$, and the diffusion coefficient of ionized matter must be properly defined. Ferraro [10] found that the effect of molecular diffusion is negligible in the E- and F_1-layers, and small and possibly negligible also in the F_2-layer. Similarly, Bagge [1] showed that molecular diffusion may influence the number density of ions above 200 km. The effect of eddy diffusion was not considered; when reference is made to Section 7, it follows that eddy diffusion should be considered in the E-layer and for the distribution of atomic oxygen around 100 km.

Since the diffusion problems of dissociated and ionized gases are highly complex and decisively determined by forced oscillations due to solar radiation, they will not be studied in this article.

9. Diffusion of Particulate Matter. Let ν_i be the number concentration of particles and let us consider horizontal uniform ν_i-distributions. Then diffusion is in the vertical and in (29)

$$n_i \mathfrak{c}_i = \Gamma \mathfrak{k}. \tag{57}$$

In steady states without sources or sinks, $\Gamma = 0$; thus, $\mathbf{c}_i \cdot \mathbf{k} = 0$; when $\mathbf{k} \cdot \mathbf{V} = 0$,

$$0 = -c_i - D \frac{1}{\nu_i} \frac{\partial \nu_i}{\partial z}, \tag{58}$$

$$\nu_i = \nu_{i0} \exp\left(-\int_0^z \frac{c_i}{D}\, dz\right). \tag{59}$$

By definition, $\nu_i = n_i/n_{N_2} = n_i/n$, when

$$n = n_0 e^{-z/H}, \tag{60}$$

if $T = T_0$ and $g = g_0$. The number of particles per cubic centimeter therefore is

$$n_i = n_{i0} \exp\left(-\frac{1}{H}\int_0^z \frac{D_c + D}{D}\, dz\right), \tag{61}$$

where

$$D_c \equiv c_i H. \tag{62}$$

Equation (58) shows that the vertical concentration gradient is everywhere negative. The term $\ln \nu_i$ decreases in the ratio of D_c/D per 8 km. In layers of high turbulence, ν_i will be practically constant when n_i decreases proportionally to the density of the atmosphere. When $\partial D/\partial z < 0$, ν_i must decrease rapidly with height. Such effects might be the cause of "dust horizons." With regard to Fig. 2, one would expect "dust horizons" at approximately 15 and 70 km, provided dust particles exist in the stratosphere. The optical phenomena due to scattering of solar radiation prove the presence of dust in the stratosphere. With reference to Gutenberg [14], the duration of civil and astronomical twilight leads to the assumption of dust boundaries at approximately 15 and 60–80 km.

Difficulties arise with regard to the rate of descent of atmospheric dust particles. Stokes' law, equation (28), is true for small spheres when Reynolds' number—defined as $Re = a_i c_i \rho/\mu$—is smaller than a fixed value. In air at NTP, $\mu/\rho = 0.2$ cm^2 sec^{-1}; therefore a_i must be smaller than 10^{-2} cm. More accurate is the formula of Oseen, which is valid also for $a_i > 10^{-2}$ cm. Both of these aerodynamic formulas fail for nonspherical particles and for values of a_i of the order of the free path of air molecules (10^{-4} cm at $z = 20$ km and 10^{-2} cm at $z = 60$ km). Humphreys [18] considered Cunningham's corrections of Stokes' law as applied to the stratospheric conditions.

As an example, we assume $c_i = 0.063$ cm sec$^{-1} \approx 50$ m day$^{-1} \approx 20$ km yr^{-1} corresponding to $a_i = 10^{-4}$ cm when, in equation (28), $\rho_i = 2$. Such a_i- and ρ_i-values correspond to volcanic dust particles which can exist in varying amounts in the stratosphere (see §14). Equation (62) then yields $D_c = 5 \times 10^4$ cm^2 sec^{-1}, which equals normal D-values in the troposphere when, in the stratosphere, D is alternately larger and smaller than 5×10^4. The study of optical phenomena appears to be useful for the discussion of turbulence in the stratosphere.

DIFFUSION IN THE VERTICAL—UNSTEADY CONDITIONS

10. The General Equation. In unsteady states, $\nabla \cdot (n_s \mathbf{c}_s) \neq 0$ in equation (15). For nonpermanent gases the equation of continuity yields

$$\frac{\partial n_s}{\partial t} = -\nabla \cdot (n_s \mathbf{c}_s) + q_s^{(+)} + q_s^{(-)}. \tag{63}$$

For the solution of (15) with regard to (63), we assume the gas to be permanent ($q_s^{(+)} = 0$, $q_s^{(-)} = 0$), and

(*a*) the N_2-atmosphere to be steady and at rest when $\nu_s \ll 1$,

$$\frac{\partial n}{\partial t} = \frac{\partial n_{N_2}}{\partial t} = 0, \qquad \frac{\partial n_s}{\partial t} = n \frac{\partial \nu_s}{\partial t}, \qquad \mathbf{k} \cdot \mathbf{V} = 0, \tag{64}$$

(*b*) $T = T_0$, and $g = g_0$,

$$\frac{\partial \ln p}{\partial z} = -\frac{1}{H}, \qquad n = n_0 e^{-z/H}, \tag{65}$$

(*c*) diffusion to be vertical,

$$n_s \mathbf{c}_s = \gamma \mathbf{k}, \qquad \nabla \cdot (n_s \mathbf{c}_s) = \frac{\partial \gamma}{\partial z}, \tag{66}$$

(*d*) gravity to be the only external acceleration,

$$\mathbf{f}_s = 0. \tag{67}$$

When equations (64)–(67) are considered, the differentiation of (23) with respect to z yields

$$\frac{\partial \nu_s}{\partial t} = \frac{d_{sN_2}}{Q_s}\left[\frac{\partial^2 \nu_s}{\partial z^2} + \frac{\partial \nu_s}{\partial z}\left(\frac{\mu_s Q_s}{H} - \frac{\partial \ln Q_s}{\partial z}\right)\right], \tag{68}$$

or

$$\frac{\partial \nu_s}{\partial t} = (d_{sN_2} + D)\left\{\frac{\partial^2 \nu_s}{\partial z^2} - \frac{\partial \nu_s}{\partial z}\left[\frac{\partial \ln d_{sN_2}}{\partial z} - \frac{\partial \ln (d_{sN_2} + D)}{\partial z}\right]\right\} + d_{sN_2}\frac{\mu_s}{H}\frac{\partial \nu_s}{\partial z}. \tag{69}$$

Let Q_s be constant throughout the layer under consideration, $Q_s = \bar{Q}$. Then D is proportional to d_{sN_2}, or D varies with height in the same way as does $1/\rho$. Consequently, the product ρD (usually called the austausch coefficient A) is independent of elevation. We introduce a new variable Z,

$$Z \equiv e^{-z/2H}, \tag{70}$$

such that

$$d_{sN_2} = (d_{sN_2})_0 e^{z/H} \equiv (d_{sN_2})_0 Z^{-2}. \tag{71}$$

For steady cases, equation (46) yields

$$\nu_s = \nu_{s0} Z^{2\mu_s \bar{Q}}. \tag{72}$$

With the aid of equations (68)–(71),

$$a^2 \frac{\partial \nu_s}{\partial t} = \frac{\partial^2 \nu_s}{\partial Z^2} + \frac{1 - b}{Z}\frac{\partial \nu_s}{\partial Z}, \tag{73}$$

where

$$a^2 \equiv 4H^2\bar{Q}/(d_{sN_2})_0 = \text{const}, \tag{74}$$

$$b = 2\mu_s\bar{Q} = \text{const}. \tag{75}$$

Equations (68), (69), and (73) prove that the form $\partial\nu_s/\partial t = K^*\partial^2\nu_s/\partial z^2$—usually called Fick's equation of vertical diffusion when K^* defines the "diffusion power" of the medium—is an inadequate expression of atmospheric diffusion. The differential equation (73) is solved by Bessel functions, and the solution expresses the concentration as a function of z and $(t - t_0)$ when allowance is made for the appropriate boundary conditions of the problem. For example, let $\bar{Q} = \bar{Q}_i$ at $t < t_0$, and $\bar{Q} = \bar{Q}_{ii}$ at $t \geqq t_0$; then, according to (72),

$$\nu_s = (\nu_{s0})_i Z^{2\mu_s\bar{Q}_i} \qquad \text{at } t < t_0, \tag{76}$$

when

$$\lim_{(t-t_0)\to\infty} \nu_s = (\nu_{s0})_{ii} Z^{2\mu_s\bar{Q}_{ii}}, \tag{77}$$

such that the total amount of s-molecules in vertical columns remains constant.

11. The Time Required for Establishing Equilibrium. The "transition period" is the time needed to transform state (76) into (77). The special transformation, where $\bar{Q}_i = 0$ (maximum mixture) and $\bar{Q}_{ii} = 1$ (Dalton atmosphere), was studied by Epstein [9] in terms of Bessel functions, and by Maris [24] and Mitra and Raksit [26] with the aid of numerical and graphical integrations. Lettau [21] derived (68) and (73) in the forms given above, which allow us to discuss more general cases where

$$0 \leqq \bar{Q}_i < \bar{Q}_{ii} \leqq 1. \tag{78}$$

Since the gas is permanent, the total amount of s-molecules in vertical columns of infinite height must be constant,

$$\Psi_s = \int_0^\infty n_s\, dz = \text{const}. \tag{79}$$

We define the number integral of s-molecules as

$$\psi_s = \int_0^z n_s\, dz, \qquad \lim_{z\to\infty} \psi_s = \Psi_s. \tag{80}$$

Since $n_s = \nu_s n$, it follows from (65) and (72) that

$$n_s = n_{s0} \exp[-(1 + \mu_s\bar{Q})z/H], \tag{81}$$

and

$$\psi_s = \frac{(n_{s0} - n_s)H}{1 + \mu_s\bar{Q}}, \qquad \Psi_s = \frac{n_{s0}H}{1 + \mu_s\bar{Q}} = \text{const}. \tag{82}$$

Equation (82) relates n_{s0} to $\bar{Q}$. Therefore, a general value of the ground concentration is defined:

$$n_{s00} = \Psi_s/H = \nu_{s00} n_0\,. \tag{83}$$

Then,

$$n_{s0} = n_{s00}(1 + \mu_s\bar{Q});$$

$$n_s = \frac{n_{s00}}{1 + \mu_s\bar{Q}} \exp[-(1 + \mu_s\bar{Q})z/H], \tag{84}$$

$$(\nu_s)_i = \frac{\nu_{s00}}{1 + \mu_s\bar{Q}_i} \exp(-\mu_s\bar{Q}_i z/H), \tag{85}$$

$$(\psi_s)_i = Hn_{s00}\{1 - \exp[-(1 + \mu_s\bar{Q}_i)z/H]\}. \tag{86}$$

When diffusion (initially characterized by $\bar{Q}_i$) is $\bar{Q}_{ii}$ at $t \geqq t_0$, it follows from (86) that the difference of the number integrals is a height function:

$$(\psi_s)_i - (\psi_s)_{ii} \equiv \Delta\psi_s = Hn_{s00}e^{-z/H}[\exp(-\mu_s\bar{Q}_i z/H) - \exp(-\mu_s\bar{Q}_{ii}z/H)]. \tag{87}$$

The term $\Delta\psi_s$ is zero at $z = 0$ and at $z = \infty$; consequently, $\Delta\psi_s$ must have an extreme value at a finite level, $z = z^*$. The level z^* is defined by $\partial\Delta\psi_s/\partial z = 0$. The differentiation of (87) with respect to z yields

$$z^* = \frac{H}{\mu_s(\bar{Q}_{ii} - \bar{Q}_i)} \ln\frac{1 + \mu_s\bar{Q}_{ii}}{1 + \mu_s\bar{Q}_i}, \tag{88}$$

and

$$\lim_{(\bar{Q}_{ii}-\bar{Q}_i)\to 1} z^* = \lim_{\left(\frac{\bar{Q}_{ii}\to 1}{\bar{Q}_i\to 0}\right)} z^* = \frac{H}{\mu_s}\ln(1 + \mu_s), \tag{89}$$

$$\lim_{(\bar{Q}_{ii}-\bar{Q}_i)\to 0} z^* = H. \tag{90}$$

By definition, $n_s = \partial\psi_s/\partial z$; therefore, z^* is the level where $(n_s)_i = (n_s)_{ii}$ or $(\nu_s)_i = (\nu_s)_{ii}$; that is, where the initial ν_s-curve intersects the final ν_s-curve (compare equations (76) and (77) and, as an illustration, Fig. 3).

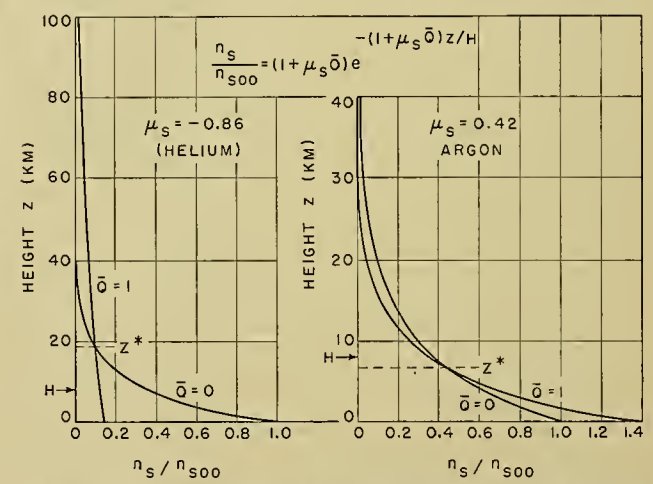


FIG. 3.—The height variation of density (relative to the ground density) of a light and a heavy gas for the border cases of maximum mixture ($\bar{Q} = 0$), and of maximum separation or Dalton atmosphere ($\bar{Q} = 1$). It is assumed that the gases have no sources and sinks.

Our result, equation (90), corresponds to the fact that $z = H$ represents the isopycnic level which was discovered by Wagner and explained by Linke (see Humphreys [18] and Doporto and Morgan [7]).

The transition period has no finite value since $(\nu_s)_{ii}$ will be approached asymptotically even if $\bar{Q}_i$ changes suddenly into $\bar{Q}_{ii}$ at $t = t_0$. The transition period can be studied only in relative terms. Epstein and Maris measured the transition period by the time necessary for accomplishing a certain fraction of the whole transformation. Lettau [21] considered the final time interval

$$\Delta t_s = \frac{\Delta\psi_s}{(n_s)_i(c_s)_{ii}^i}, \tag{91}$$

where $(n_s)_i$ refers to the initial state when $(c_s)^i_{ii}$ is the vertical diffusion velocity as fixed by the final diffusion conditions $(\bar{Q}_{ii})$ and the initial concentration-distribution $(\nu_s)_i$. The changes of both ν_s and c_s with the progress of the transition are not considered; therefore, Δt_s in equation (91) represents a minimum value of the transition period.

Expressions more accurate than equation (91) could be obtained; however, the gain in accuracy does not justify the added complexity of computation in view of the crude assumptions that Q is independent of height and changes its value suddenly and simultaneously throughout the entire atmosphere.

With the aid of (87) and (91),

$$\Delta t_s = \frac{H}{(c_s)^i_{ii}[1 + \mu_s \bar{Q}_i]} \{\exp[-\mu_s(\bar{Q}_{ii} - \bar{Q}_i)z/H] - 1\}. \quad (92)$$

Let us investigate the three special transformations:

$$\left.\begin{aligned} &(a)\ \text{complete mixture } (\bar{Q}_i = 0) \rightarrow \\ &\qquad \text{Dalton atmosphere } (\bar{Q}_{ii} = 1) \\ &(b)\ \text{Dalton atmosphere } (\bar{Q}_i = 1) \rightarrow \\ &\qquad \text{partial separation } (\bar{Q}_{ii} = \epsilon_1) \\ &(c)\ \text{partial separation } (\bar{Q}_i = \epsilon_1) \rightarrow \\ &\qquad \text{partial separation } (\bar{Q}_{ii} = \epsilon_2), \end{aligned}\right\} \quad (93)$$

when $0 < \epsilon_2 < \epsilon_1 << 1$. The diffusion velocity follows from (15)–(19) and from the special assumptions above,

$$\left.\begin{aligned} &(a) \qquad (c_s)^i_{ii} = -\frac{\mu_s}{H}(d_{sN_2})_0 e^{z/H} \equiv c_a \\ &(b) \qquad (c_s)^i_{ii} \approx c_a/\epsilon_1 \\ &(c) \qquad (c_s)^i_{ii} \approx c_a\, \epsilon_1/\epsilon_2\,. \end{aligned}\right\} \quad (94)$$

As an abbreviation, we define $\Delta\tau_s \equiv H^2/\mu_s\,(d_{sN_2})_0$; then equation (92) yields

$$\left.\begin{aligned} &(a)\ \Delta t_s = \Delta\tau_s\, e^{-z/H}[1 - \exp(-\mu_s z/H)] \\ &(b)\ \Delta t_s \approx \Delta\tau_s \frac{\epsilon_1}{1 + \mu_s} e^{-z/H}[\exp(\mu_s z/H) - 1] \\ &(c)\ \Delta t_s \approx \Delta\tau_s \frac{\epsilon_2(\epsilon_1 - \epsilon_2)}{\epsilon_1} \frac{\mu_s z}{H} e^{-z/H}. \end{aligned}\right\} \quad (95)$$

The transition period is zero at the lower and upper boundary of the atmosphere, thus attaining a maximum value at the level defined above, that is, at $z = z^*$. Figure 4 illustrates the conditions for argon when in (b), $\epsilon_1 = \frac{1}{3} \times 10^{-5}$; that is, $D_0 = 0.54 \times 10^5$ cm^2 sec^{-1} or $\bar{A} = 65$ g cm^{-1} sec^{-1}; and in (c), ϵ_1 remains as before, $\epsilon_2 = 0.30 \times 10^{-5}$; that is, $D_0 = 0.60 \times 10^5$ cm^2 sec^{-1} or $\bar{A} = 72$ g cm^{-1} sec^{-1}.

For argon, $z^* = 6.7$ km in (a) and (b). At this level, the transformation of a complete argon-nitrogen mixture into a Dalton atmosphere takes at least 34,000 years. The reverse transformation of a Dalton atmosphere into a steady state of partial separation characterized by a constant austausch coefficient of 65 g cm^{-1} sec^{-1} takes at least 1.5 months. The changes owing to a 10 per cent variation of austausch of the order given above are accomplished in not less than 5 days.

In case (a) the transition period above 200 km is less than one hour whereas it is measured by years below 100 km. Evidently, very slight turbulence will prevent

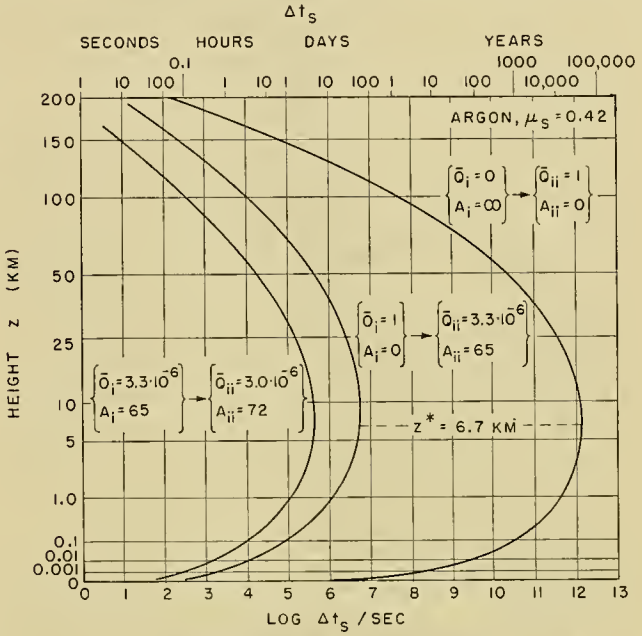


FIG. 4.—The height variation of the minimum values of the transition period Δt_s in three different cases of initial and final diffusion conditions. The computation was carried out for argon; other gases will have different values for Δt_s and for z^*, the height of maximum transition period (see Fig. 3), but the order of Δt_s will not differ from the above. The height scale is proportional to $z^{1/3}$.

the establishment of a Dalton atmosphere below 100 km when the effects of occasionally enlarged turbulence will be extinguished after less than one minute above 200 km. This result supports the findings concerning the height of the datum level (see §7).

When similar computations are based on more accurate assumptions $(\partial Q/\partial z \neq 0)$, the results might be interesting with regard to seasonal and long-time variations of turbulence, or with regard to the intensity of the general circulation and its effect on the composition of the upper atmosphere. Time variations due to variable rates of production of gases should also be investigated with the aid of equations (63) and (73). However, the effect of natural or artificial external sources on the composition of the atmosphere will be very small and measured only within geological epochs.

THREE-DIMENSIONAL DIFFUSION

12. The General Causes of Three-Dimensional Diffusion. As a result of the earth's shape, the use of spherical coordinates (longitude, latitude, height) is natural. Atmospheric diffusion is a three-dimensional problem when the physical state of the atmosphere and/or the strength of sources and sinks depends on height, latitude, and longitude. As indicated by equation (4), the coefficient of molecular diffusion is a func-

tion of density and temperature; the coefficients of eddy diffusion depend on horizontal and vertical gradients of pressure and temperature. These elements will vary at surfaces of constant height in the upper atmosphere while diurnal, interdiurnal, and seasonal variations will be different at different latitudes. However, apart from the troposphere and substratosphere, data are scanty and the effects of meridional variations of d and D on the composition of the upper atmosphere are hypothetical.

In the discussion of diffusion in the vertical, under steady conditions, large-scale external area sources and sinks were considered which covered the entire earth's surface (*e.g.*, helium) or considerable parts thereof (*e.g.*, carbon dioxide). External and internal point or line sources can exist (*e.g.*, meteor trails); production and destruction of matter may be instantaneous, continuous, or periodical as exemplified by volcanic eruptions, lithospheric helium production, or photosynthesis. As demonstrated by the first and second versus the third of these examples, the process can or cannot be independent of the state of the atmosphere. It must be concluded from the above that diffusion in general is not only a three-dimensional, but a time-varying process. Solutions of equations (15), (22), and (63) must be found which satisfy appropriate initial and boundary conditions.

13. The Horizontal Coefficients of Eddy Diffusion and the Effect of the Representative Wind Field on Diffusion Processes. In contrast to molecular diffusion, the coefficients of eddy diffusion at a fixed point may be different in horizontal and vertical directions; with regard to equations (12)–(14), $D_{x,y} \neq D$. Another important factor is that the eddy-diffusion coefficients depend on the averaging process.

Defant [6] considered the traveling cyclones and anticyclones of middle latitudes as individual turbulent elements in an intensive meridional mixing process where $\zeta_D^{(D)} \approx 10^3$ cm sec^{-1} and $\lambda_D^{(D)} > 10^8$ cm are horizontal values; thus, $D_{x,y}^{(D)} = \zeta_D^{(D)} \lambda_D^{(D)} \approx 10^{11}$ cm^2 sec^{-1} (*cf.* Fig. 1). The magnitude of the virtual time-terms, $\lambda_D^{(D)}/\zeta_D^{(D)} > 10^5$ sec, manifests the differences of $D_{x,y}^{(D)}$ in comparison with $D_{x,y}$ due to normal turbulence, since $\lambda_D/\zeta_D = 10$ to 10^2 sec. Representative with regard to $D_{x,y}^{(D)}$ are monthly means; with regard to D, hourly means. The coefficient $D_{x,y}^{(D)}$ effectively equalizes the meridional differences of physical properties and composition; however, very little is known about $D_{x,y}^{(D)}$ in the upper atmosphere.

Even in normal turbulence, D_x or D_y may differ from D, denoting nonisotropic turbulence caused by thermal stratification [22]. In the surface layer, the isotropy coefficient deviates only slightly from unity and it can be expected that throughout the entire atmosphere D_x and D_y have the same order of magnitude as D.

Another important consequence results when the medium of atmospheric turbulence is not "at rest" (*cf.* §4). In reality, $\mathbf{k}\cdot\mathbf{V}$ equals zero for the atmosphere as a whole. The term $\mathbf{k}\cdot\mathbf{V}$ and the components of $\nabla\cdot\mathbf{V}$ and $\mathbf{V}$ are functions of height, latitude, and longitude, since hemispheric circulation cells are superimposed on the general zonal motions. The most general equation of three-dimensional diffusion is equation (63) or

$$\frac{\partial n_s}{\partial t} = -\nabla\cdot[n_s(\mathbf{c}_d + \mathbf{c}_p + \mathbf{c}_F + \mathbf{C}_s + \mathbf{V})] + q_s^{(+)} + q_s^{(-)}, \tag{96}$$

which can be expressed by the individual time variation

$$\frac{dn_s}{dt} = \frac{\partial n_s}{\partial t} + \mathbf{V}\cdot\nabla n_s = -\nabla\cdot[n_s(\mathbf{c}_d + \mathbf{c}_p + \mathbf{c}_F + \mathbf{C}_s)] - n_s\nabla\cdot\mathbf{V} + q_s^{(+)} + q_s^{(-)}, \tag{97}$$

where $q_s^{(+)}$ and $q_s^{(-)}$ and the velocity terms as defined earlier (see §§2–4) are functions of space and time. Lettau [23] found that the term $\nabla\cdot(n_s\mathbf{C}_s)$ depends on the wind shear and that the effects of local wind derivatives increase with time and with the geometrical dimensions of the air mass under consideration. This explains why Fick's simple equation of diffusion

$$\frac{dn_s}{dt} = \nabla\cdot(K\nabla n_s) \tag{98}$$

or

$$\frac{dn_s}{dt} = K^*\nabla\cdot\nabla n_s \tag{99}$$

results in a "diffusion power" K or K^* which must depend on arbitrary physical parameters, such as time of the diffusion process and the geometrical dimensions of the distances under consideration. Richardson [32] and Stommel [36] maintain the point of view that the diffusion coefficient is a function of distance.

Observations of moisture variations along isentropic surfaces in large-scale tropospheric currents to the north or to the south led Rossby [33] and Grimminger [13] to the concept of lateral diffusion processes where the effective coefficients $D_{x,y}^{(R)} \approx 10^9$ cm^2 sec^{-1} are such that $D_{x,y}^{(D)} \approx 10^2\, D_{x,y}^{(R)} \approx 10^4\, D_{x,y}$. Lettau [23] considered, as a criterion of real coefficients of eddy diffusion, that $D = \zeta_D\lambda_D$ when ζ_D and λ_D can be statistically related with observable properties of irregular motions and depend on physical parameters dealing with the energy and the heat transfer of the flow. The term $D_{x,y}^{(R)}$ does not satisfy the requirements of a real diffusion coefficient. Consequently, Lettau [23] termed it an "apparent" coefficient which may be caused by steady deformation fields superimposed on the mean current. This explanation appears to be supported by the findings of Miller [25] that a strong positive correlation exists between the vertical wind velocity and the mean speed of the south-to-north components of the air flow, and the finding of Grimminger [13] that there is a definite indication that the coefficients of lateral mixing increase downstream.

14. Examples of Three-Dimensional Diffusion. All terms of equation (96) were accounted for in Dütsch's study of ozone; the only simplification of (96) was that $\mathbf{c}_F = 0$. Dütsch [8] systematically investigated the ν_{O_3}-equilibrium in vertical columns. Direct radiation and also direct plus diffuse radiation explained only

the basic facts of the vertical distribution of ozone, when incorrect meridional distributions of the total ozone (vertical number integral, as defined by equation (79)) were obtained. Molecular diffusion proved to be negligible; the addition of vertical eddy diffusion improved the results with regard to the meridional distribution. However, the annual variation was still incorrect. Finally, the addition of vertical and horizontal advection due to the average general circulation in the lower stratosphere yielded fairly satisfactory theoretical values. Tables V and VI condense the findings of Dütsch

Table V. Meridional Variations of Total Ozone* (in cm)
(*After Dütsch* [8])

Basis of computation	Latitude North					
	Midsummer			Midwinter		
	25°	45°	70°	25°	45°	60°
Direct radiation	0.52	0.33	0.28	0.31	0.17	0.07
Direct + diffuse radiation	0.39	0.27	0.20	0.26	0.17	0.14
Direct + diffuse radiation + vertical eddy diffusion	0.19	0.20	0.21	0.19	0.20	0.22
Observation	0.20	0.24	0.26	0.19	0.23	0.26

* Under the assumption of equilibrium in vertical columns.

and may be regarded as proving the influence of terms like $n_s \mathbf{C}_s \cdot \mathbf{k}$ and the horizontal components of $\nabla \cdot \mathbf{V}$ in equation (96).

Other studies based on all the terms of equations (96) and (97) do not exist. The main reason is the lack of appropriate observations of gas concentration, including moisture, in the upper atmosphere. Particulate matter and its diffusion appear to be more easily observable.

As an example of dust transportation, a paper by Brandtner [3] may be mentioned. On March 29, 1947, dust from North Africa (latitude 32–35°N, longitude 0–3°E), which was brought into the upper troposphere by the passage of two cold fronts, was transported with south-southwesterly winds and arrived approximately 15 hours later at latitude 50° in western and central Europe. Brandtner's study of the upper-air weather maps proved that the rising of the dust was due to horizontal convergence, the settling to divergence of the wind.

Another example is the eruption of Krakatoa on August 26–27, 1883. The extremely violent catastrophe and the attendant optical phenomena commanded general attention in all parts of the world. Reports like those of Kiessling [19] and Symons [37] offer excellent bases for future studies of atmospheric diffusion.

Great masses of fine volcanic dust were ejected to levels of more than 30 km. Floating with the currents in the stratosphere, the haze caused extraordinary twilight glows and Bishop's ring around the sun. From this, the diameter of the average particle was deduced to be 0.1×10^{-4}—$0.4 \times 10^{-}$

Table VI. Annual Variations of Total Ozone at 60°N (in cm)
(*After Dütsch* [8])

Basis of computation	Month											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Direct + diffuse radiation + vertical eddy diffusion (under the assumption of equilibrium in vertical columns)	0.22	0.22	0.22	0.22	0.22	0.22	0.21	0.21	0.22	0.22	0.22	0.22
Direct + diffuse radiation + vertical eddy diffusion + average horizontal and vertical circulation in the lower stratosphere	0.23	0.25	0.27	0.25	0.22	0.19	0.18	0.17	0.17	0.17	0.18	0.20
Observation	0.27	0.30	0.31	0.30	0.29	0.26	0.24	0.23	0.23	0.23	0.23	0.25

Systematic differences between the last two lines of Table VI can be disregarded; they may be caused by unreliable values of physical constants in the term $q^{(+)}$.

Annual averages of the vertical ozone distribution show a slow increase of O_3-concentration in the troposphere and lower stratosphere when a rapid increase occurs a few kilometers above the tropopause. This appears to be due to rapidly decreasing D-values in these layers. Thus, another proof for the shape of the D-curve about 15 km above sea level, as shown in Fig. 2, is obtained [21]. In this connection, the observed correlation between total amount of ozone and pressure at sea level appears to be explained by vertical oscillations of the tropopause and layers of $\partial D/\partial z < 0$ when the layers of $q^{(+)}$ remain unchanged.

cm. The mean height of the glow stratum decreased 15 km fairly continuously during 5 months which is approximately 0.1 cm/sec [*c.f.* §9]. The glow stratum travelled several times around the globe completing one circuit in approximately 13 days. On the first circuit the band of twilight glows was centered at the latitude of Krakatoa, 6°S., and the mean extension north and south was 15°. During the second circuit the limits were not so determinate. Up to Oct. 5th the rate of lateral expansion was maintained, but after this epoch a distinct retardation in the latitudinal spread of the main body of haze occurred. In November, a sudden rush took place, which by the end of this month, caused the phenomenon to be seen on the major parts of North America and Europe up to latitude 60°. While the material was crossing 30°N it was simply spreading north and south, and afterwards turned around to move from SW to NE [19].

The Krakatoa eruption and the subsequent phenomena show that three-dimensional diffusion problems in the upper atmosphere cannot be solved without considering the average horizontal and vertical motions and the pertinent steady or unsteady deformation fields of representative motion.

PROSPECTS FOR FUTURE WORK

The present-day inadequacies of the field require experimental as well as theoretical investigations.

More plentiful and improved observations of the composition and the geophysical conditions of the upper atmosphere with the aid of high-altitude balloons and rockets will stimulate considerably the interest in diffusion problems. To be certain of the representativeness of the data, it is imperative that upper-level soundings should be distributed more uniformly than before with regard to season and geographic latitude.

The deficiencies of our knowledge are fairly well illustrated by Fig. 2. The purpose of this graph was to demonstrate how an assumed $D(z)$-function affects the height variation of gas concentration in the stratosphere. Only when $D(z)$ can be verified more soundly than by arbitrary assumptions can the composition of the stratosphere be computed.

The most promising method of direct attack requires chemical analyses of the air at levels above 15 km, especially from the layers of presumably small turbulence at 20–30 and 80–100 km approximately. Because of certain facts discussed in Section 4, permanent gases like argon are preferable in such an analysis; nonpermanent gases like helium, ozone, water vapor, etc., must be compared to permanent gases in order to ascertain the location and characteristics of sources and sinks.

Difficulties encountered in the mathematical analysis of time-varying one- or three-dimensional diffusion must, of necessity, confine the discussion to simplified models of the processes. More thorough and critical considerations than before appear necessary in order to avoid misleading results owing to oversimplification of the models used.

REFERENCES

1. Bagge, E., "Die Bedeutung der Ionendiffusion für den Aufbau der Ionosphäre." *Phys. Z.*, 44: 163–167 (1943).
2. Bartels, J., "Überblick über die Physik der hohen Atmosphäre." *Elekt. Nachr.-Tech.*, Bd. 10 (Sonderheft), 40 SS. (1933).
3. Brandtner, E., "Der Staubfall in West Europa am 29. März 1947." *Meteor. Rdsch.*, 1: 222–223 (1948).
4. Chapman, S., and Cowling, T. G., *The Mathematical Theory of Non-uniform Gases.* Cambridge, University Press, 1939.
5. Chapman, S., and Milne, E. A., "The Composition, Ionisation and Viscosity of the Atmosphere at Great Heights." *Quart. J. R. meteor. Soc.*, 46: 357–396 (1920).
6. Defant, A., "Die Zirkulation der Atmosphäre in den gemässigten Breiten der Erde." *Geogr. Ann., Stockh.*, 3: 209–265 (1921).
7. Doporto, M., and Morgan, W. A., "The Significance of the Isopycnic Level." *Quart. J. R. meteor. Soc.*, 73: 384–390 (1947).
8. Dütsch, H.-U., *Photochemische Theorie des atmosphärischen Ozons unter Berücksichtigung von Nichtgleichgewichtszuständen und Luftbewegungen.* Zürich, Leemann, 1946.
9. Epstein, P. S., "Über Gasentmischung in der Atmosphäre." *Beitr. Geophys.*, 35: 153–165 (1932).
10. Ferraro, V. C. A., "Diffusion of Ions in the Ionosphere." *Terr. Magn. atmos. Elect.*, 50: 215–222 (1945).
11. Glueckauf, E., "Carbon Dioxide Content of Atmospheric Air." *Nature*, 153: 620–621 (1944).
12. Götz, F. W. P., "Die Atmosphäre, Beschaffenheit, Schichtung, Erstreckung," *Lehrbuch der Meteorologie*, J. v. Hann und R. Süring, Hsgbr., 5 Aufl. Leipzig, Keller, 1939. (See pp. 3–24)
13. Grimminger, G., "The Intensity of Lateral Mixing in the Atmosphere as Determined from Isentropic Charts." *Bull. Amer. meteor. Soc.*, 22: 227–228 (1941).
14. Gutenberg, B., "Der Aufbau der Atmosphäre," *Handbuch der Geophysik*, Bd. IX. Berlin, Gebr. Bornträger, 1932. (See pp. 1–89)
15. *Handbook of Chemistry and Physics*, 30th ed. Cleveland, Chemical Rubber Publ. Co., 1947.
16. Hann, J. v., "Die Zusammensetzung der Atmosphäre." *Meteor Z.*, 20: 122–126 (1903).
17. Haurwitz, B., "The Physical State of the Upper Atmosphere." *J. R. astr. Soc. Can.*, 30: 315–330, 349–366, 397–415 (1936); 31: 19–42, 76–92 (1937). Also reprinted with additions in monograph form by the University of Toronto Press, 1937.
18. Humphreys, W. J., *Physics of the Air*, 3rd ed. New York, McGraw, 1940.
19. Kiessling, J., *Die Dämmerungserscheinungen im Jahre 1883 und ihre physikalische Erklärung.* Hamburg und Leipzig, L. Voss, 1885, 1888.
20. Lettau, H., *Atmosphärische Turbulenz.* Leipzig, Akad. Verlagsges., 1939; Ann Arbor, Michigan, J. W. Edwards, 1944.
21. —— "Zur Theorie der partiellen Gasentmischung in der Atmosphäre." *Meteor. Rdsch.*, 1: 5–10, 65–74 (1947).
22. —— "Isotropic and Non-isotropic Turbulence in the Atmospheric Surface Layer." *Geophys. Res. Pap.* No. 1, Air Force Cambridge Res. Lab., Cambridge, Mass. (1949).
23. —— *A Theory of Eddy Diffusion.* Unpublished manuscript, 1949.
24. Maris, H. B., "The Upper Atmosphere." *Terr. Magn. atmos. Elect.*, 33: 233–255 (1928); 34: 45–53 (1929).
25. Miller, J. E., "Studies of Large Scale Vertical Motions of the Atmosphere." *Meteor. Pap. N. Y. Univ.*, Vol. 1, No. 1 (1948).
26. Mitra, S. K., *The Upper Atmosphere.* Calcutta, Royal Asiatic Society of Bengal, 1948.
27. Paneth, F. A., and Glueckauf, E., "The Helium Content of Atmospheric Air." *Proc. roy. Soc.*, (A) 185: 89–98 (1946).
28. Penndorf, R., "Die Zusammensetzung der Luft in der hohen Atmosphäre." *Meteor. Z.*, 55: 28–31 (1938).
29. Petersen, H., "On the Influence on the Composition of the Air of a Possible High Temperature in the Highest Strata." *Publ. danske meteor. Inst.*, No. 6 (1928).
30. Regener, E., "Oxygen Content of the Stratosphere." *Nature*, 138: 544 (1938).
31. —— "Ozonschicht und atmosphärische Turbulenz." *Meteor. Z.*, 60: 253–269 (1943).
32. Richardson, L. F., and Stommel, H., "Note on Eddy Diffusion in the Sea." *J. Meteor.*, 5: 238–240 (1948).
33. Rossby, C.-G., and Collaborators, "Aerological Evidence of Large-Scale Mixing in the Atmosphere." *Trans. Amer. geophys. Un.*, 18: 130–136 (1937).

34. Senftleben, H., und Gladisch, H., "Zur Frage der Einwirkung elektrischer Felder auf den Wärmeübergang in Gasen." *Naturwissenschaften*, 34: 187–188 (1947).
35. Spitzer, L., Jr., "The Terrestrial Atmosphere Above 300 Km" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed., pp. 213–249. Chicago, University of Chicago Press, 1949.
36. Stommel, H., *Diffusion Due to Oceanic Turbulence*. Woods Hole Oceanographic Institution, Tech. Rep. No. 17, Woods Hole, Mass. (1949).
37. Symons, G. J., and Collaborators, *The Eruption of Krakatoa and Subsequent Phenomena*. Krakatoa Committee, Royal Society. London, Trübner and Co., 1888.
38. Warfield, C. N., "Tentative Tables for the Properties of the Upper Atmosphere." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1200 (1947).

THE IONOSPHERE

By S. L. SEATON

Geophysical Institute, University of Alaska

Introduction

The ionosphere is that portion of the earth's atmosphere containing a sufficient free-electron population to noticeably affect propagation of electromagnetic waves in the radio-frequency spectrum. It might be described also as that region of our atmosphere in which the refractive index has a value perceptibly less than unity.

The free-electron concentration is brought into being through detachment of electrons from negative ions and by ionization of neutral molecules and atoms. Contemporary thought indicates that the principal electron-liberating agent is ultraviolet light from the sun, but other factors are probably active, for example, bombardment by particles, freeing of electrons through molecular recombination processes, and so on.

There is a tendency towards horizontal stratification of electron distribution. Important concentrations of free electrons exist from about 90 km to over 400 km above sea level. In a broad sense, greatest electron densities are to be found at the subsolar point. Because of the high probability of collision and the estimated maximum energy falling on the atmosphere, it is believed that large free-electron populations cannot exist continuously at altitudes less than about 10 km. Balloon-carried detectors indicate no large concentrations of electrons to 30 km, and radio measurements point to no continuous free-electron populations below about 60 km.

There seems to be no reason to restrict the upper limit of the ionosphere, although above about 600 km the atmospheric density is so low that the free-electron concentration is probably limited. Thus the upper and lower limits of the ionosphere are not well-defined.

Historical

As early as 1880 those studying terrestrial magnetism and its variations postulated an electrically conducting region high in the atmosphere in which electric currents flowed, inducing at the earth's surface a small magnetic field. This postulate seems to have been stimulated by speculation upon the causes of the aurora polaris. However, it was not until 1925 that direct evidence was obtained by Appleton and Barnett in England, and by Breit and Tuve in America, proving the existence of a conducting region in the earth's upper atmosphere. These investigators identified electromagnetic waves returned from a reflecting region above their energy source and determined grossly the height of this reflecting region. At first only one layer was detected but subsequently echoes were found to be returned from other heights, giving rise to the idea of separated strata. More recent investigations point towards a continuous region with maxima and minima.

In using the term *free-electron density*, it must be realized that a free electron is one which has become separated from its original environment as part of a negative ion, a positive ion, or a neutral particle. This free electron is likely to be captured soon by another similar environment, there to remain until set free again. Electrons per se, perhaps arriving from space, are usually neglected in considerations of the ionospheric electron density, since the origin of specific electrons cannot be determined at present. Thus the free-electron concentration is made up of the average number of free electrons per cubic centimeter measured over convenient lengths of time, usually of the order of a fraction of a second or more. The contribution made by electrons other than those supposed to belong naturally to the environment has not been investigated, and indeed the means for such an investigation does not seem to be at hand. The same thing is true in regard to the flux of energy through the region.

Neglecting philosophical utterances, which may have struck home by chance, the mathematician Gauss, followed by the physicists Balfour Stewart and Sir Arthur Schuster, showed from studies of terrestrial magnetic-field variations that there should be a region of high electrical conductivity in the upper atmosphere. Their works were known to only a few, and when Marconi succeeded in sending electromagnetic waves over long distances in 1900 it was not understood how these waves could bend round the earth. Unaware of the earlier suggestions, Professor A. E. Kennelly of Harvard University, and the British engineer, O. Heaviside, independently proposed that an electrically conducting region must exist high in the atmosphere which would bend electromagnetic waves back to earth at a distance. Of these two approaches, that is, inference from terrestrial magnetic variations, and propagation of electromagnetic waves, only the latter has given direct evidence of the ionosphere. The former may have unexpected possibilities. The latter gives information about the former.

Methods for Studying the Ionosphere

In addition to the tools already mentioned, the instrument-carrying rocket and examination of energy arriving at the earth from space promise useful information. Other means of exploring the ionosphere, for example, study of compressional wave propagation, light scattered from modulated searchlight beams, meteorological variations, and cosmic ray changes, are worthy of serious consideration. Further seemingly more remote possibilities exist.

Since 1925 only electromagnetic-wave propagation has given adequate information about the ionosphere. In 1925 there were two laboratories, one near London, England, and one near Washington, D. C., at which direct measurements of the ionosphere were being made. In 1949 the reasonable hope of direct measurement by instrument-carrying rockets presented itself. In 1949 over sixty ionospheric measuring stations scattered throughout the populated regions were in operation. Despite these advances an adequate number of observing stations does not yet exist. There is a serious lack of knowledge over a twelve-degree-wide belt in equatorial regions, and both north and south polar areas are virtually unexplored from an ionospheric standpoint. Regions of especial interest are where geomagnetic and geographical equators cross, where they are farthest apart, and polar regions, particularly in the zones of greatest auroral activity. Intermediate locations are also necessary.

Because the electromagnetic wave has so far been of greatest usefulness in exploring the ionosphere, theoretical developments have occurred allowing the alterations suffered by the wave in its passage through an ionized atmosphere to be interpreted. In 1912 Eccles laid the foundation for determining the effect of charged particles upon the propagation of radio waves. The theory was incomplete and in 1924 Larmor supplied improvements. The Eccles-Larmor theory forms the foundation for understanding electromagnetic-wave propagation in the ionosphere, if the effect of the earth's magnetic field is neglected. The problem becomes much more complicated with inclusion of the effect of collisional friction and of the earth's magnetic field traversing the ionosphere. The magneto-ionic theory developed by Appleton, Nichols and Schelleng, Hartree, Goldstein, and others, together with the theories of Lorentz, Booker, and contemporaries, give the broad foundations for interpretation of information collected experimentally. There are still uncertainties, and without doubt additional elegant theoretical work and experimentation must be done before our knowledge approaches completeness. But notable progress has been made since 1880.

Elementary Concepts

Originally the ionized region of the upper atmosphere was called the Kennelly-Heaviside layer. Later, when stratification was indicated, each investigator named his "layer." The ensuing arguments over names became so troublesome that by mutual consent the whole structure was designated the *ionosphere*, with letters and subscripts to indicate the principal regions within it. There are three principal maxima of free-electron concentration: the E-region at about 100 km, the F_1-region at about 225 km, and the F_2-region at about 350 km. The normal E- and F_1-regions develop maximum electron density with greatest solar altitude and exist only by day. Behavior of the F_2-region is more complicated. With low solar altitude F_1- and F_2-regions merge to form a general F-region which persists throughout the night. In the E-region sporadic ionization occurs with quite large free-electron concentrations. Sporadic E-region phenomena are not understood but the ionization appears to be patchy and does not follow solar altitude control. Figure 1 is drawn to a distorted scale in order to permit visualization of the principal ionospheric regions.

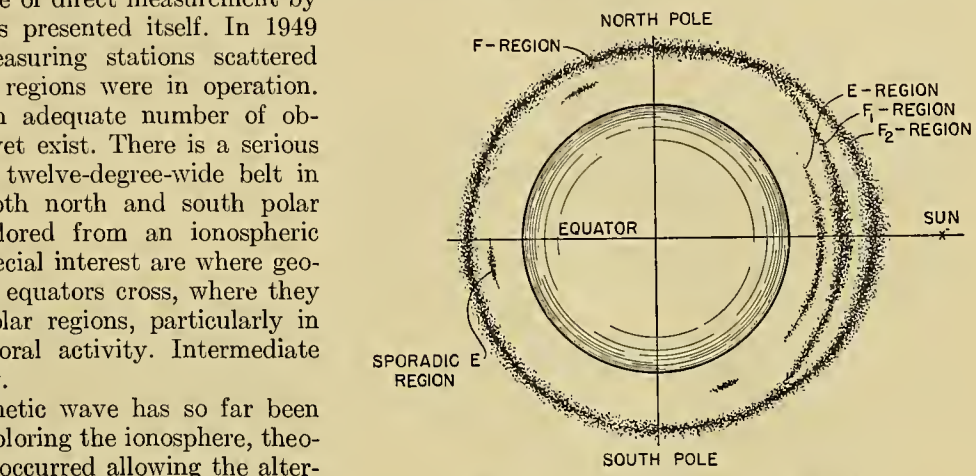


FIG. 1.—Principal ionospheric regions.

In addition to diurnal variations dependent upon solar altitude there is a pronounced change related to solar activity as measured by sunspot numbers. Arguments in favor of some ionization being caused by charged particles bombarding the atmosphere are called up by behavior of the F_2-region. Because the earth's magnetic field extends far into space, charged particles entering the upper atmosphere from space will be deflected by this field. Average maximum F_2-region ionization, studied with respect to the geomagnetic coordinates, shows values in accordance with action of charged particles under the influence of the magnetic field. However, the origin of the particles, their charge, mass, and velocity, are not known. It has not been definitely shown, either, that the observed behavior is surely caused by charged particles. If particles do influence the ionospheric layers, it is not known whether they come from space or whether they are parts of our own atmosphere thrown high up by thermal agitation and returning under gravitational influence.

Typical Data

In exploring the ionosphere by means of the electromagnetic wave at radio frequency, it is usual to emit short wave-trains approximately 10^{-4} sec long of known radio frequency and polarization. The time of departure of each wave packet is noted. One or more wave packets may be expected to return from the ionosphere when the radio frequency emitted lies below a value limited by the number of free electrons and described approximately by

$$N = 1.24\ (10)^4 f^2, \qquad (1)$$

where N is the number of electrons per cubic centimeter and f is the radio frequency in megacycles per second. The returning wave packets are altered during their flight. The elapsed time between transmission and reception, magnitude of returned energy, polarization of the returned wave, splitting of the original wave packet

into two or more returned components, and the angle of arrival give information concerning the medium responsible for return of the wave energy.

In terms of the penetration frequency f_0 and the wave frequency f, the index of refraction for elementary purposes is given by

$$\mu^2 = 1 - \frac{f_0^2}{f^2}. \tag{2}$$

Clearly the index of refraction is a function of frequency and therefore the ionosphere is a dispersive medium. Usually, also, each successive wave packet emitted is of a radio frequency differing incrementally from that preceding. When this process is carried out over a large frequency range, the results can be recorded on a graph as in Fig. 2.

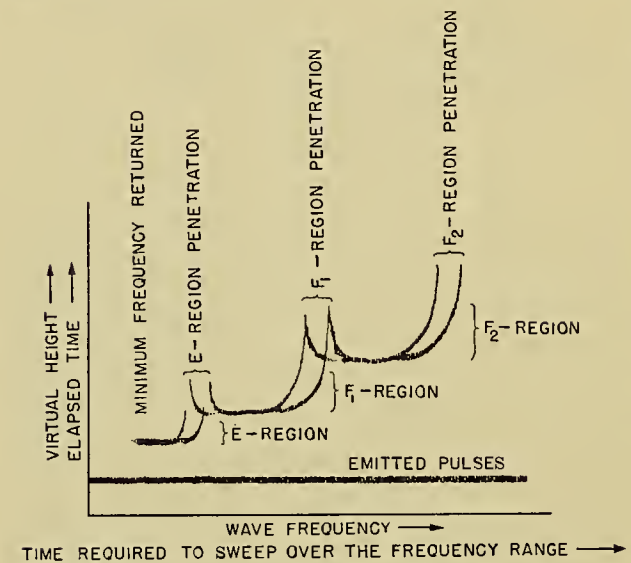


FIG. 2.—Typical data obtained by radio sounding of the ionosphere.

This is the usual form taken by data embracing time of flight versus frequency in temperate latitudes during daylight. Splitting of the original wave packet into two or more components is caused by the effect of the earth's magnetic field. Transmission and reception are at vertical incidence. This technique is the most widely used and is an elegant development of the original Breit-Tuve experiment. Study of the ionosphere at vertical incidence immensely simplifies the general theory for a special case and, although involving some approximations, has given the bulk of useful information about the ionosphere. This procedure fails at low frequencies and is usually replaced by the Appleton-Barnett phase-interference method. Polarization and angle-of-arrival measurements are quite important but have been little used systematically because of inherent interpretative and instrumental difficulties. More attention should be given to them because some uncertainties in the theory cannot be resolved otherwise.

The amount of information to be derived from data represented in Fig. 2 has not been fully realized. For years it was known that the time of flight of a wave packet to and from the ionosphere did not measure the height of the region returning the wave energy. It was not until about 1940, however, that the heights and reasonably sure distributions of ionization were possible. The theoretical investigations of Booker and Seaton, and especially of Rydbeck, now permit distributions to be determined from data in the form of Fig. 2. Briefly, the Booker-Seaton relationship is as follows.

Define a function

$$\varphi(x) = \frac{x}{2} \ln \left| \frac{1+x}{1-x} \right| - 1, \tag{3}$$

and write

$$h' = h_M + \tau \cdot \varphi(f/f_0), \tag{4}$$

where h' is the virtual (or apparent) height at a wave frequency f, and f_0 is the penetration frequency. If heights at two convenient values of f are taken, τ (the semithickness of the parabolic distribution) and h_M (the height of maximum electron density) may readily be determined.

Because of the discontinuities evident in Fig. 2 it was believed at first that the ionosphere was divided into discrete layers. Application of Rydbeck's equations gives a different viewpoint and indicates a general region wherein maxima and minima of ionization occur, but probably no sharp separation into strata.

The difference between penetration frequencies in a region is a measure of the strength of the earth's magnetic field at that height. The reason for this is given in the complicated magneto-ionic theory.

Probable Distribution of Ionization

Consider Fig. 3, in which the heavy curves represent oft-measured quantities, heavy broken lines infrequent

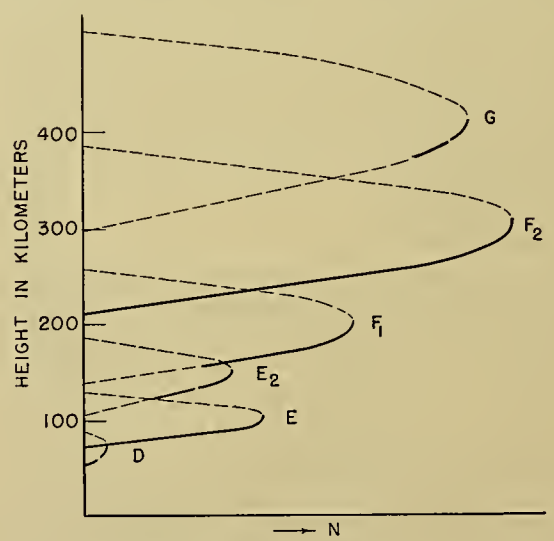


FIG. 3.—Contemporary ideas of free-electron distribution in the ionosphere.

determinations, and dashed lines estimates based upon what seem at present to be reasonable assumptions. Parabolic distributions are shown, but may be replaced

by other distributions as analysis may indicate. Appleton has detected tidal effects in the E-region heights, and Jones has recently found a lunar tide in F-region thickness of some 10 km for a station in high latitudes.

It is to be noted that additional regions are indicated beyond and between those shown in Figs. 1 and 2. The distribution in Fig. 3 represents the most likely ionospheric situation according to contemporary thought. (There is a considerable variation in heights from this typical situation.) The penetration frequencies of Fig. 2 occur whenever the ionization gradient along the direction of wave travel becomes zero in Fig. 3.

The D-Region

Starting at the bottom in Fig. 3, the D-region is not well-defined. There is some evidence that very low radio frequencies are reflected at oblique incidence from heights in the vicinity of 60 km. It is also known that medium-to-high frequencies are absorbed at times in a region below 100 km. The latter information gives no evidence of the height of the absorbing area except that it must be where the mean free path of the electron is small. Wulf and Deming believe that formation of the D-region is the result of photo-ionization of ozone by ultraviolet light. Maximum electron concentration is probably of the order of 2×10^5 electrons per cc. Recombination of electrons with positive ions near the 60-km height is very rapid so that within a few seconds after the ionizing agent ceases to operate any existing free-electron population will disappear. It is in some zone below 100 km that the direct relationship between solar chromospheric eruptions and absorption of radio-wave energy appears. At such times some evidence is available pointing to enhancement of low-frequency oblique-incidence radio-wave propagation by virtue of reflections from a height of about 60 km. Height variations and details of the distribution are unknown. Electron density varies diurnally, seasonally, and with solar activity.

The Normal E-Region

Above the D-region there is to be found a zone in which changes in electron density follow solar altitude quite closely. In the E-region maximum free-electron concentration occurs shortly after local noon; reaching values of the order of 3×10^5 electrons per cc. For the normal E-region, ionization disappears within a few minutes after ultraviolet light from the sun is cut off by the earth. Heights display some variation with geographic location and vary diurnally, seasonally, and with the sunspot cycle. Lowest E-region heights are near 90 km, with maxima near 130 km. Electron distribution within the E-region is not known accurately but probably follows either a Chapman distribution or a parabolic law. The lower boundary is usually well-defined. The half-thickness is of the order of 5 to 20 km. It is in this region that the transition from predominantly molecular oxygen to predominantly atomic oxygen occurs. It is thought that electrons are freed in the E-region by photo-ionization of oxygen molecules.

Sporadic E-Region

In addition to the normal E-region, there occurs an E-region ionization sporadic in occurrence and density. Some of this sporadic ionization is brought about by meteors encountering the earth's atmosphere. However, at present it is believed that meteors are not the only source, but that some other mechanism is present. There is a semblance of system in occurrences of sporadic E-region ionization such that the phenomenon appears most frequently in the arctic regions at night, particularly near zones of maximum auroral activity.

Maximum electron concentrations reach values for the sporadic E-region of over 3×10^6 electrons per cc. Heights appear to be about the same as those of the normal E-region. Duration of sporadic E-region ionization may be from a few seconds to several hours. There is some indication that the sporadic E-region may be of patchy form and at times may move or propagate rapidly. The recombination coefficient between electrons and positive ions is of the order of 1×10^{-8} cc sec^{-1} for the E-region generally.

The E_2-Region

Until recent times it was thought that E- and F_1-regions were virtually separate entities with essentially no intermediate free-electron density. However, enough experimental evidence is now available to indicate the probability of an E_2 distribution centering near 150 km in height. This distribution is not always seen because the maximum concentration is normally less than that of the E-region. Little is known of the E_2-region, but it may be existent by virtue of photo-ionization of atomic oxygen. It appears to occur most frequently in the daytime.

The F_1-Region

The F_1-region exists in identifiable form in the daytime with large solar altitudes. In general behavior the F_1-region resembles the E-region. Maximum densities occur a little after local noon and attain values of the order of 4×10^5 electrons per cc. With low solar altitudes F_1- and F_2-regions merge to form the general F-region which persists throughout the night at all latitudes where measurements have been made. Maximum electron density varies diurnally, seasonally, and with sunspot activity. There is at present no conclusive evidence that ionization is caused by other than ultraviolet light. Wulf and Deming favor one of the nitrogen-molecule absorptions as responsible for formation of the F_1-region. Mohler has pointed out that, because the recombination coefficient between electrons and positive ions is a function of pressure, the height at which maximum production of ionization takes place is probably below the height at which maximum concentration occurs. Heights of the F_1-region vary over wider limits than do those of the E-region. The range in height is roughly from 160 to 280 km. There is a diurnal variation with lowest heights around midday. This region rises slightly to merge with the F_2-region near sunrise and sunset. Seasonal and longer-period variations in height occur and there is a variation with latitude.

The half-thickness of the F_1-region ranges roughly between about 10 and 30 km, depending largely upon time of day, season, and solar activity. Very occasionally during disturbance a well-formed F_1-region appears at night. This latter phenomenon is not understood. The recombination coefficient between electrons and positive ions is of the order of 1×10^{-9} cc sec^{-1} for the F_1-region.

The F_2-Region

The F_2-region of the ionosphere has been extensively studied because its variations present peculiarities not immediately evident in the lower regions. When the F_2-region is mentioned it is customary to think of the general F-region at night and the F_2-region in the daytime as being of the same character. During intervals when the F_2- and F_1-regions are not separately identifiable the region unresolved is thought of as the general F-region and as having F_2-region characteristics predominantly.

An outstanding oddity of the F_2-region (or the general F-region) is that maximum electron density occurs everywhere in the Northern Hemisphere in midwinter when the solar altitude at noon is much lower than in midsummer. From this behavior it is evident that F_2-region maximum electron concentration does not follow solar altitude changes in the same manner as do the E- and F_1-regions. Berkner and Wells, and later Seaton and Berkner, demonstrated the existence of a nonseasonal component in the F_2-region electron concentration which is world-wide in appearance and may be caused by the annual change in distance between earth and sun. Maximum free-electron density in the F_2-region is quite sensitive to solar activity, varying some 300 per cent during the sunspot cycle, but this nonseasonal variation persists throughout.

Frequently, depending upon location, season, and solar activity, the maximum concentration of electrons does not occur near noon. While there is always an afternoon maximum, the time of its occurrence varies from about an hour to several hours after midday and there frequently is a pre-noon maximum. The effect of such variations is often to create a minimum at noon. This noon decrease in F_2-region electron population was originally thought to be caused by heating and consequent expansion of the region. Later it was suggested that such a proposal was untenable when Northern and Southern Hemisphere data were examined over the same time interval. In more recent years a different approach to the description of the ionosphere has arisen in which electron concentration is replaced by total electron content in the region. This concept is particularly important for the F_2-region. When F_2-region behavior is examined by this newer measure, much if not all of the peculiar bimodal behavior is removed, leaving a single maximum occurring after noon. It has been demonstrated that the F_2-region may grow thinner over an interval of time, leaving the electron concentration unchanged but reducing the total electron content. In other words, maximum concentration is not necessarily a correct indication of the behavior of this region, and the use of concentration alone may lead to false concepts especially when rates of electron production, recombination processes, and other factors are sought.

Early unexplained asymmetry of the F_2-region has to some extent been clarified through consideration of data in terms of geomagnetic coordinates. It is in this layer that strongest evidence exists pointing towards ionizing agents other than ultraviolet light from the sun. All characteristics of the F_2-region display wide variations when compared with those of the lower layers.

One of the important methods for study of the ionosphere is examination of electron-concentration variations during eclipse of the sun. The E- and F_1-regions show behaviors in good agreement with theory, but eclipse effects in the F_2-region have been difficult to detect. One reason for difficulty in isolating such effects in the F_2-region is of course the small recombination coefficient between electrons and positive ions, which is of the order of 1×10^{-10} to 1×10^{-11} cc sec^{-1} at these great heights. Another reason for difficulty is the immense variability which is always evident in the electron density of the F_2-region. Some investigators have pointed out that if an important part of the F_2-region ionization is caused by corpuscular bombardment, eclipse of the ultraviolet light would not coincide with eclipse of the particle stream. Corpuscular eclipse has been sought without success. There is some evidence, although not wholly satisfactory, of ultraviolet light eclipse in the F_2-region. The semithickness of this region ranges between about 30 and 100 km.

Free-electron densities in the F_2-region range between about 2×10^4 electrons per cc for winter night and 3×10^6 electrons per cc for daytime during maximum solar activity. The free-electron population in the F_2-region is probably the result of photo-ionization by ultraviolet light of the nitrogen molecules. Ionization of atomic nitrogen and atomic oxygen may also contribute during the middle of the day.

The G-Region

The highest ionospheric region for which any evidence exists is the G-region (see Fig. 3). This region may be present in reality. Occasionally data of the type shown in Fig. 2 develop an additional retardation beyond the F_2-region penetration frequency. In some records where additional reflections appear, the character of the data suggests Rayleigh scattering from patches of ionization of the order of a wave length or less in diameter. On other occasions, however, the data are so suggestive of another higher region that credence must be given this possibility. There is certainly no a priori reason why such a G-region should not be present. It may always be present but, because of rarity of the atmosphere in the vicinity of 400 km, may only develop free-electron concentrations greater than those in the F_2-region during unusual conditions. So little is known of the G-region that it can only be said of the free-electron concentration there that it is almost always less than that of the F_2-region. Height of maximum concentration is probably of the order of 400 to 500 km.

Relations between the Ionosphere and Surface Meteorology

When it is remembered that postulates concerning the existence of the ionosphere originated with mathematicians and magneticians and that proof of the reality of the ionosphere was given by engineers and physicists, it is not astonishing to observe that many years passed before the ionosphericists and the meteorologists found a common interest in the ionosphere from a meteorological point of view. The meteorologist with his nose to the ground and the ionosphericist with his head above the clouds have, since about 1946, permitted themselves to recognize the possibility of mutually profitable cooperative studies of the high atmosphere.

Oliver Wulf investigated F-region variations in connection with upper-air meteorological soundings. The Australians studied ionospheric phenomena together with movement of fronts across the continent of Australia. Appleton searched for and found a small tidal effect in the E-region, and other investigators sought connections between the behavior of upper and lower portions of the earth's atmosphere. One of the greatest hindrances to appreciation of important variations in pressure, temperature, and state of the upper atmosphere has been the concept of an isothermal upper atmosphere in diffusive equilibrium. Serious suspicion was first cast upon this concept by the Norwegian students of the aurora polaris and by Whipple and his colleagues in study of meteor trails. Contemporary thought has all but discarded the idea of such a quiescent upper atmosphere.

Available evidence points towards an ionosphere in which thorough mixing of atmospheric gases is the rule. Proportions of nitrogen, oxygen, hydrogen, and so on, appear to be about the same in the ionosphere as at sea level with the exception that the atomic states may prevail in the ionosphere at higher levels. The amount of water vapor in the ionosphere is thought to be negligible.

From the meteorological standpoint the ionosphere appears to have immense possibilities. It is clear that energy entering the environment of the earth from space first strikes the high atmosphere and at least some of this energy causes large changes in the ionosphere. Ionospheric behavior is quite sensitive to variations in energy in the ultraviolet light portion of the spectrum. The ionosphere probably is influenced to a noticeable extent also by corpuscular streams of energy. Jones has recently isolated variations in thickness of the general F-region which correlate inversely with ground barometric pressures. While this work is incomplete, tentatively it appears that variations in ground barometric pressures lag several hours behind changes in F-region thickness in some latitudes.

Ionospheric Temperatures

Many attempts have been made to deduce, by various means, temperatures in the ionosphere. The first speculations favored an isothermal condition with temperatures near 230K. Later when the noon decrease in F_2-region electron concentration was noted in the Northern Hemisphere, temperatures of the order of 2000K were proposed for this region.

Several approaches have been employed in trying to determine ionospheric temperatures. The Norwegians, remembering that the aurora polaris occurs in the same space as does the ionosphere, have estimated (from spectroscopic measurements) nighttime temperatures of 228K in the vicinity of 100 km and have argued on the basis of similar measurements that this temperature increases as soon as the sun's rays impinge on atmospheric gases at these heights. Penndorf, using scale height of the ionosphere, gives values for the E-region temperature near 350K and for the F_2-region near 700K. Other students give values for the E-region ranging from 100K to 1000K, and for the F-region between 120K and 4000K. Many of the results have been highly speculative and based upon other than conservative estimates.

One of the most promising methods to emerge in recent years is deduction of ionospheric temperatures on the basis of hour-to-hour changes in data of the form of Fig. 2. By means of such data, application of the Booker-Seaton method of reduction to true heights, use of Appleton's arguments, and the invoking of Thomson's relationship between recombination coefficient and absolute temperature, it has been possible to make an objective approach to this problem. The time rate of change of electron density is given by

$$\frac{dN}{dt} = q - \alpha N^2, \tag{5}$$

where q is the rate of production, and α the recombination coefficient between electrons and positive ions. The relationship between the recombination coefficient and the absolute temperature is given by

$$\alpha = \alpha_0 \, (P_0/P)^\gamma \, (T/T_0)^3, \tag{6}$$

where α_0, P_0, and T_0 are reference values of recombination coefficient, pressure, and temperature respectively, and γ probably has a value near $\frac{1}{3}$. Neither of the foregoing expressions is exact, but in this simplified form they serve to illustrate the central line of reasoning. While the results cannot be considered entirely satisfactory, the method has been tested for some twenty geographic locations over a range of conditions. In most instances results are reasonable and self-consistent. One of the unexpected concepts to come from this method has been that of a cellular arrangement of temperature isopleths, indicating the probability of systematic wind systems in the ionosphere. While by no means conclusive, application of this objective method at the same location and time as soundings by rockets gives excellent correspondence at E-region heights.

It is clear from the foregoing discussion that it is no longer acceptable to consider the ionospheric temperature as a simple function of height. Rather it appears to be necessary to examine the data from many stations over the world throughout the seasons and, from the derived information, to construct isopleths descriptive of the arrangement of ionospheric temperatures.

By this method, temperatures for January 1947 taken on a world-wide basis range between 100K and 500K for the E-region; between 50K and 1500K for the F_1-region; and between 100K and 1000K for the F_2-region. The ranges indicated are a function of latitude and of local time; and probably of season, solar activity, and other factors. Thus there seems to be a strong possibility of being able, in the near future, to calculate from the dynamics of the situation the probable wind systems in the ionosphere and to investigate possible connections between the ionosphere and lower-atmosphere meteorology.

Aside from electromagnetic-wave propagation, the study of propagation of sound waves from large explosions throws some light upon E-region temperatures. Preliminary results indicate E-region temperatures of the order of 400K for temperate latitudes.

Probable Formation Mechanisms

Wulf and Deming have given what appears to be the most acceptable discussion of causes of the ionospheric structure. Their work, taken with that of Mohler, constitutes a good basis for further study as knowledge of the ionosphere develops. Only the D-, E-, and F-regions are examined by these authors. Knowledge of the probable existence of the G-region came at a time after these original papers had been presented.

Wulf and Deming have shown that above about 100 km oxygen atoms predominate and that below about 100 km oxygen molecules predominate (other constituents being the same) and have assumed a temperature of 219K up to 100 km, with a temperature of 700K above 100 km, and have consequently decided that the F_2- and F_1-regions are caused by two distinct nitrogen-molecule absorptions and that the E-region is the result of oxygen-molecule absorption. According to these authors, the D-region is the result of a broad absorption by ozone. By inference, using arguments of Wulf and Deming not elaborated upon by them in terms of later experimental evidence, it may be that the E_2-region is formed by oxygen-atom absorption and that the G-region is developed through nitrogen-atom absorption.

It appears in light of later work on ionospheric temperatures that the temperatures assumed by Wulf and Deming may not be correct. However, alteration of the temperature arrangement to what seem to be more appropriate values will only change somewhat the heights where maximum absorptions occur. In fact, the adoption of the more recent temperature postulates will improve the results of Wulf and Deming, giving better agreement between their deductions and the experimental information about the location of the regions.

Need for Additional Studies

While many observations of the ionosphere as deduced from electromagnetic-wave propagation have been made, a great portion of this information is poorly controlled from the standpoint of accuracy. Much of the available information is in the form of short series of measurements over restricted frequency ranges. There is a great need for accurate, long-continued observations over all geographic locations. Only in this way can sufficient data be accumulated to permit satisfactory study of the ionosphere and its relationship to the lower atmosphere. In addition, the various theories need to be re-examined and extended to remove uncertainties now present. The region of ten to twelve degrees above and below the equator at all longitudes is in great need of exploration. Both north and south polar areas are virtually unexplored. In all but polar regions it is entirely feasible for the governments of countries considered civilized to establish permanent observatories for study of the ionosphere. Such programs must lead to results of importance from many viewpoints.

It is quite difficult, on the other hand, to establish stations within about 25 degrees of either pole. It is suggested that with modern high-speed ionospheric recorders it is entirely within the realm of possibility to make systematic ionospheric measurements from long-range aircraft. Such a program, ideally, should be a cooperative one engaged in jointly by all nations, and particularly by those bordering upon the polar areas.

REFERENCES

Examination of the following basic books containing bibliographies will permit all references in the text to be identified, and will permit location of a large body of additional material.

Chapman, S., and Bartels, J., *Geomagnetism*. Cambridge, University Press, 1940.

Fleming, J. A., ed., *Physics of the Earth—VIII, Terrestrial Magnetism and Electricity*. New York, McGraw, 1939.

Manning, L. A., *Final Engineering Report on High Altitude Radio Frequency Propagation*. Palo Alto, Electronics Research Laboratory, Stanford University, 1947.

Mitra, S. K., *The Upper Atmosphere*. Calcutta, The Royal Asiatic Society of Bengal, 1948.

NIGHT-SKY RADIATIONS FROM THE UPPER ATMOSPHERE

By E. O. HULBURT

Naval Research Laboratory, Washington, D. C.

Introduction

In places remote from artificial illumination at night, with no moon, the luminosity of the sky is due to several sources of light, all of which are at a considerable distance. The sources are (1) radiations from the gases of the upper atmosphere, (2) polar aurorae, (3) zodiacal light, (4) comets and possibly scattered sunlight in interplanetary space (if there is something there to scatter the sunlight), and (5) stars and nebulae in interstellar space.

If we omit polar aurorae from consideration, the radiations from the high atmospheric gases are the strongest source of night illumination, being four or five times as intense as all of the other sources combined. In more detail, the intensity of the night-sky light averaged over the sky is divided as follows: For the photographic spectrum [5] region from 3500 to 4500 A about $\frac{1}{6}$ is due to the stars, $\frac{1}{6}$ to the high atmospheric emissions, and possibly $\frac{1}{6}$ to emissions from interplanetary space (the amount attributed to sources in interplanetary space has decreased as the measurements have improved; it is now down to about $\frac{1}{6}$ of the whole and may eventually become a few hundredths or even less); for the visible spectrum [17] seen with the dark-adapted eye, about $\frac{1}{5}$ is due to the stars and $\frac{4}{5}$ due to the high atmosphere; for the entire spectrum from about 3000 to 10440 A, because of its strong infrared nitrogen emissions, the high atmosphere contributes more than $\frac{9}{10}$ of the nocturnal radiation and the stars less than $\frac{1}{10}$. These fractions are average values; for areas in the sky such as the Milky Way where there are many stars the fractions of the sky luminosity due to the stars are greater than the average values, and for areas in the sky where there are few stars the fractions are less.

Spectrum

The night-sky light is feeble, and even with spectrographs of the highest light-gathering power, relatively low dispersion and long exposures are necessary to obtain spectra. We need to refer only to recent work and mention that the best spectra at present available were those obtained by Cabannes and Dufay [8] in 1933–34 in France; by Elvey, Swings, and Linke [12] in 1939 at the McDonald Observatory, University of Texas; by Barbier [4] in 1942–44 at the Observatory of Haute Provence, France; and by Meinel [19] in 1948 at the Lick Observatory, University of California.

The dispersion of the spectrograph of Barbier was 150 and 630 A mm^{-1} at 3200 and 5000 A, and the length of the spectrum from 3200 to 5000 A was about 7 mm; exposures from 100 to 200 hours were used. The McDonald spectrograph had slightly higher dispersion and much greater light-gathering power; the dispersion was 115 and 500 A mm^{-1} at 3200 and 5000 A, the length of the spectrum from 3200 to 5000 A was about 10 mm, and exposures of 9 hours were used. These were prism spectrographs and therefore had low dispersion for the longer wave lengths. To obtain better data for wave lengths longer than 6000 A, Meinel built a spectrograph with an f/1 camera and a 7500 line per inch transmission grating with a dispersion of 250 A mm^{-1} in the first order; exposures of 30 hours were used. Night-sky spectra are shown in Fig. 1, in which *A* is a spectrum photographed by Barbier [18], *B* by Elvey, Swings, and Linke [12], and *C* by Meinel [19]. In Fig. 1*C* the narrow

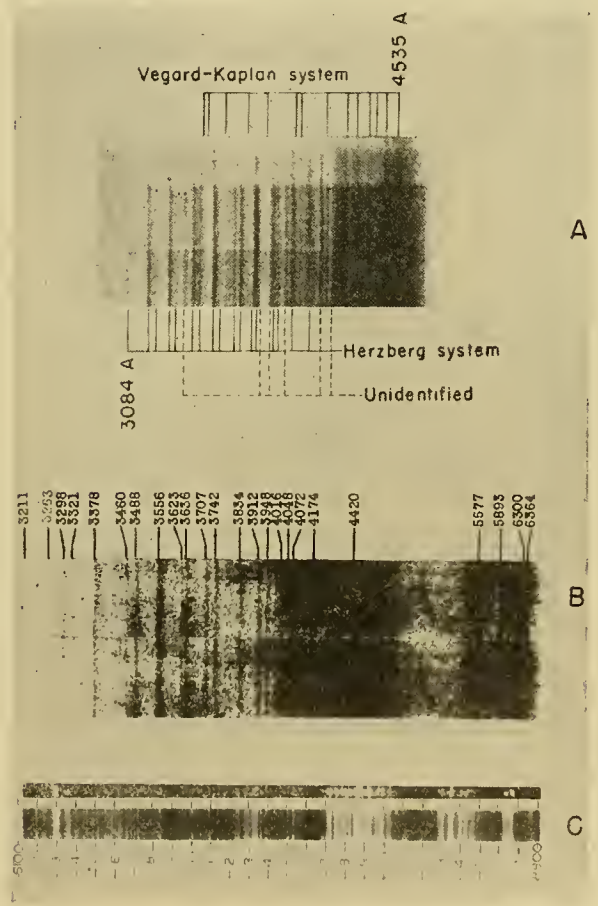


FIG. 1.—Spectra of the night sky, *A* by Barbier [18] (*reproduced by permission of The University of Chicago Press*); *B* by Elvey, Swings, and Linke [12]; and *C* by Meinel [19].

upper strip was the original spectrum from which the lower strip was obtained by spreading; Meinel remarked that many of the features of the lower strip were spurious and were caused by grains of the photographic plate.

The night-sky spectrum is very complex and is composed of many lines and bands, for the most part in-

completely resolved. In Table I are listed the wave lengths of the night-sky emissions and their identifications which are probably agreed to at present; "h" indicates the Herzberg system of molecular oxygen, and "vk" the Vegard-Kaplan system of molecular nitrogen. Wave lengths 8829 to 6300 A in Table I were from the list of Meinel [19], and wave lengths 5896 to 3029 A from the lists of Déjardin [9]. The sources of the emissions whose identification seemed certain are given in Table I. The only atomic lines are the green line 5577 and the red pair 6364, 6300, which are forbidden transitions of oxygen, and the yellow pair 5896, 5890 of sodium. All the remaining lines and bands are of molecular origin. The line at 10440 A of the first positive system of N_2 is by far the strongest emission thus far observed, being about thirty times as intense as the green line 5577, the next most intense emission. It was discovered in 1940 by Stebbins, Whitford, and Swings [25], by means of a photoelectric cell and filters, and its wave length was determined within ±25 A. It was also discovered independently by Herman, Herman, and Gauzit [15]. A group of lines 6580, 6530, and 6470 A, also due to the first positive system of N_2, were mentioned by Déjardin, but were not listed by Meinel; they cannot be seen in Fig. 1*C*, and have not been included in Table I.[1] About 35 bands, which include the strongest blue-violet bands, were identified with the Vegard-Kaplan system of N_2, and about 33 bands, which include the strongest ultraviolet bands, were identified with the Herzberg system of O_2.

Table I. Wave Lengths in Night-Sky Spectrum

λ	Intensity	Source	λ	Intensity	Source	λ	Intensity	Source
10440 A		N_2	4837	4	N_2,vk	3982	1	N_2,vk
8829	3		4824	2	O_2,h	3960	2	
8770	4		4810			3949	3	N_2,vk
8694	1		4798			3914	4	
8659	6		4772	2	N_2,vk			
8628	6		4739	0	N_2,vk	3901	2	
8515	1		4715	1	N_2,vk	3888	2	N_2,vk
8496	3					3873	2	
8466	5		4693			3854	1	N_2,vk
8431	8		4682	1	O_2,h	3848		O_2,h
			4670	3	O_2,h	3833	3	O_2,h
8398	6		4650	1	N_2,vk	3818	2	
8379	2		4632	1		3778	2	
8346	12		4615	1	N_2,vk	3752	2	N_2,vk
8311	4		4604		N_2,vk	3740	4	O_2,h
8287	3		4581					
8102	1		4569			3704	2	
8066	2		4552		O_2,h	3673	1	
8026	4					3661	2	N_2,vk
7992	7		4535	5	N_2,vk	3637	3	O_2,h
7965	6		4520	2		3623	1	
			4488		N_2,vk	3597	1	N_2,vk
7916	10		4478	3	N_2,vk	3584	1	N_2,vk
7885	5		4469			3571		
7856	7		4461			3554	5	O_2,h
7821	6		4442	2	O_2,h	3545	4	O_2,h
7789	4		4421	5	N_2,vk			
7752	6		4408	4	O_2,h	3512	1	
7717	8		4396			3497	1	N_2,vk
7405	1					3483	3	O_2,h
7336	3		4378		N_2,vk	3469	2	
7283	2		4360	2	N_2,vk	3460	1	O_2,h
			4348			3452	2	O_2,h
7249	1		4327	2		3433	1	
7148	1		4316		N_2,vk	3424	1	N_2,vk
7093	1		4286	1	O_2,h	3393	1	
6364	2	*O*	4278		O_2,h	3385	1	O_2,h
6300	4	*O*	4270	4	N_2,vk			
5896		*Na*	4257	2		3373	4	O_2,h
5890		*Na*	4239	1		3319	2	O_2,h
5775						3296	3	O_2,h
5675			4219	0	N_2,vk	3284	1	O_2,h
5577		*O*	4203			3264	2	
			4188			3230	1	O_2,h
5460			4180			3220	2	
5320			4169	4	N_2,vk	3212	3	O_2,h
5250			4158		O_2,h	3201	1	N_2,vk
5160			4140	2	N_2,vk	3191	1	
5130		O_2,h	4131	1				
5090	2	N_2,vk	4117	1		3164	1	O_2,h
5040	2		4111			3157	1	
5002	1	O_2,h				3144	3	O_2,h
4960	1	N_2,vk	4088	3		3133	1	O_2,h
4931	3		4071	5	N_2,vk	3103	1	
			4063		O_2,h	3095	1	
4904		N_2,vk	4048	3	N_2,vk	3084	1	O_2,h
4889			4018	3		3029	2	O_2,h
4869	1	N_2,vk	4004					

In addition, in Table I, there are a number of unidentified lines or bands due to unknown systems. A weak continuous background with a trace of Fraunhofer absorptions, which was mentioned in earlier work, was not observed in the better, more recent spectrograms. It was this background which was attributed to sunlight scattered by material in interplanetary space. But since its existence is now uncertain one may be doubtful of inferences based on it. The Shumann-Runge system of O_2 and the Lyman-Birge system of N_2 have been looked for, but no certain identifications have been made; the same is true for *CO* and CO_2, H_2, O_3, NO_2, N_2O_4, N_2O, N_2O_3, NH_2, NH_3, *CH*, *CN*, H_2O, and compounds of *Na*, *Si*, and *S*.

The spectral intensity distribution received at the surface of the earth [6] is given approximately in Table II.

Table II. Intensity of Night-Sky Emissions

Wave length (A)	Emitter	Intensity (erg cm^{-2} sec^{-1})
10440	N_2	190 × 10^{-4}
6300, 6363	*O*	5 × 10^{-4}
5890, 5896	*Na*	0.7 × 10^{-4}
5577	*O*	7 × 10^{-4}
Main band systems	N_2, O_2	10 × 10^{-4}

The only reported spectral energy distribution of the night-sky light is that of Babcock and Johnson [3] derived from measurements of a small photographic spectrum of dispersion 1100 A mm^{-1} at $H\gamma$. Their curve is given in Fig. 2; it rises with increasing wave length throughout the extent of the spectrum from 3800 to 6500 A, with bumps at the various lines and bands A color temperature of 3450K was ascribed to it. Thus the night-sky light may be said to be yellowish in color.

Variations

If we exclude polar aurorae, the variations in the total intensity of night-sky emission are relatively

1. In a private communication Dr. Meinel stated that photometric traces of his spectra showed a faint feature centered at 6562 A.

small, rarely as much as a factor of 2. During the full nighttime hours there appear to be many small irregular and complex intensity fluctuations which vary with the

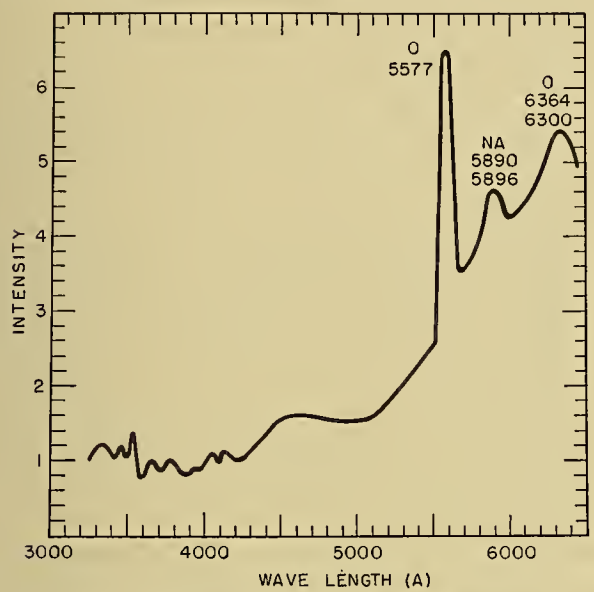


FIG. 2.—Spectral energy distributions of night-sky light. (*After Babcock and Johnson* [3].)

place in the sky and which are different for the various wave lengths. There are strong intensity changes during twilight.

Diurnal Variations. In France, Dufay and Tcheng Mao-lin [10] obtained 588 spectrograms during 189 nights from October 1940 to January 1944. From these spectrograms systematic measurements were made of the intensity variations of the oxygen lines 5577 and 6300 A and of the sodium pair 5893. The oxygen line 5577 A usually showed a weak maximum around midnight (not more than 40 per cent) often obscured by fluctuations; 6300 and 5893 A merely weakened slowly during the night, the enfeeblement for 6300 being greater than that for 5893 A, which amounted only to 10 or 15 per cent. In Russia [23] photocell measurements of radiation in the infrared region 850 to 11000 A indicated that the intensity was at a maximum around midnight, being about twice as great then as at 10 P.M. and 2 A.M.

Twilight Effects. The green line 5577 was observed to exhibit no twilight enhancement. On the other hand, the sodium yellow emission 5893 dropped suddenly in intensity to one per cent of its former value at about ten minutes after sunset [7] with a corresponding recrudescence near dawn. Likewise the red oxygen lines 6300 showed a similar but slower change during twilight [13]. The strong N_2^+ flash at twilight discovered by Slipher is intense in aurorae but absent from the night sky.

Annual Variations. In temperate latitudes 5577 has slight maxima in October and February, and a 27-day variation in intensity, hence a faint correlation with solar phenomena; 6300 and 5893 have maxima near midwinter and minima in summer of amplitude factors of about 2.5 and 5, respectively; they seem indifferent to solar phenomena [2, 10, 12].

Latitude Variations. A survey [17] of the visual brightness of the night sky in various latitudes was carried out with standardized visual photometers which had a field of view about 11° in diameter. Visual measurements of the night-sky luminosity refer mainly to the strong atomic oxygen line 5577. This is because, by multiplying the night-sky spectral energy curve of Fig. 2 by the sensitivity curve of the dark-adapted eye, one finds that 5577 has the main effect (8/10) with only minor influence from the oxygen red lines 6300, 6363, the sodium yellow lines 5896, 5890, and the molecular bands in the blue and green.

Observations were made at stations in Bocaiuva (Brazil), Bikini, Maryland, Whiteface Mountain (New York State), and Greenland, at latitudes −17°, +12°, +39°, +45°, and +63°, respectively. The results are plotted in Fig. 3, in which the ordinates are the night-

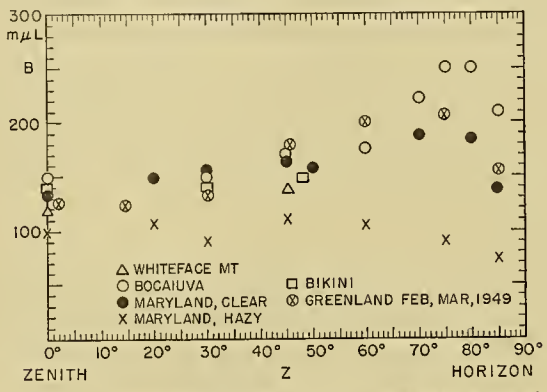


FIG. 3.—Average night-sky brightness at several localities [17].

sky brightness B in millimicrolamberts (mμL; 1 mμL = 10^{-9} lambert = 2.96×10^{-7} candle ft^{-2}) and the abscissas are the zenith angle Z of the place in the sky. Each point of Fig. 3 was the average of values observed in several directions for several nights when no polar aurorae were visible [17].[2] From the data of Fig. 3 it was concluded that there were no changes with latitude in the visual night-sky brightness which could not be attributed to variations in haze.

In connection with the geographical distribution of night-sky intensity, Farnsworth and Elvey [13] concluded that their photographic observations with a glass-prism spectrograph indicated no major differences at Bosque Alegre, Argentina (32°S), Portrerillos, Chile (27°S), and in the southern New England states; their data were too meagre for much generalization. Rayleigh [21] concluded from his visual observations that the average intensities of the blue, green, and red spectral regions at Terling, England (51°N), Capetown, Africa (34°S), and Canberra, Australia (35°S) were of comparable magnitude, but tended to be highest at Terling

2. Data for latitude +63° were observed in February and March, 1949, by personnel at a United States Air Force weather station in Greenland using a Naval Research Laboratory visual low brightness photometer.

and lowest at Capetown. Visual measurements of the sky near Polaris, reported by Fessenkoff [14], indicated that the brightness increased by a factor of 4 from latitudes +44° to +80°, but the result may have been influenced by polar aurorae; it is not in accord with the observations of Fig. 3.

Variation with Zenith Angle. The variation of the visual brightness of the night sky with zenith angle is also shown by the data of Fig. 3, which as has been pointed out refer mainly to the oxygen green line 5577 A. It is seen that the variation of the brightness with zenith angle Z was about the same for latitudes from −17° to +63°, and that the brightness increased from about 130 mμL at the zenith to a maximum of about 230 mμL at about 15° above the horizon for a clear atmosphere; with slight haze the maximum near the horizon became less pronounced, and with more haze it disappeared [17]. The variation was observed at the Pic du Midi Observatory, France, to be somewhat different for red and green wave lengths [1], the sky intensity observed through a red filter being about 2.2 times as bright at $Z = 80°$ as at the zenith, and through a green filter being about 1.7 times as bright at $Z = 80°$ as at the zenith. The difference was probably due to the greater scattering of the lower atmosphere for green wave lengths than for red wave lengths.

Altitude

The altitudes of the regions in which the high atmospheric emissions originate are not known with certainty. The method which has been used has been well described [17] and need not be detailed here. It is based on the observed variation of the night-sky intensity with zenith angle, and when applied to the night-sky observations led to erratic and discrepant values from fifty to several hundred kilometers. The uncertainties were due to nonuniformity of the emissions and to incompletely worked-out corrections for atmospheric attenuations. But even with better correction formulas no improved or more trustworthy values of the altitudes were determined [17]; and in no case was the atmospheric attenuation measured at the same time that the sky intensity was observed. Swings [26] stated the situation thus:

> There are great irregularities in the distribution of the emission over the sky. These irregularities appear consistently when simultaneous exposures in different azimuths at the same zenith distance are compared. No layer of uniform brightness exists, but rather a set of bright clouds, which move around and change brightness. The only remaining hope is that, by taking a sufficiently large number of observations, the erratic fluctuations will average out.

From 1941 to 1944 Dufay and Tcheng Mao-Lin [11] in France made several series of measurements with their spectrograph of the ratios of the intensities of various wave lengths at the zenith to the intensities near the horizon. From averages of the observations with approximate corrections for estimated atmospheric attenuation, they obtained an altitude of 103 km for 5577 A, 80 km for 5892 A, and 181 km for 6300 A. In 1948 Roach and Barbier [22] carried out surveys of the night-sky intensity in California with recording photocell equipment and interference filters for isolating narrow regions of the spectrum. By using average values, with correction for scattered starlight and estimation of atmospheric attenuation, they obtained an altitude of 100 km for 5577 A and about 300 km for 5892 A. Plans are under way to carry out direct photocell measurements from rockets to determine whether the nocturnal radiations arise in regions accessible to the rockets.

The altitude of the yellow sodium line 5892 A in twilight was determined in another way based on the fact that these radiations decreased suddenly in intensity at twilight when the region in which they originated was shielded by atmospheric ozone from the direct ultraviolet rays of the sun. Calculations made by Penndorf [20] from the observations of Vegard and Tønsberg [27] in February and March 1939, at Tromsö and Oslo, indicated definitely that the altitude of the sodium emission at twilight in those places was in the region from 80 to 116 km.

At present one should keep in mind the probability that the various night-sky radiations may arise at different levels, and that the levels may change in the course of the night and may vary with the latitude and season. Further, it seems reasonable to think that the night-sky emissions originate mainly in the atmospheric levels of the aurorae and the ionosphere, that is, from about 80 to 300 km, for it is in this region that strong photochemical effects occur.

Theory

The spectral identifications show that most of the night-sky emissions come from the two most abundant gases of the atmosphere, oxygen and nitrogen; a small portion comes from sodium. No one has questioned the general notion that the nocturnal emissions derive their energy from the sun, but no processes or quantitative details have been agreed upon because information is lacking about radiation transitions and atmospheric composition. In fact, without exception, all theoretical processes, qualitative or quantitative, thus far proposed, have been weighted down with a wealth of criticism [6].

The excitation energy of the line systems 5577 *O* I, 6300 *O* I, 5893 *Na* I are 4.2, 2.0, and 2.1 ev, respectively; and of the band systems Vegard-Kaplan N_2, Herzberg O_2, Schumann-Runge O_2, and Lyman N_2 are 7.0, 4.7, 6.2, and 8.7 ev, respectively. Three theoretical suggestions for the source of the energy in the high atmosphere are that it arises from (1) the association energies of atomic oxygen and nitrogen which are 5.09 and 7.38 ev, respectively; (2) the energies of ionization, which are about 13.5 and 14.5 ev for the first ionization potentials of atomic oxygen and nitrogen, respectively; (3) particles of solar origin or particles swept up from the dust of interplanetary space. It appeared that (1) was sufficient for the line systems but for none of the band systems except that of Herzberg (perhaps the Shumann-Runge and the Lyman band systems need no longer be considered because, as has been said, they are not listed in

the more recent night-sky spectral identifications); (2) seemed energetically satisfactory, but not completely in accord with the Frank-Condon probabilities; (3) is speculative and has not been put in quantitative terms.

The relatively small amount of sodium in the upper atmosphere necessary to account for the 5890, 5896 emissions may come either from the surface of the earth, as the sea, or from interplanetary space. Both views have been suggested and the correct one is not known. There have been no explanations of the remarkable winter enhancement of the sodium emissions observed in temperate latitudes. It would be of interest to observe the emissions in low latitudes. Whether the sodium emissions are merely a minor geophysical phenomenon or have wider astronomical interest is not yet clear.

Zodiacal Light

The zodiacal light is a well-known luminous phenomenon of the night sky which adds appreciably to the brightness of certain regions of the sky near the ecliptic. At present the facts available are not adequate to allow us to say whether the zodiacal light arises in the terrestrial atmosphere or at a greater distance, and therefore space will not be taken here to describe it. There are two theories to explain the zodiacal light. An older theory [24] attributes it to sunlight scattered from a lens-shaped disk of dust particles in the plane of the ecliptic extending well beyond the orbit of the earth; a more recent theory [16] attributes it to sunlight absorbed and re-emitted by a band of atmospheric ions surrounding the earth in the outer fringes of the upper atmosphere. Each theory has points in its favor but further investigations of the zodiacal light with modern spectrographic and photocell techniques and better determinations of its parallax are necessary to reach a correct explanation.

Aurora

We have attempted to exclude the aurora from the foregoing survey of the night-sky light knowing that this cannot be done, and in the end should not be done. It is probable that a complete understanding of the night sky cannot be reached without at the same time having a complete understanding of the aurora. Both, it is believed, derive their primal energy from the sun. The aurora emissions are mainly from the same familiar gases of our atmosphere, oxygen and nitrogen, as are those of the night sky, but with different spectral distributions, with enormously greater intensities and in quite different geometrical patterns. The auroral energy is believed to be received in concentrated form from explosions or outbursts on the sun, in contrast to the relatively gentle and steady solar-energy stream which supports the normal night-sky light. The kinship between the aurora and the night-sky light may be analogous to that between a hurricane and a gentle rain.

Suggested Experiments

At the present time enough is known about the light of the night sky to indicate that it is an important and interesting phenomenon of the upper atmosphere. Some further experiments may be suggested which may increase understanding of the nature and origin of the light. However, the experiments are not easy and require the best of modern equipment. This is true, of course, of most experimental sciences, for in any field the easy experiments are done so quickly by the pioneers that later experimenters almost always find themselves in the stage where further progress is difficult.

Improved spectra of high dispersion of the night-sky light should be obtained. For such a purpose powerful spectrographs and the facilities of a first-class observatory are required. The spectra would provide needed additional information about the types of emitting atoms and molecules, and perhaps might answer the important question whether *any* of the night-sky light is sunlight scattered by material sufficiently distant from the earth to be in sunshine. Spectra should also be taken of the zodiacal light in order to discover whether it is a luminous apparition which originates in the outer reaches of the terrestrial atmosphere or in some more distant place. Spectra of the aurorae should be included.

The distribution of energy in the spectrum of the night-sky light should be redetermined with improved equipment; this has been done only by means of spectrographs of very low dispersion. Detailed measurements of the intensities of the lines and bands of the spectrum should yield information about the excitation phenomena of the radiations and the physical state of the regions of the atmosphere in which they occur.

Surveys should be made over the entire sky in the several wave-length bands in order to determine the nature and behavior of the variations of the luminosity. Sensitive photocells and filters which pass narrow bands of wave lengths could be used for this purpose. Such surveys should be carried out at several stations for a period of time to bring out geographical and temporal variations and to establish relations, if any, with other phenomena, such as season of the year, solar activity, and ionospheric variations. From the variation of the intensity with zenith angle the data might yield better determinations of the altitude of the luminosity. The strong infrared radiations should be included in the survey.

The survey of the night-sky brightness in several wave lengths with carefully standardized and calibrated photometric apparatus should be extended to high latitudes and regions near the magnetic pole in order to determine the geographical distribution of the luminosity. Such observations might throw further light on the differences between the night-sky light and aurorae.

Measurements with photocells should be made from rockets fired at night to heights of 130 km and above in order to discover whether the rocket enters or traverses the regions of the nocturnal luminosity.

REFERENCES

1. Abadie, P., Vassy, A., et Vassy, É., "Altitude de l'émission lumineuse du ciel nocturne." *Ann. Géophys.*, 1: 189–223 (1944).

2. Babcock, H. W., "Radiations of the Night Sky Photographed with a Grating." *Publ. astr. Soc. Pacif.*, 51: 47–50 (1939).
3. —— and Johnson, J. J., "A Spectrophotometric Study of the Light of the Night Sky." *Astrophys. J.*, 94: 271–275 (1941).
4. Barbier, D., "Le spectre du ciel nocturne dans la région des longueurs d'onde inférieures à 5000 A." *Ann. Géophys.*, 1: 224–232 (1945).
5. —— "Mesures spectrophotométriques sur le spectre du ciel nocturne (λλ 4600–3100)." *C. R. Acad. Sci., Paris*, 224: 635–636 (1947).
6. Bates, D. R., "Theoretical Considerations Regarding the Night Sky Emission," in *Emission Spectra of Night Sky and Aurora.* Gassiot Comm. Rep., Phys. Soc., London, 1948. (pp. 21–33); "The Origin of the Night Sky Light." *Mon. Not. R. astr. Soc.*, 106: 509–514 (1946).
7. Bernard, R., "Das Vorhandensein von Natrium in der Atmosphäre auf Grund von interferometrischen Untersuchungen der D-Linie im Abend- und Nachthimmellicht." *Z. Phys.*, 110: 291–302 (1938).
8. Cabannes, J., et Dufay, J., "Le spectre du ciel nocturne dans les régions bleue et violette." *Ann. Géophys.*, 1:1–17 (1944).
9. Déjardin, G., "Le rayonnement du ciel nocturne; description du spectre et identification des radiations observées." Gassiot Comm. Rep., Phys. Soc., London, 1938. (pp. 3–8).
10. Dufay, J., et Tcheng Mao-lin, "Recherches spectrophotométriques sur la lumière du ciel nocturne dans la région visible." *Ann. Géophys.*, 2: 189–230 (1946).
11. —— "Tentative détermination de l'altitude des couches lumineuses de l'atmosphère pendant la nuit." Gassiot Comm. Rep., Phys. Soc., London, 1948. (pp. 62–69)
12. Elvey, C. T., Swings, P., and Linke, W., "The Spectrum of the Night Sky." *Astrophys. J.*, 93: 337–348 (1941).
13. Elvey, C. T., and Farnsworth, A. H., "Spectrophotometric Observations of the Light of the Night Sky." *Astrophys. J.*, 96: 451–467 (1942).
14. Fessenkoff, B., "On the Luminosity of Nocturnal Sky in Different Latitudes." *C. R. (Doklady) Acad. Sci. URSS*, 32: 320–322 (1941).
15. Herman, R., Herman, L., and Gauzit, J., "Infra-red Spectrum of the Night Sky." *Nature, Lond.*, 156: 114–115 (1945).
16. Hulburt, E. O., "The Zodiacal Light and the *Gegenschein* as Phenomena of the Earth's Atmosphere." *Phys. Rev.*, 35: 1098–1118 (1930).
17. —— "Night Sky Brightness Measurements in Latitudes below 45°." *J. opt. Soc. Amer.*, 39: 211–215 (1949).
18. Kuiper, G. P., ed., *The Atmospheres of the Earth and Planets.* Chicago, University of Chicago Press, 1949. (See Fig. 56, p. 161)
19. Meinel, A. B., "The Near Infrared Spectrum of the Night Sky and the Aurora." *Publ. astr. Soc. Pacif.*, 60: 373–377 (1948).
20. Penndorf, R., "Effects of the Ozone Shadow." *J. Meteor.*, 5: 152–160 (1948).
21. Rayleigh, Lord, and Jones, H. S., "The Light of the Night-Sky; Analysis of the Intensity Variations at Three Stations." *Proc. roy. Soc.*, (A) 151: 22–55 (1935).
22. Roach, F. E., and Barbier, D., "The Height of Upper Atmospheric Emissions." *Publ. astr. Soc. Pacif.*, 61: 89–91 (1949); "The Height of the Emission Layers in the Upper Atmosphere." *Trans. Amer. geophys. Un.*, 31: 7–12 (1950).
23. Rodionov, S. F., and Pavlova, E. N., "Ob infrakrasnom izluchenii nochnogo neba." *Doklady Akad. Nauk, SSSR*, 65: 831–834 (1949).
24. Russell, H. N., Dugan, R. S., and Stewart, J. Q., *Astronomy*, revised ed. Boston, Ginn, 1945. (See Vol. 1, p. 358); for recent references see Whipple, F. L., and Gossner, J. L., "An Upper Limit to the Electron Density near the Earth's Orbit." *Astrophys. J.*, 109:380–390 (1949).
25. Stebbins, J., Whitford, A. E., and Swings, P., "A Strong Infrared Radiation from Molecular Nitrogen in the Night Sky." *Astrophys. J.*, 101: 39–46 (1945).
26. Swings, P., "The Spectra of the Night Sky and the Aurora" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed. Chicago, University of Chicago Press, 1949. (See p. 178)
27. Vegard, L., and Tønsberg, E., "Investigations on the Auroral and Twilight Luminescence Including Temperature Measurements of the Ionosphere." *Geofys. Publ.*, Vol. 13, No. 1 (1940).

AURORAE AND MAGNETIC STORMS

By L. HARANG

Norwegian Defense Research Establishment

Introduction

In this survey of our present knowledge of aurorae, the first two sections which follow will be concerned mainly with the appearance, or morphology, of the auroral forms. The first section contains material obtained for the most part from observations at a single station, while in the second section the material (mainly height determinations) was obtained from two or more stations. The principal instrument used was the auroral camera. The third section gives a survey of spectral investigations, and the fourth section outlines the corpuscular theory of aurorae and magnetic storms. Although this theory must be regarded as only a first approximation to the real conditions, it is of the greatest value in discussing observations and in developing new points of view for further research. Two sections are devoted to a discussion of the connection between aurorae and magnetic storms, and a new field of research—the scattering of radio waves in the VHF-band from aurorae—is briefly mentioned. We can assume that new and important knowledge of the details of the auroral processes will emanate from the fields of research mentioned in these two sections. The final section lists some promising lines for further research.

Appearance of the Aurorae

Forms. The variety of auroral forms and the gradual shift from one form to another have made necessary a general classification, involving international cooperation. The atlas of auroral forms [10] published under the auspices of the International Union of Geodesy and Geophysics (Prague, 1927), is based mainly on the photographs and descriptions presented by Störmer. In this atlas the forms are divided into two main classes: forms without ray-structure and forms with ray-structure. An additional class, flaming aurorae, can be added.

I. Forms without Ray-Structure.

1. Homogeneous Quiet Arc (*HA*). This form may appear near the horizon, usually in the northern sky. It is diffuse along the upper border, sharp along the lower. This is the most stable of all auroral forms, and may maintain its shape for a period of time ranging from several minutes to half an hour. A double arc often appears, or several arcs may simultaneously be seen in the sky.

2. Homogeneous Bands (*HB*). These forms are irregular in shape and often move rapidly. As with *HA*, the upper border is diffuse while the lower is sharp. The homogeneous bands often change into bands with ray-structure.

3. Pulsating Arc (*PA*). This form is seldom visible. The arc or parts of an arc show more or less regular pulsations in luminosity. The period of pulsation ranges from about twenty seconds down to a few seconds. In some cases *PA* may disappear almost completely and then reappear, at approximately the same place, at regular time intervals.

4. Diffuse Luminous Surfaces (*DS*). These appear in the form of a diffuse veil over parts of the sky. They have no distinct limitations. These surfaces often appear *after* intense displays of rays and curtains. The appearance of *DS* is therefore usually restricted to the later hours of the night, after brilliant auroral displays.

5. Pulsating Surfaces (*PS*). These appear in the form of diffuse patches which come and go rythmically at the same place, retaining the same irregular form. They usually appear after a display of flaming aurorae.

6. Feeble Glow (*G*). This form resembles the dawn. It often appears as the upper part of an arc whose lower border is below the horizon.

II. Forms with Ray-Structure.

These consist of short or long rays which can be arranged in various ways.

1. Bands with Ray-Structure (*RB*). These resemble *HB*, but consist of a series of rays close to one another along the band. When a band is near the magnetic zenith it may have the form of a corona.

2. Draperies (*D*). If the rays along a band become longer, the form resembles a curtain or drapery, the lower border of which is often very luminous. With the exception of the homogeneous arc, the drapery is the most common auroral form. When it appears near the magnetic zenith, perspective may give *D* a fanlike form.

3. Rays (*R*). These forms may be isolated, or parallel in bundles. In the latter case they often resemble *D*. The vertical extension may vary considerably, ranging from short rays to the very long, sunlit rays.

4. Corona (*C*). When rays or draperies approach the magnetic zenith they appear to converge, because of the perspective, and a corona is formed. This corona is often incomplete, and only one-half of it may be developed. It can also be formed by bands which converge towards the magnetic zenith.

III. Flaming Aurorae (*F*).

This is a form consisting of strong waves of light which characteristically move rapidly one after the other in the direction of the magnetic zenith. The waves may take the form of detached arcs which move upwards towards the magnetic zenith. Flaming aurorae appear frequently after strong displays of rays and curtains and are often followed by the formation of a corona.

PLATE I.—Typical auroral forms: (a) Homogeneous Arc; (b) Band; (c) Arc with Ray-Structure; (d) Corona; (e) Drapery; (f) Rays. (*After C. Störmer.*)

Variations in Latitude. The most remarkable fact about the geographical distribution of the aurorae is that the greatest frequency, in both arctic and antarctic regions, occurs along a zone lying about 20°–25° from the earth's magnetic axis point. For the arctic region this was stated as early as 1881 by Fritz [4]; White and Geddes [19] have drawn up a similar zone of maximum frequency for the antarctic region (Fig. 1). However, the records are too few for a complete construction of the lines of equal frequencies over the antarctic region, and additional observations from conveniently distributed points would be of great interest.

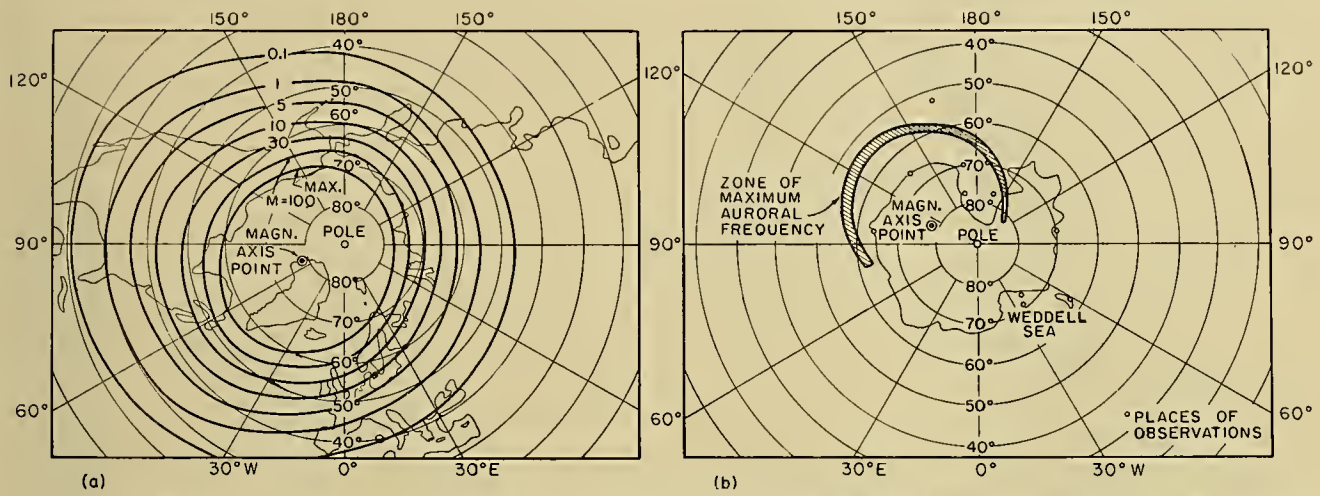


FIG. 1.—The geographical distribution and zone of maximum auroral frequencies in the north and south polar regions. (*After Fritz* [4], *and White and Geddes* [19].)

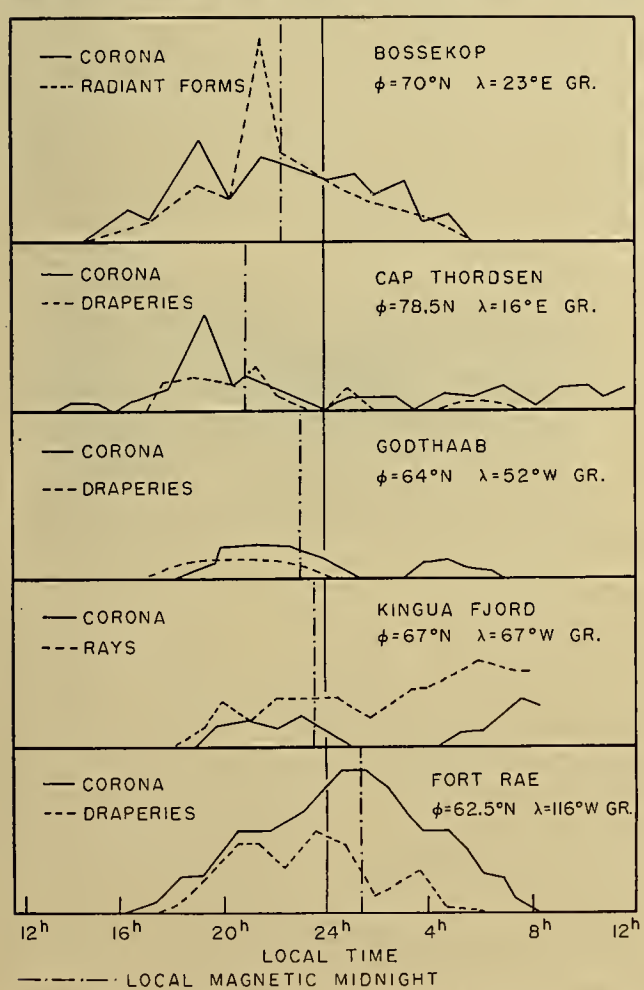


FIG. 2.—Diurnal variation of the aurorae, observed during the Polar Year 1882–83. (*After Vegard* [15].)

Variations in Time. The mean diurnal appearance of the aurorae seems to be correlated with *local magnetic time.* The maxima apparently occur about 1.3 hours before magnetic midnight (Fig. 2). (Magnetic midnight is here defined as the time at which the plane through the point of observation and the magnetic axis passes the sun.)

Position in Space of the Aurorae

The Method of Height Determination. Reliable height determinations can be made only by means of parallactic photography. If objectives of high light-power (about f 1:1.5 and less) and sensitive plates are used, the time of exposure for aurorae of medium intensity is even less than a second. Graphical methods, chiefly developed by Störmer [12],[1] are used for reduction of the plates. The base line should not be less than 20–30 km.

Direction of the Arcs and Position of the Radiation Point. The position and the direction of the arcs are closely connected with the direction of the earth's magnetic field at the place of observation. The position of the radiation points of coronae is similarly dependent on the inclination and declination of the earth's magnetic field at the place of observation.

Height Statistics of Aurorae. The positions in space of the auroral forms are usually characterized by the lower and upper limits of the luminosity. In the photographs of such forms as arcs, bands, and draperies, the lower limit is usually easy to determine, since a sudden decrease in the luminosity is apparent. The determination is more difficult with such forms as rays and pulsating surfaces. It is always difficult to determine the upper

1. Methods for the reduction of a pair of parallactic photographs are described in detail in [5]. (See also [16].)

limit precisely for all forms. In Fig. 3 the *lower* limits of all forms of polar aurorae are shown. The measurements are the result of three series of observations.

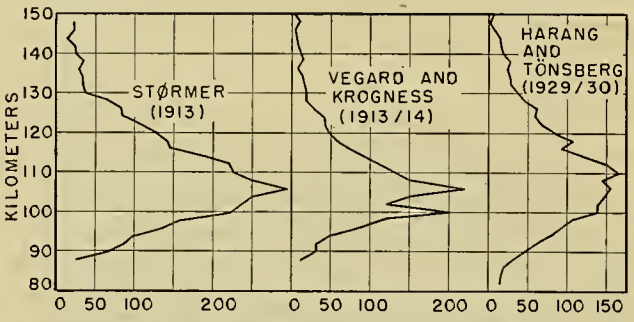


Fig. 3.—Height statistics. Lower limits of polar aurorae, all forms. (The number of points appearing in 2-km height intervals are used as abscissae.)

The lower limits lie between 80 and 140 km, with a pronounced maximum at about 106 km. Störmer has made an extensive study of the lower limits of different forms and has also compared the lower limits of the polar aurorae and aurorae appearing at more southerly latitudes. There appear to be only small systematic changes in the *lower* limits for the various types and latitudes.

There is a wide range of variation in the *vertical* extent of the luminosity. In polar regions the rays extend to heights up to 200–250 km. In lower latitudes the vertical extent may increase to as much as 800–1000 km. Aurorae appear only at these latitudes during very intense magnetic storms, and radio-echo measurements indicate that there is a considerable increase in the reflection heights of the ionosphere during the storm phase. The increases in vertical extent of the aurorae at lower latitudes may therefore be due partly to the special conditions of the ionosphere during strong magnetic storms.

Sunlit Aurorae. Height measurements in arctic regions near the auroral zone have shown that the rays and draperies in the *dark* atmosphere very seldom attain heights over 250–300 km. Störmer has shown, however, that faint rays, often of a grayish color, appear-

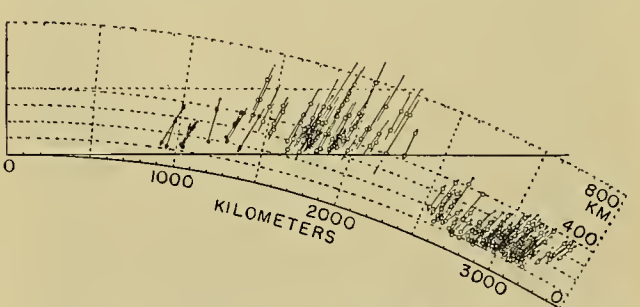


Fig. 4.—Aurorae situated in the border region between the dark and sunlit atmosphere. The horizontal line indicates the shadow line. (*After Störmer* [13].)

ing over southern Scandinavia at a considerable distance from the auroral zone, may attain heights up to 800–1000 km. Aurorae appearing at such great heights are always situated in the sunlit atmosphere and, at these low latitudes, appear only during great magnetic storms. Figure 4 shows how the lower limits concentrate along the shadow line between the dark and the sunlit atmosphere. Spectrographic observations by Störmer have shown that the nitrogen band in sunlit aurorae is strongly enhanced relative to the green auroral line.

Spectrum of the Aurorae

The intensity of the auroral luminosity is low, and if spectrographs are to be used, they must be of considerable light-gathering power. The spectrum consists of a number of lines and bands; more than one hundred are listed in the wave-length tables. They extend from 8100 A (infrared) to 3100 A (ultraviolet) where ozone absorption cuts off the spectrum.

Wave Lengths and Average Intensities of Spectral Lines. By exposing a spectrum over a number of nights, an average spectrum with intensities approximating the values given in Table I is obtained. Here only the *stronger* of the auroral lines, together with their identifications, are listed. The spectrum is dominated by the nitrogen band systems, the negative group ($N\ G$) and the first and second positive groups ($1\ P\ G$ and $2\ P\ G$), which are all well known from the study of gas discharges. The visible color of the aurorae, however, is produced by the atomic oxygen lines 5577 A in the green (Fig. 5) and the doublet 6363, 6300 A in the red. In addition to the stronger lines a considerable number of weak lines and bands have been listed, some of which seem to be atomic lines of O and N. Of special interest is the appearance of the forbidden nitrogen line $N\text{I}$ ($^4S - {}^2P$) at 3466 A.

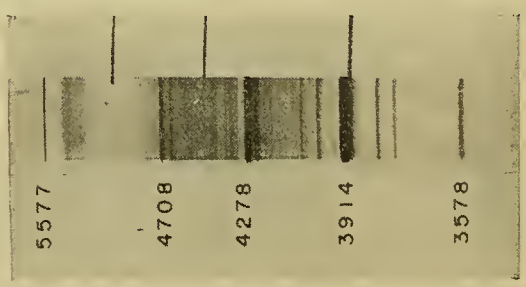


Fig. 5.—Spectrum of aurorae. (*After Vegard* [15].)

Intensity Effects. The study of intensity variations within the auroral spectrum is of great importance, but because of the rapid movements and low intensity it is difficult to carry through. Here one has to use small spectrographs of the highest light-power, permitting short exposures, and the measurements are therefore limited to the stronger lines. The following intensity measurements have been made: (1) intensity variations within a single auroral form from the lower edge to the upper limit; (2) intensity variations from one auroral form to another; and (3) latitudinal variations between polar aurorae and aurorae at lower latitudes.

1. Altitude Effects. In all auroral forms there is a distinct increase in the intensity of violet nitrogen bands relative to the green auroral line as one goes

TABLE I. AURORAL LINES AND BANDS
(*After Vegard* [15])

A	Intensity	Identification	
7906	47	1 *P G*	7–6*
7867		—	—
6619		—	—
6605	40	1 *P G*	6–3
6592		1 *P G*	13–11
6363	—	O_I	$(^1D_2 - {}^3P_1)$
6300.30	28	O_I	$(^1D_2 - {}^3P_2)$
5990.8	15	1 *P G*	15–12
5891	13	1 *P G*	9–5 $(Na: D_1D_2)$
5577.35	100	O_I	$(^1S_0 - {}^1D_2)$
5238	6	1 *P G*	16–11
4709	8	*N G*	0–2
4652	5	*N G*	1–3
4596	3	*N G*	2–4
4551	2	*N G*	3–5
4278	24	*N G*	0–1
4236	6	*N G*	1–2
4059	3	2 *P G*	0–3
3998	4	2 *P G*	1–4
3942	2	2 *P G*	2–5
3914	47	*N G*	0–0
3805	5	2 *P G*	0–2
3755	4	2 *P G*	1–3
3578	10	2 *P G*	0–1
3537	5	2 *P G*	1–2
3466	—	N_I	$(^4S - {}^2P)$
3371	9	2 *P G*	0–0
3159	6	2 *P G*	1–0
3126	4	2 *P G*	2–1

* Vibrational quanta numbers for the transition.

from the lower edge to the upper limit of the aurorae. There is also a distinct increase in the relative intensity of the red doublet at 6300 A compared to the green line 5577 A. At the same time the red nitrogen bands 1 *P G* at 6500 A are relatively stronger at the lower part of the border than at the upper limit, when compared with the green line.

2. Type Effects. Forms such as diffuse and pulsating surfaces, compared with strong and radiant forms such as draperies, bands, and arcs, show violet nitrogen bands of considerably greater intensity than the green line. In the sunlit aurorae, where the upper limits may attain heights of 800–1000 km, the violet region is also strongly enhanced relative to the green line, and the color may appear to be gray-violet.

The red coloring of aurorae is due either to an enhancement of the red oxygen line 6300 A or to the red nitrogen bands at 6500 A. The former effect occurs especially at lower latitudes; the red coloring at the lower edge of strong aurorae is usually due to the enhancement of the nitrogen bands. When the aurorae appear in uniform red-colored forms, especially at low latitudes, the coloring is due to the oxygen line. However, the red color of the lower edge of strong aurorae is due to the nitrogen bands.

3. Latitudinal Effects. These effects were summarized by Vegard in a comparison between 70°N (Tromsö) and 60°N (Oslo), as follows: (1) The intensity of 6300 A relative to the green line increases towards lower latitudes. This relation is even more pronounced when the green line is compared with the negative bands. (2) The intensity of the green line relative to the negative bands increases towards the lower latitudes.

Appearance of Hydrogen and Sodium Lines. Among the fainter lines in the auroral spectrum are the two Balmer lines, $H\alpha$ (6563 A) and $H\beta$ (4816 A), which may appear with varying intensities. Furthermore, the *Na*-line (5891 A) also appears. This line is strongly enhanced in the sunlit part of the night-sky spectrum.

Determination of Temperature from the Nitrogen Bands. According to the quantum theory, the width of the nitrogen bands depends on the temperature of the gas. By measuring the intensity distribution within the *R*-branch of a band, the apparent temperature of the gas can be calculated. In laboratory experiments these measurements are based on spectra taken with great dispersion. In auroral spectroscopy, the dispersion is limited, and the accuracy is accordingly limited. Quantitative measurements by Vegard based on the photometry of the nitrogen band 4278 A give a temperature of about −45C.

The Corpuscular Theory of the Aurorae

The coincidence between the appearance of magnetic storms, aurorae, and sunspot activity leads naturally to the assumption that the primary cause of both magnetic storms and aurorae must be a corpuscular radiation emitted by the sun. The fact that aurorae appear on the night side of the globe may be explained by the effect of the earth's magnetic field on an electrically charged stream of corpuscles. Different views have been expressed as to the nature of the particles emitted by the sun, and at the present time we have no definite evidence as to whether they are electrons, positive rays, or a mixture of both held together by electrostatic action.

Störmer [14] has treated extensively the case of the movement of an electrically charged particle approaching the earth's magnetic dipole field. The theory can be applied to both negative and positive charges, but it has been especially worked out for the assumption that corpuscles are fast electrons. Chapman [3] and later Alfvén [1] developed a corpuscular theory in which a cloud of negatively and positively charged particles leaves the sun at a comparatively moderate velocity. This ion cloud is polarized in the earth's magnetic field, and electrostatic fields are set up. The detailed analysis of the orbits of the particles is difficult because of the complexity of the problem.

The Model Experiments of Birkeland. A first approach to a corpuscular theory of aurorae and magnetic storms was made through the model experiments of Birkeland. A small uniformly magnetized sphere, the "Terrella," was placed in a vacuum and a stream of electrons was directed against the sphere. At low pressures the paths

of the electron stream can be made visible, and these model experiments gave a striking picture of the possible orbits for an incoming electron stream approaching the earth. These model experiments were later extended by Brüche and Malmfors (see [1]).

The Movement of a Charged Corpuscle in the Earth's Magnetic Field—Störmer's Calculations. The movement of an electron with a certain initial velocity is well known for a number of simple types of magnetic fields, such as movement in a homogeneous field parallel or transverse to the direction of the field. In a radial field which is produced by a single magnetic pole the orbits of the electrons will be like geodetic lines on a cone. It has not been possible to solve completely the more general case represented by the field of a dipole such as that of the earth. The differential equations for the movement of a particle in the dipole field are easily set down, but a general integration of these equations has not been possible. Störmer has discussed the equations in several papers and given many numerical solutions which he has applied to auroral problems. It may be added that the same problem appears in the theory of cosmic rays, where fast electrons from space enter the earth's magnetic field. Störmer's calculations explain several of the effects associated with the appearance of the aurorae. According to theory, the electrically charged particles will impinge on the earth along two zones, symmetrical with respect to the magnetic axis points, which represent the auroral zones. Further families of trajectories of the particles will end on the earth's night side, thus making the appearance of the aurorae on the night side possible.

Different views have been expressed concerning the motion of the electrically charged particles in an auroral form. In the case of a fine auroral ray it may be assumed that the earth's magnetic field can be regarded, to a first approximation, as equivalent to the field of a single pole. The motion should be along a geodetic line. When approaching the earth, the electrically charged particle will reach a certain minimum height, where the motion will be in a circle lying at right angles to the field. The radius ρ of this circle is a measure for the stiffness ($H\rho$) of the rays, given by the formula,

$$H\rho = \frac{mv}{e},$$

where H is the magnetic field intensity, and m and e are, respectively, the mass and the charge of a ray-corpuscle. Measurements of the width of fine auroral rays show that the stiffness ($H\rho$) may attain values of 10^5. This corresponds to stiffness of the order of β-rays or of fast positive rays in gas discharges. The measurements of $H\rho$ therefore do not give any definite information on the nature of the corpuscles producing the aurorae. But if, after entering the earth's atmosphere, the primary rays are spiraling down along the lines of the earth's magnetic field, the velocity must be considerable, and slow electrons or positive ions seem to be excluded. Störmer's mathematical theory treats only the highly idealized case, that is, the movement of a *single* electrically charged particle in the earth's field. In nature we have to consider the movement of a stream or cloud of charged particles, perhaps of both signs and of different masses and charges, leaving the sun and approaching the earth. During the passage, the cloud will induce changes due to electrostatic attraction or repulsion. In addition, upon entering the earth's magnetic field, the positive and negative charges will be displaced relative to each other and polarization effects will occur. Chapman [3] and, later, Alfvén [1] have extensively treated the passage of an ion cloud emitted by the sun and approaching the earth. Chapman has paid especial attention to the terrestrial effects appearing as magnetic storms, while Alfvén has discussed the auroral effects.

The Aurorae and Magnetic Storms

Physically, one must regard the aurorae and the polar magnetic storms as two effects with the same primary cause—an ionizing corpuscular radiation penetrating the atmosphere to a level of 80–100 km. Magnetic storms have been studied thoroughly and different types and phases within a storm have been classified. The appearance of the polar magnetic storm is of the greatest interest for auroral studies. This type of storm appears regularly and is mainly confined to the belt of the auroral zones. Physically it can be explained as the effect of a current sheet flowing along or parallel to the auroral zone at a height of 100–150 km. Birkeland [2] studied this storm type extensively, using synoptic charts. The records from the Polar Year 1932–33 gave a unique opportunity of studying this storm type in more detail than ever before. McNish [9], Vestine [11], Chapman [18], and others have made synoptic and statistical studies of the polar magnetic storm. Most interesting in this connection is the study by Vestine. He showed that the mean position of the current sheet coincided almost exactly with the position of the auroral zone, both lying at a distance of about 23° from the earth's magnetic axis point.

A most remarkable effect is the increase of the angular diameter of the auroral zone which coincides with increasing strength of magnetic storms. During great storms the aurorae are displaced towards the south; at places along the auroral zone the aurorae are then usually seen towards the south. Störmer has explained this widening of the auroral zone as due to the effect of a ring current appearing in the equatorial plane of the earth, but lying far outside the atmosphere.

The Aurorae and the Ionosphere

The normal polar aurorae appear at heights ranging from 80 to 300 km above the surface of the earth. At lower latitudes and during great storms auroral forms may appear at much greater heights. The ionosphere consists mainly of two regions of ionization, the E-layer at 120 km above the earth, and the F_2-layer at about 230 km. The normal ionosphere is produced by the ionizing effect of the sun's ultraviolet spectrum. Owing to the stratification in the air's composition at different levels, ionization maxima are produced when the sun's rays pass through the atmosphere, giving rise to

the normal stratification of the ionosphere—the E-, F_1-, and F_2-layers. In addition to the ultraviolet rays from the sun, corpuscular rays emitted by the sun may also have an ionizing effect when they penetrate the earth's atmosphere. The penetration of these corpuscular rays is made visible by the appearance of aurorae. A marked increase in the ionization of the layers during auroral displays might therefore be expected, and this is what actually happens. The effects of aurorae on the ionosphere are treated below.

The Appearance of the Abnormal E-Layer during Auroral Displays. We must here distinguish between the effects of *faint* and *strong* auroral displays. During a faint auroral display the electrically charged corpuscles will emit light visible as aurorae, and in addition they will produce ionization by impact which will increase the electron density of the layers, especially the E-layer. In polar regions there is a very close connection between the appearance of faint or medium aurorae, small magnetic storms, and a simultaneous increase in the electron density of the E-layer. This increase of the electron density of the layers can be followed by radio-echo observations which measure the critical penetration frequencies of the layers. From the critical penetration frequencies one can easily calculate the *maximum* electron densities of the layers. Figure 6 shows how the maximum electron densities of the F_2- and E-layers change at a polar station during a small magnetic storm accompanied by a faint aurora overhead.

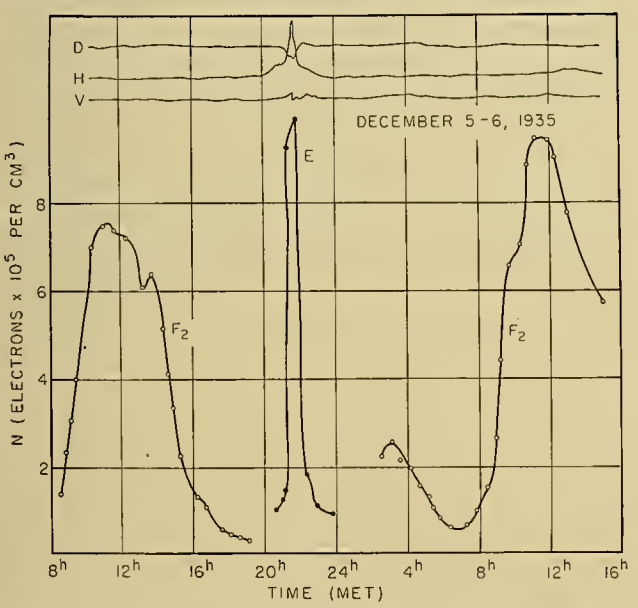


Fig. 6.—Appearance of abnormal E-layer during a small magnetic storm accompanied by faint aurorae (observed in Tromsö). Magnetic declination (*D*), horizontal intensity (*H*), and vertical intensity (*V*) of the magnetic field are shown above. (MET = Middle European Time.)

During stronger magnetic storms and aurorae the conditions are more complex, and the effect of absorption is of the greatest importance. During strong auroral displays, the radio echoes reflected from the ionosphere are usually weak, and there may be a complete cessation of radio echoes during the greatest displays. At the same time there may be a complete breakdown of medium- and short-wave commercial radio transmission. The explanation of these two effects—the increase in electron density in the E-region during small storms and faint aurorae, and the complete cessation of radio echoes during strong magnetic storms—is the following: During small storms the impact of the electrically charged particles increases the ionization of the E-layer down to a height of 100–110 km. This increase in ionization is measured by the increase in critical penetration frequencies of the layer. During strong magnetic storms and aurorae, this increase in ionization is displaced farther down in the atmosphere, to a height of about 80 km. Here the density of the air will be considerably greater than in the E-layer, and this will increase the collision frequency between the free electrons and the surrounding gas molecules. The increase in collision frequency will cause a strong absorption of radio waves in the medium- and short-wave bands. The irregular disturbances in radio communications show very close connections with the appearance of magnetic storms and aurorae, and it is possible, to a certain degree, to forecast the disturbances by carefully observing magnetic storms and aurorae.

The propagation of long and very long radio waves is only slightly influenced by irregular changes in the ionosphere. The long waves are propagated mainly as ground waves over short distances. Over longer distances there will be a fraction—about 10–20 per cent of the field strength at the receiving station—which is propagated as a reflected wave. For such long waves the ionosphere acts almost as a reflecting mirror. An increase in the electron density in the lowest part of the E-layer only increases the reflective power of the layer for these waves, and receiving conditions for very long radio waves are therefore usually more favorable during disturbed periods than during quiet periods.

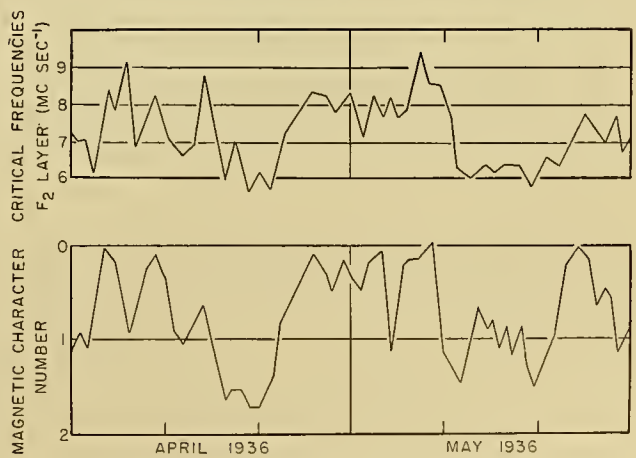


Fig. 7.—Critical frequencies of the F_2-layer and magnetic character numbers (observed in Tromsö). During periods of great magnetic activity the critical frequencies decrease.

The Effects on the F_2-Layer. The effect of magnetic storms and aurorae on the F_2-layer is more complex, since disturbances at this height (about 250 km) are accompanied by changes in the structure of the layer.

There is an inverse correlation between magnetic activity and critical penetration frequency, that is, the maximum electron density of the layer. During disturbed periods the maximum electron density decreases. This effect is commonly explained by assuming a vertical expansion of the ionosphere during disturbed periods. Figure 7 shows the inverse correlation between the critical frequencies of the F_2-layer and the magnetic character numbers, as recorded at a polar station.

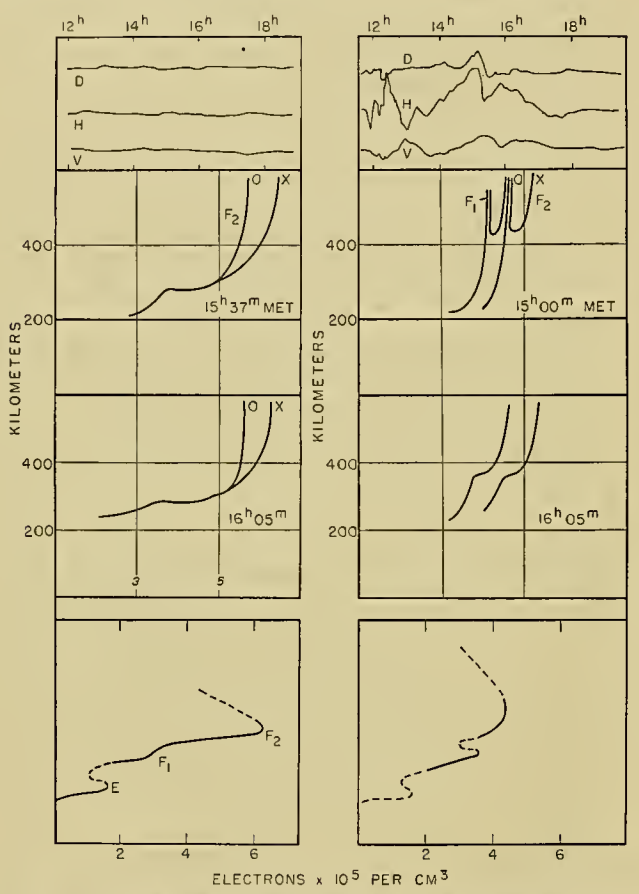


Fig. 8.—Changes in the structure of the ionosphere in the auroral zone from a quiet day to the following disturbed day (from records in Tromsö). The magnetic records are shown above. The echo curves and the structure of the ionosphere are given below.

The changes in the structure of the ionosphere during intense magnetic storms are evident from the echo records. A typical example indicating the conditions of the ionosphere on a quiet day and a following disturbed day at a polar station is given in Fig. 8. The structure of the ionosphere can be deduced from the character of the echo curves (virtual heights of echoes versus frequency). The expansion of the F_2-layer, and the more pronounced stratification between the F_2- and F_1-layers, are evident.

Scattered Reflections from Aurorae in the VHF-Band. It has already been mentioned that strong aurorae and magnetic activity are accompanied by an increase of ionization in the lower edge of the E-layer. This increase in ionization at a level at which the density is comparatively high produces a strong absorption of the radio waves in the usual high-frequency band 1–15 mc sec^{-1}, which is usually in operation for ionospheric radio-echo recording. During strong auroral displays this absorbing region usually screens off the higher part of the ionosphere, and the echo records give no information about the conditions of the ionosphere and auroral region during the phases of strong absorption.

The amount of absorption will, however, decrease with increasing frequency, and it thus becomes possible to use waves in the VHF-band for studying *scattering* effects from the dense ionization in the aurorae. Scatter from aurorae has been obtained by Harang and Stoffregen [6], using 41 mc sec^{-1} waves. Lovell, Clegg, and Ellyett [8] used frequencies of 72 and 46 mc sec^{-1} and obtained scattering effects. The use of VHF-waves for the study of scattering in the ionosphere must be regarded as a new and promising field of research for obtaining information during the phases of strong absorption.

Concluding Remarks

Auroral phenomena are the visual results of an action of a stream of solar particles on the upper layers of the earth's atmosphere. Magnetic storms and changes in the structure of the ionosphere occur simultaneously with the aurorae.

A quantitative theory of the aurorae, magnetic storms, and the irregular changes within the ionosphere would presuppose a knowledge of the nature of the solar particles and a detailed knowledge of the physical condition of the upper atmosphere.

In the preceding sections it has been pointed out that auroral phenomena give no definite indication of the nature of the solar particles producing the aurorae. Furthermore, the physical conditions of the upper atmosphere—density, pressure, temperature, chemical composition, vertical and horizontal displacement of the air masses, and the variations of these quantities during the day and the year—have mainly been deduced in indirect ways. The picture of the physical condition of the upper atmosphere has therefore been put together like a mosaic, with information supplied by the various branches of geophysics. If one of these questions—either the nature of the solar corpuscles or the physical state of the upper atmosphere—could be solved independently, an important step forward would have been taken in the theory of the aurora. Through V-2 rocket experiments, reliable values for densities and pressures up to heights of 90 km have now been obtained [7]. Further measurements will remove many uncertainties concerning the physical properties of the upper atmosphere. A direct experimental investigation of the nature of the solar particles should be possible by means of V-2 rocket flights through aurorae in polar regions.

The position of the auroral zone and the distribution of the auroral frequency with the latitude has hitherto been based on Fritz' chart, published in 1881 [4]. In the last few decades a vast amount of statistical information has been gathered which should pro-

vide possibilities for more reliable determinations of the auroral frequency curves.

A detailed morphological study of auroral phenomena has hitherto been carried on mainly over Scandinavia. It would be of the greatest interest to extend this detailed study of types of aurorae and their position in space to other places along the auroral zone. In this connection it would be of great interest to study the horizontal extension of a single quiet auroral form, such as a homogeneous arc or band, along the auroral zone. From a single station one can determine the extension of such a form up to 600–700 km, but it is still an unsolved question whether or not such a form may have even greater extension, or may even cover a greater part of the auroral zone. The question could perhaps be solved most conveniently by using a series of airplanes flying at fixed distances from each other along the auroral zone.

The spectrum of the aurorae shows an increasing number of lines with each new improvement in the spectrographic equipment used [17]. The most important new features of the spectra are the number of faint atomic lines from *N* and *O* which seem to be present. Further, the hydrogen lines $H\alpha$ and $H\beta$ show great changes in intensity together with a strong broadening of the line width, which indicate Doppler movements of the emitting atoms.

A new and promising field of research on ionization processes within the aurorae is the study of the scattering of radio waves from aurorae in the VHF-region. In this case, the ion clouds within an auroral display would act as scattering centers, and a detailed study would give information about the dimensions of these centers.

REFERENCES

1. Alfvén, H., *Cosmical Electrodynamics.* Oxford, Clarendon Press, 1950.
2. Birkeland, Kr., *Norwegian Aurora Polaris Expedition, 1902–03.* Christiania, Vol. 1, Part 1, 1908; Part 2, 1913.
3. Chapman, S., and Bartels, J., *Geomagnetism,* Vol. II. Oxford, Clarendon Press, 1940. (See pp. 850–890)
4. Fritz, H., *Das Polarlicht.* Leipzig, Brockhaus, 1881.
5. Harang, L., *The Aurorae.* London, Chapman (in press).
6. —— und Stoffregen, W., "Echoversuche auf Ultrakurzwellen." *Hochfrequenztech. u. Elektroakust.*, 55: 105–108 (1940).
7. Havens, R., Koll, R. T., and Lagow, H., *Pressures and Temperatures in the Earth's Upper Atmosphere.* Naval Res. Lab. Rep., Washington, D. C., March, 1950.
8. Lovell, A. C. B., Clegg, J. A., and Ellyett, C. D., "Radio Echoes from the Aurora Borealis." *Nature*, 160: 372 (1947).
9. McNish, A. G., "Heights of Electric Currents near the Auroral Zone." *Terr. Magn. atmos. Elect.*, 43: 67–75 (1938).
10. *Photographic Atlas of Auroral Forms,* 24 pp. Oslo, International Union of Geodesy and Geophysics, 1930.
11. Silsbee, H. B., and Vestine, E. H., "Geomagnetic Bays, Their Frequency and Current-Systems." *Terr. Magn. atmos. Elect.*, 47: 195–208 (1942). (See also Vestine, E. H., "On the Analysis of Surface Magnetic Fields by Integrals." *Ibid.*, 46: 27–41 (1941).)
12. Störmer, C., "Résultats des mesures photogrammétriques des aurores boréales observées dans la Norvège méridionale de 1911 à 1922." *Geofys. Publ.*, Vol. 4, No. 7 (1926).
13. —— "Sonnenbelichtete Nordlichtstrahlen." *Z. Geophys.*, 5: 177–194 (1929).
14. —— "Über die Probleme des Polarlichtes." *Beitr. Geophys.*, Supp. 1: 1–84 (1931).
15. Vegard, L., "Die Deutung der Nordlichterscheinungen und die Struktur der Ionosphäre." *Ergebn. exakt. Naturw.*, 17: 229–281 (1939).
16. —— "The Aurora Polaris and the Upper Atmosphere," Chap. XI, in *Terrestrial Magnetism and Electricity,* J. A. Fleming, ed. New York, McGraw, 1939.
17. —— *New Important Facts Relating to the Composition and State of the Ionosphere Derived from Auroral Studies.* Report to the Sec. Meeting of the Mixed Commission on the Ionosphere, URSF, Zürich, 1950.
18. Vestine, E. H., and Chapman, S., "The Electric Current System of Geomagnetic Disturbance." *Terr. Magn. atmos. Elect.*, 43: 351–382 (1938).
19. White, F. W. G., and Geddes, M., "The Antarctic Zone of Maximum Auroral Frequency." *Terr. Magn. atmos. Elect.*, 44: 367–377 (1939).

METEORS AS PROBES OF THE UPPER ATMOSPHERE

By FRED L. WHIPPLE

Harvard College Observatory

INTRODUCTION AND ASTRONOMICAL BACKGROUND

In this short discussion of the use of meteors in upper-atmospheric research no attempt is made to achieve historical completeness. For an interesting historical account of meteors and the basic theory of the earth's attraction, the reader is referred to Olivier's book on the subject [61]. Watson's *Between the Planets* [85] covers in semipopular style the related subjects of comets, meteors, meteorites, and minor planets. Pioneering work on meteoric and atmospheric theory such as that by Schiaparelli [76] in 1871 is discussed by Kopff [37] in his section on meteors in the *Handbuch der Astrophysik*. Hoffmeister [30] has summarized much of the other material in his book on the subject. A recent account, concentrating mostly on the early work by Lindemann and Dobson, is given by Mitra [59] in his valuable volume *The Upper Atmosphere*.

The reality of stones falling from space was not generally recognized in the scientific world until the first decade of the nineteenth century. Previously, the French Academy had remained particularly obdurate in denying the phenomenon, even though Halley had been convinced of the general idea in 1686 and the concept was very commonly accepted in ancient times.

Once the fact is accepted that meteors arise from an interaction between the atmosphere and bodies falling from space, certain limits on the phenomenon can be set immediately. The lower limit of the velocity near an altitude of 100 km is about 11.1 km sec^{-1}, the velocity of fall from rest at infinity. The earth itself is moving in a nearly circular orbit about the sun with a mean velocity of 29.7 km sec^{-1}, while the velocity of escape from the sun at the earth's distance is about 42.1 km sec^{-1}. Hence, the maximum velocity of encounter is nearly 73 km sec^{-1}, representing the head-on case and allowing for the earth's attraction [61].

The average velocity rises statistically from 6:00 P.M. to 6:00 A.M. as the observer is turned from the following hemisphere of the earth to the leading hemisphere. Meteors of maximum velocity cannot be observed before midnight.

There is still no proof from photographic or radio meteor studies that *any* meteoric bodies originate outside the solar system. The Harvard two-station photographic studies of some sixty brighter meteors give vectorial velocities with a precision of about one per cent. The radio-doppler studies of some 3000 meteors by McKinley [52] provide scalar velocities of somewhat less accuracy, but to a brightness limit well below the capacity of the naked eye. Lovell[1] has observed scalar velocities of more than eighty radio meteors coming from directions including specifically the apex of the earth's motion. None of these studies yield a statistically decisive excess of velocities above the solar-system limit within the visual magnitude range from −4 to possibly +7 or +8.

To clarify terminology, the term *meteor* will be used here to indicate the meteoric phenomenon and the term *meteoroid* to indicate the material body producing a meteor. Meteors sufficiently bright to cast shadows are called *fireballs*, while detonating fireballs are *bolides*. In case a meteoric body is sufficiently large for part of it to reach the earth's surface as a solid, the resultant body is a *meteorite*. These distinctions are necessary because of the rapidly accumulating evidence that the meteoroids producing the common meteors observed visually, photographically, and by radio techniques have a different origin than the meteorites. Hence, there is no justification for assuming that the chemical and physical structures of the meteoroids commonly seen as meteors are exemplified by the meteorites.

A large percentage of meteors are observed in *showers*, repeated with variable intensity at the same dates from year to year. Since the meteoroids in a shower strike the earth in parallel paths relative to the observer, the resultant meteors appear to radiate from a *radiant* on the sky. The various meteoric *streams* or showers are usually named for the constellations in which the radiants are located. Schiaparelli first showed that the Perseid shower, observable for more than half the month of August, is associated with Comet 1862 III. Since then a number of the major meteor showers have been associated with comets [85] and there is every reason to believe that essentially all shower meteors arise from comets.

The remaining meteors not associated with known showers are called *sporadic* meteors. The Harvard photographic meteor studies show that the major percentage of the bright sporadic meteors move in very elongated and randomly oriented cometlike orbits about the sun before striking the earth. The remainder move in orbits of low eccentricity, small major axis, and low inclination to the fundamental plane of the planets. These orbits are similar to those of the few minor planets or asteroids that approach the sun within the earth's distance, but also similar to the orbits of short-period comets. Wylie [95] has shown that several fireballs followed similar orbits. It is not possible to state definitely whether these sporadic meteors are of cometary, of asteroidal, or of other origin.

The recent studies of meteorites by Brown and Patterson [9] and Bauer [6], on the other hand, indicate strongly that meteorites represent debris from a broken planet or planets. The writer doubts seriously that comets can have such an origin. Hence, the commonly

1. Private communication.

observed meteoroids are probably of a different structure than meteorites.

The studies of meteor spectra, largely by Millman [55, 56, 57], indicate predominance of low-excitation *Fe*I lines, with *Fe*II lines [85] possibly showing in one case, probably for a very high-velocity meteor. Present also are lines of the atoms *Na*I, *Ca*I, *Mg*I, *Mn*I, *Cr*I, *Si*I, *Ni*I, *Al*I, *Ca*II, *Mg*II, *Si*II, and possibly *Fe*O. No atmospheric constituents have been observed. The relatively low states of excitation correspond to temperatures of only a few thousand degrees, in spite of the high energies of the impinging atoms; for example, a nitrogen atom at a velocity of 70 km sec^{-1} carries an energy of 410 ev.

The average meteoroid is probably an irregularly shaped stone or stony iron. There is no reason to believe that it is structurally strong or homogeneous in physical or chemical structure. Roughly 3 per cent of the meteors photographed by Harvard split into two or more pieces in the upper atmosphere, while nearly 10 per cent showed *flares* in brightness. Jacchia [35] has shown definitely that flares arise from quick losses of material, probably by crumbling or breakage of the meteoroid. Direct evidence as to the nature of meteoroids may sometime be obtained from micrometeorites [38, 92], those meteoroids that are sufficiently small to be stopped by the atmosphere without vaporizing. Until then we can only postulate the structure of meteoroids and check our postulates indirectly by various observations and deductions. An acceptable theory as to the nature and origin of comets would also be of value in the study of meteors.

ATMOSPHERIC RESEARCH BASED ON VISUAL OBSERVATIONS

Visual Observations of Meteors. As early as 1798 Brandis and Benzenberg set out to determine the heights of meteors by simultaneous observations from two separated stations. In all they observed 402 meteors, of which 22 appeared to be identical. Newton [60], a major contributor to early meteor studies, analyzed 21 of these "pairs" and calculated heights, ranging from 12 to 245 km, with a mean of about 100 km. The mean is in good agreement with modern measures but the range is far too great. Average visual meteors become observable at a height of about 100–110 km with a large scatter, and persist to 90 km or lower, depending upon their brightness. High-speed photographic meteors first show at greater heights, about 120 km, while the chief activity observed by electronic techniques is generally near the E-layer. The slowest and brightest photographic meteors observed at Harvard disappear in the neighborhood of 40 km.

Lindemann and Dobson [41, 43] made the pioneer application of meteoric theory and observation to a study of the density and temperature of the upper atmosphere. They developed the first comprehensive theory of the meteoric process and then utilized the extensive visual observations of meteor heights that had been made by that most assiduous of meteor observers, W. F. Denning, and by his co-workers. Since Mitra [59] has recently reproduced *in extenso* the theoretical developments by Lindemann and Dobson, only a few remarks concerning their work will be made in this article. Briefly, Lindemann and Dobson recognized and stated clearly the fundamental meteoric processes, namely, surface heating by impact with air molecules, stressing the importance of effective mean free path, vaporization of the meteoroid, luminosity produced by encounter of the vaporized material with the air, and the relatively slow deceleration of the nucleus remaining.

Although their thermodynamic, rather than kinetic, approach to the problem of heat transfer through the gas cap led to erroneously high values of the calculated air densities at great altitudes, as pointed out by Sparrow [81] and later workers, nevertheless the basic processes as visualized by Lindemann and Dobson are in other respects fundamental. Progress towards a completely satisfactory theory is still surprisingly slow.

Lindemann and Dobson showed clearly that the previous concept of a constant stratospheric temperature in the upper atmosphere must be replaced by the recognition of higher temperatures at great altitudes. They noted further that from Denning's observations the end heights of meteors were markedly less frequent in the region from 50 to 60 km than immediately above or below. From this fact they concluded that the atmospheric temperature must rise abruptly at a height of about 60 km so that "As the meteor passes into colder air the temperature of the air cap will fall so that the heating will be reduced." This latter argument and consequent conclusion must be revised slightly. It is more accurate to conclude that a region of maximum temperature must be present between 50 and 60 km because the existence of a smaller logarithmic density gradient prolongs the lifetime of a meteoroid reaching this height; the meteoroid has, therefore, a smaller chance of disappearing in the critical range of altitude. The air temperature, per se, is not the dominating factor; it is the consequent reduced density gradient which is important.

Related effects of a variable logarithmic gradient in atmospheric density show markedly in meteor trails photographed at Harvard and account largely for the phenomenon observed by Hoffleit [29] (see also [23]) that the point of maximum brightness moves systematically forward in meteor trails with increasing velocity (for discussion see [87]).

Lindemann and Dobson also pointed out the fact that meteor heights are systematically greater during the summer than during the winter. They could not conclude positively, however, that the height of a point of given density in the atmosphere is greater in summer than in winter. It is possible that there are systematic differences in velocity between the seasons and that the heights of meteors are very dependent upon velocity. These general problems concerning visually observed meteors still constitute a source of considerable discussion or disagreement [44, 51, 72, 73].

The Arizona Expedition for the Study of Meteors was planned by Shapley, Öpik, and Boothroyd [77].

The goal of the expedition was to obtain measures of meteor velocities, heights, radiants, and statistical data visually by the use of the "rocking mirror" technique. In applying this technique the observer determines the angular velocity of a meteor by looking at its reflection in a plane mirror, the mirror being made to oscillate in such a fashion that a normal to its surface describes a conical surface of small apex angle, with the apex near the center of the mirror. Trajectories and heights were measured by simultaneous visual observations from two stations separated by 36 km.

The active direction of the expedition and the analysis of the results were conducted by Öpik. From October 1931 to the end of July 1933, some 22,000 individual meteors were observed, and heights were determined for 3540. Most of the latter were sporadic meteors (about 80 per cent) and only 7 per cent came from the major showers. In analyzing the heights, Öpik [65, 66] made statistical corrections for all effects that he thought might possibly introduce systematic errors into a seasonal phenomenon, if present. He found that an average meteor corrected to standard conditions of brightness, velocity (actually elongation of radiant from apex of earth's motion) etc., appears at a greater altitude in summer than in winter. From this he concluded that the data "suggest an annual fluctuation of the height of the atmosphere, more or less corresponding to the annual temperature curve, of an (total) amplitude of 3.7 $\pm$ 0.7 km," applying at a mean altitude of about 88 km above sea level.

From his theory [67, 69] of the meteoric process he concluded that an atmospheric temperature of +100C at an altitude of 90 km and a mean molecular weight as at sea level would be consistent with the meteor observations by the Arizona Expedition.

Öpik's further conclusions as to a "night effect," in which the height of the "normal" meteor decreases during the course of the night, has not yet been confirmed or disproved. His conclusion [68] as to the appreciable percentage of "hyperbolic" velocities among visual meteors appears not to be consistent with the radio observations of meteors mentioned earlier.

Long-Enduring Meteor Trains and Winds in the Upper Atmosphere. The most comprehensive study of atmospheric winds and turbulence as reflected in the motions of long-enduring meteor trains has been made by Olivier [62, 63], who gives data for nearly 1500 trains. Of great importance also is the work by Fedynsky [20], who reports on 41 night trains observed systematically for the purpose.

Olivier finds from more than fifty cases that the average beginning height of night trains is 104 km, with average end heights of 80 km, while the corresponding data for twilight trains are 77 and 45 km respectively (twenty-six cases) and for daylight trains, 45 and 27 km (nine and fourteen cases). He calculates from the observed motions that the average wind velocity is 203 km hr^{-1} for the fifty best-observed night trains, 182 km hr^{-1} for night trains whose heights have been assumed, and 173 km hr^{-1} for day or twilight trains. A more detailed analysis suggests that the winds increase with altitude, that is, from daylight to night trains. It is clear, from both photographs and drawings, that "layers of the atmosphere, lying quite close together, have winds blowing in different directions and at quite different velocities." In very brief intervals of time after the appearance of trains "the photographs show that the trains have become zigzag in shape. . . . There seem to be inescapable proofs, in certain cases, of vertical components."

As for the directions of the train motions, Olivier finds that for America the greater number drift north with strong east-west components. See Fig. 1. For all

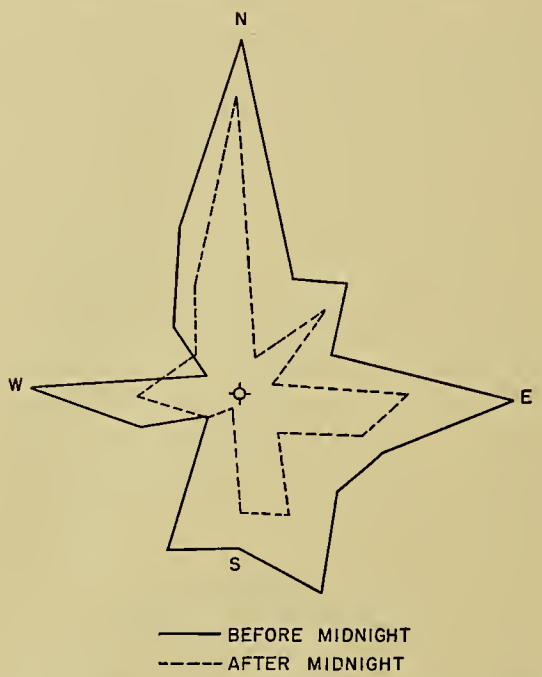


Fig. 1.—Frequency of night wind vectors over America.

Europe an eastern tendency is strong, while other areas of the earth show various preferential drifts. Unfortunately, Olivier has not searched for seasonal effects —a search that almost certainly must lead to important conclusions concerning upper-atmospheric circulation.

The results given by Fedynsky include a number of unusually high velocities, of the order of 1000 km hr^{-1}, and it is difficult to appraise the accuracy of the observations. His average deduced velocity is 391 km hr^{-1}. Additional confirmation would be desirable before accepting completely his separation of the winds into two groups of low and high velocity.

PHOTOGRAPHIC METEOR STUDIES

Methodology. Since photographic meteors have been used as a tool for research on the upper atmosphere almost solely at the Harvard Observatory, the discussion of this section will be confined largely to a short résumé of that work.

Systematic photography of meteors by the use of two cameras equipped with rotating shutters was begun by Elkin [16, 17] at the Yale Observatory in 1893 and continued until 1909. The astronomical results of this

work were analyzed by Olivier [64]. Unfortunately, the base line used (approx. 3.5 km) was inadequate for precision results, and in addition the original photographs have been destroyed. Lindemann and Dobson [42] used this method to a limited extent. Whitney [94] experimented successfully with electronically synchronized shutters for paired meteor cameras, but obtained few observational results. Fedynsky and Stanjukovitsch [21] measured the deceleration of a meteor photographically.

In the Harvard two-station photography of meteors the base line has been 37.9 km, between the Cambridge and Oak Ridge stations in Massachusetts [88, 90]. The two Ross Xpress lenses of aperture 1.5 inches, focal ratio F/4, were occulted 20 times per second by single-bladed shutters powered with synchronous motors from commercial power lines. Meteor trails show 5 to 80 breaks spaced uniformly in time. Exposure times were normally 1^h or 2^h, the cameras being driven on polar axes to follow the stars. The camera settings were so chosen as to produce an overlap in areas at meteor altitudes.

Analysis of the photographic meteor trails is simplified because the trails are amazingly straight [3, 22, 78, 89], that is, great circles; a change in direction of 20′ along a trail is rare indeed, occurring probably in less than one per cent of the (1700) trails in the Harvard plate collection, except in a few cases, very near the end. Hence the measurement and reduction of two-station trails are straightforward though tedious. The general method has been described by the writer [88] although the detailed formulas have not been published. In the most common case the data consist of the trajectory referred to sea level, the velocity V and deceleration $-V'$ at a known altitude, and the light curve. Longer and brighter trails often provide deceleration measures at several points, although the deceleration is barely measurable in most cases. The precision of the trajectory is extremely high (within a few meters) when the instant of the meteor is observed; the velocity is determinate to better than one per cent on the average; the deceleration may be well or poorly determined, while the uncertainty in the brightness ranges perhaps from 20 to 40 per cent.

Because the available wide-angle lenses can photograph only very bright meteors (-1 to -2 visual magnitude or brighter) the rate of photography is slow, one meteor in perhaps 50 to 100 hours of total exposure [86]. Hence, in ten years of continuous photography during clear moonless nights, only about sixty meteors were doubly photographed in the Harvard program. However, a program sponsored by the U. S. Naval Ordnance Department is under way to photograph fainter and more frequent meteors [91].

The Theory of Photographic Meteors. Theories of the meteoric process have been presented by Lindemann and Dobson [41], Hoffmeister [31, 32], Sparrow [81], Maris [50], Öpik [67, 69], Hoppe [33], Levin [39, 40], and others. Of these investigators Öpik has delved most deeply into the detailed physical processes involved, particularly that of the production of light. The writer [88, 90] has used physical data and concepts based largely on Öpik's work but has combined them in a mathematical framework modeled on that of J. Hoppe. The present account will present a very abbreviated version of this composite theory, influenced by the simplifications introduced by Jacchia [34] for the case where the meteor trail is completely observed. A criticism of the basic assumptions follows the formal presentation of the theory.

In its passage through air of density ρ at a velocity V, a meteoroid of mass m and effective cross section $Am^{2/3}$ will, in time dt, encounter an air mass

$$dm_a = Am^{2/3} \rho V dt. \tag{1}$$

We may assume from the kinetic viewpoint, at the high velocities involved, that in large measure the atmospheric molecules in the direct path of the meteoroid are trapped by the meteoroid or the gases escaping from its surface, momentarily carried along, and then swept backward at a relatively low velocity with respect to the meteoroid. It can be shown that for the photographic meteors the effective mean free paths of the air molecules (reduced, as shown by Lindemann and Dobson [41], by the ratio of the temperature velocity to the meteoric velocity) are generally smaller than the dimension of the meteoroid, except at extremely high velocities or faint magnitudes. Hence, in most cases there is a concentration of air (and meteoric gas) in front of the meteoroid in excess of that which might be expected from the relatively slow escape velocity of the molecules. Some sort of shock wave must form, but the significance of the shock-wave concept has not yet been useful in the theory.

The acceleration of the meteoroid may be written in the form:

$$V' = \frac{dV}{dt} = -\gamma \frac{V}{m} \frac{dm_a}{dt} = -\gamma A m^{-1/3} \rho V^2, \tag{2}$$

where γ is one-half the usual drag coefficient. Equation (2) expresses the conservation of momentum if $\gamma = 1$ and if the trapped air molecules are assumed to escape with negligible relative velocity.

The rate of mass lost by the meteoroid is undoubtedly a complex function of ρ, V, m, A and the physical and chemical structure of the body. Öpik [67] has shown that a meteoroid of pure iron would probably lose a considerable fraction of its mass as liquid because of its high thermal conductivity. On the other hand, the writer questions (as did Öpik) that a large fraction of the photographic or visual meteors arise from iron meteoroids. More likely the mass is a heterogeneous or conglomerate solid of low physical strength. Hence its thermal conductivity is probably low. In this case the mass loss would be directly proportional to the heat transferred to the surface and inversely proportional to the latent heat of vaporization ζ at approximately room temperature.

We may expect that the heat available for vaporizing the meteoroid will be roughly proportional to the energy released by the trapped air mass, dm_a. If the efficiency

of the heat transfer to the surface is given by λ, then the rate of mass loss will be:

$$\frac{dm}{dt} = -\frac{V^2}{2}\frac{\lambda}{\zeta}\frac{dm_a}{dt} = -\frac{\lambda}{2}Am^{2/3}\rho V^3 . \qquad (3)$$

The problem still remains to determine the mass m of the meteoroid. Equation (2) will lead to such a solution when velocity, deceleration, density shape-factor (A) and air density are known. But to determine the air density from meteor observations, we must evaluate m independently. This can be done by equation (3) if we assume that the luminosity of the meteor is proportional to the kinetic energy of the mass lost and if we can evaluate the proportionality factor. The greatest success in this evaluation has been attained by Öpik. For photographic meteors of velocity greater than about 20 km sec^{-1}; Öpik's calculations [69] are reasonably well approximated by the relation:

$$I = \tau_0 V\left(\frac{1}{2}V^2\frac{dm}{dt}\right), \qquad (4)$$

where I is the intensity of the visual radiation and $\tau_0 V$ is the luminous efficiency, τ_0 being a constant.

The rate of mass loss, from equation (4), becomes:

$$\frac{dm}{dt} = \frac{2I}{\tau_0 V^3}. \qquad (5)$$

The mass at any instant t_0 is then obtained from the integral:

$$m = \frac{2}{\tau_0}\int_{t_0}^{\infty}(I/V^3)\,dt. \qquad (6)$$

From a completely observed meteor trail in which the velocity V_0 and deceleration $-V_0'$ are well determined at time t_0 the air density ρ_0 can be determined by equations (6) and (2) in the form:

$$\rho_0 = -\frac{2^{1/3}V_0'}{\gamma A\tau_0^{1/3}V^2}\left[\int_{t_0}^{\infty}(I/V^3)\,dt\right]^{1/3}. \qquad (7)$$

This form of the solution for ρ has been used by Jacchia [34], [in equation (10), his $K = 2^{1/3}/(\gamma A\tau_0^{1/3})$] in recent Harvard results in the Naval Bureau of Ordnance program, as contrasted to the writer's direct use [90] of equations (3) with (5) and (2) for the less numerous earlier meteor observations. In practice the numerical constants γ, A and τ_0 of equation (7) may be evaluated separately from theoretical considerations, or as a single constant by comparison with atmospheric densities at the 50- to 60-km level where the densities are fairly well known. The agreement of the two methods is surprisingly good when Öpik's luminosity factor is used, γ is taken as near unity, and a reasonable value of A is chosen.

Jacchia [35] confirms Öpik's calculation of a decrease in τ_0 for velocities below about 20 km sec^{-1} by studies of bright slow meteors in the 40- to 50-km zone. At greater altitudes an attempt [34] to minimize the relative residuals in log ρ for the various meteors by varying the exponent of velocity in equation (7) indicates that a little improvement in the results can be obtained if the exponent of V in the denominator is reduced by 0.5 to 0.8.

We must conclude generally from Jacchia's work that the largest systematic error (excepting a constant multiplying factor) in atmospheric densities as calculated from meteoric decelerations from equation (7) rests in the assumed constancy of γ with respect to both velocity and density. At extremely great heights (> 100 km) where we deal with fast, small meteoroids the value of the drag coefficient may possibly in some cases exceed 2 (*i.e.*, $\gamma > 1$). Where mean free paths are relatively large the re-emitted air material will increase the drag, as shown by Tsien [82] and Heineman [24]. Furthermore, the vaporized meteoric material will cause an even greater increase by a similar process. As the height decreases, meteors of similar luminosity represent larger masses moving at lower velocities in air of greater density. Hence one should expect the drag coefficient to decrease. Consequently, the calculated values of air density from equation (7) may be somewhat overestimated at great heights if correctly evaluated at lesser heights. This drag problem is under investigation.

Less precise measures of atmospheric density can be determined from the luminosities early in the trails, while atmospheric pressure can be determined at the end points. Jacchia [36] finds, as did the writer [90], that the results obtained by these other methods are consistent with the more accurate results from the deceleration data when analyzed by the simple theory. Also, the light curves are accurately predicted except for flares. From the theoretical viewpoint it is interesting to note that the other methods involve λ/ζ, and that the ratio $\lambda/(\gamma\zeta)$ can be found by intercomparison of results. Jacchia [35] shows that this ratio requires no correction in the velocity exponent, suggesting that λ, the heat transfer coefficient, and γ, the drag coefficient, are truly constants or else show a similar velocity dependence. This conclusion is not surprising, at least from a qualitative point of view, but complete theoretical determinations of λ and γ are essential to an entirely satisfactory theory of the meteoric process.

Atmospheric Density Distribution and Seasonal Variations. As shown in the preceding section, the major uncertainty in the determination of atmospheric densities by photographic meteor observations rests in a constant multiplying factor. The density curve for meteors is, therefore, anchored to that obtained by other means in the 50- to 60-km zone. It is found that the gradient follows closely the gradient of the Tentative Standard Atmosphere of the National Advisory Committee for Aeronautics [84]. This result is not surprising in view of the fact that the N.A.C.A. atmosphere at great heights was based, to a considerable extent, upon results obtained by meteor photography.

The atmospheric densities found from the decelerations of thirty-five well-observed photographic meteors were compared by Whipple, Jacchia, and Kopal [93] with the densities of the N.A.C.A. atmosphere. The residuals show a strong correlation with season. In

Fig. 2 these residuals $\Delta \log_{10} \rho$ are plotted against the mean average ground temperature at Boston (lat. 42.5°N) at the date of each meteor. Three velocity

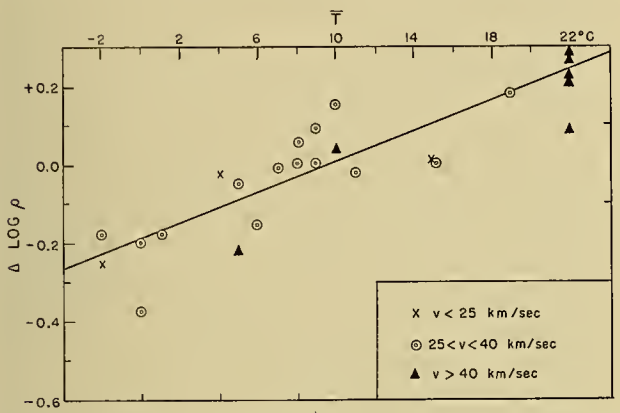


Fig. 2.—Seasonal density variations from photographic meteors.

groups among the meteors exhibit the same correlation. A least-square solution shows that $\log_{10} \rho$ increases 0.019 ± 0.0013 (P.E.) per degree centigrade of ground temperature. This result applies at a mean height of 78 km. The total seasonal range would be 0.46 in $\log_{10} \rho$, corresponding to a height variation of 8.6 km.

The seasonal correlation is less marked when the solar declination is used in place of the ground temperature. The correlation is not improved by comparison with the actual ground temperature at the date rather than the general average; hence, the correlation is truly a seasonal one which does not measure local variations. No effects associated with synoptic weather fronts, deviant temperatures in the lower stratosphere, sunspot numbers, lunar-hour angle or solar-hour angle are conspicuous. From further meteor investigations, including also atmospheric data from the beginning and end points of photographic meteors, Jacchia [36] finds evidence that the seasonal effect decreases with increasing height, becoming small and uncertain around the 100-km level.

Combining the photographic meteor data from decelerations and beginning-point data, Jacchia [36] has derived the upper-atmospheric density distribution given in Table I. Heights are given in kilometers above sea level and density in grams per cubic centimeter. The deviations of $\log_{10} \rho$ from the N.A.C.A. Tentative Atmosphere are given in the third column; the N.A.C.A. values have been adjusted slightly to avoid discontinuities in the temperature gradient, which are awkward physically. The values of atmospheric temperature in Table I are calculated with a constant molecular weight but include the decrease of gravity with height. The temperatures may be varied greatly within the range of solution, particularly the minimum values near 83 km and the values above 100 km.

The observational basis of Table I is shown in Fig. 3. The deceleration data are considerably more reliable than the beginning-point data. The results given in Fig. 3 and Table I have been derived purely from meteoric data without reference to other methods except for the zero point in $\log_{10} \rho$. The simple theory has been used, the power of the velocity having been varied to produce the best interval agreement among the meteoric data.

Table I. Adopted Density Profile

H (km)	$\log \rho$	$\Delta \log \rho$ (Derived − N.A.C.A.)	T (deg K)
70	−6.801	0.000	281
72	−6.883	0.001	266
74	−6.971	0.001	252
76	−7.067	−0.001	240
78	−7.173	−0.004	231
80	−7.287	−0.006	224
82	−7.412	−0.009	221
84	−7.543	−0.012	221
86	−7.679	−0.022	224
88	−7.816	−0.033	229
90	−7.951	−0.044	235
92	−8.083	−0.057	241
94	−8.211	−0.070	246
96	−8.334	−0.080	250
98	−8.453	−0.090	253
100	−8.569	−0.099	255
102	−8.683	−0.107	256
104	−8.795	−0.115	256
106	−8.908	−0.124	256
108	−9.019	−0.132	256
110	−9.131	−0.141	256

The determination of atmospheric densities and seasonal effects from the photographic observations of meteors can be improved by at least four methods of approach:

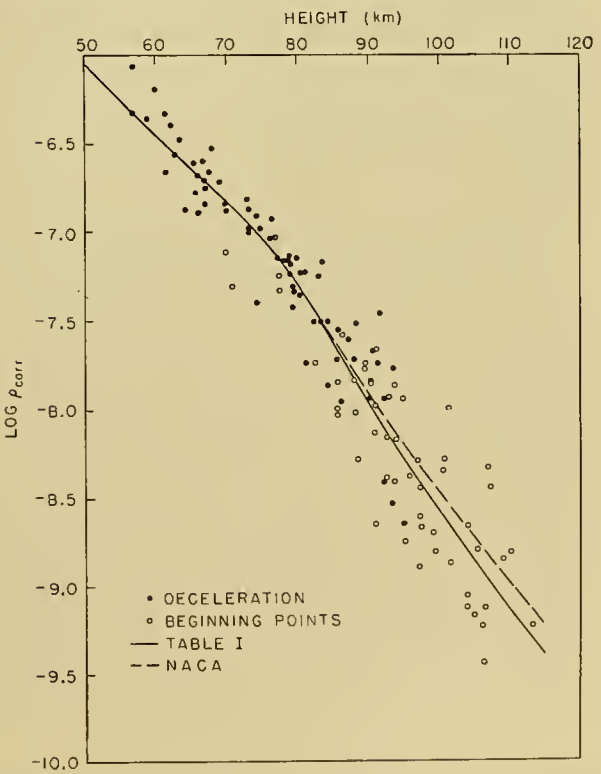


Fig. 3.—Atmospheric densities from photographic meteors.

1. More numerous and accurate observations of me-

teors, by photographic techniques and by combining photographic and radio techniques.

2. Improved theoretical evaluations of the drag coefficient γ, the heat-transfer coefficient λ, and the luminous efficiency factor τ_0.
3. Experimental determinations of γ, λ, and τ_0 by wind-tunnel, ballistic-range, or rocket procedures.
4. Studies of the physical and chemical nature of meteoroids through the study of micrometeorites.

The U. S. Naval Bureau of Ordnance has made possible the continued analysis of Harvard meteor data at the Center of Analysis at the Massachusetts Institute of Technology, carried out by Dr. L. Jacchia under the general direction of Dr. Z. Kopal and with advice from the writer. Meanwhile the Bureau of Ordnance, via the Harvard Meteor Ballistic Project, has made possible the establishment of meteor-observing stations in New Mexico under the direction of the writer [91]. This program includes the procurement of Super-Schmidt Meteor Cameras being made by the Perkin-Elmer Corporation from optical designs by Dr. J. G. Baker and with critical glass produced by the U. S. Bureau of Standards. These cameras, of focal length 8 inches, aperture 12 inches (optically F/0.66), and field of diameter 52°, should satisfy requirement (1) above, particularly if radio equipment can be operated simultaneously nearby. Requirement (4), the study of micrometeorites, will also be undertaken by the Harvard-Bureau of Ordnance program.

The cooperation of the Naval Ordnance Laboratory at White Oak, Maryland, has been obtained toward the fulfillment of requirement (3), particularly the determination of the heat-transfer and drag coefficients for a model meteoroid made from a substance of low melting point, such as CO_2, in a high-velocity wind-tunnel.

It is hoped that the interest of other scientists and research groups may be roused with respect to the various problems mentioned in requirements (2), (3), and (4). The writer is of the opinion that the potentialities of the photographic-meteor approach to upper-atmospheric research have by no means been exploited fully and that rich returns lie ahead. Besides the increased precision possible in density determinations to an altitude of 120 km or more, there are the correlations with latitude and season, as well as conceivable synoptic air-mass and other correlations. A start is just being made on the determination of winds in the region from 50–100 km by photographic meteors. Close cooperations with radio techniques (next section) should produce extremely important results concerning dissociation, ionization, recombination, absorption of quanta, composition, turbulence, and other processes in and near the E-layer of the ionosphere.

RADIO METEORS

The progress in radar techniques for observing the ion columns produced by meteors has been so active since Skellett's suggestion [79, 80] in 1932 that this review can touch on only a few points of major interest. The reader, for detailed enlightenment, must refer to survey articles, such as those by Hey [26], Lovell [45], and Herlofson [25], or to the original papers. Even the present short account will show, however, that the future possibilities of these electronic techniques are truly enormous for the study of the upper atmosphere to a height of at least 130 km via meteoric phenomena.

The basic principle in the radio observation of meteors involves the "reflection" of high-frequency radio waves from the ion column produced by a meteoroid. Pierce [70] suggested that the column should produce maximum "reflection" for normal incidence of the radio beam. The ion columns tend to be more intense in the neighborhood of the E-layer; hence the geometrical relations limit the radar observation of a given meteor. It is found that there is by no means a one-to-one coincidence between the visual and radio meteors unless the geometrical conditions are well satisfied.

Two basic electronic techniques are used for observing meteors by radio. The first involves the transmission of radio pulses a number of times per second, the range (distance) being measured by the time lag of the reflected pulse packets—the standard radar technique. The second depends upon continuous-wave transmission, the Doppler principle providing a modulation at the receiver by the beating between the ground wave and the reflected wave.

The pulse-packet technique, first used for ionospheric research by Breit and Tuve [8], has been the chief tool for radio meteor work until recently. The research previous to 1945 was mostly exploratory, finally establishing the reality and general character of transient echoes from meteor ion columns. The chief workers in this period were Skellett [79, 80], Schafer and Goodall [74, 75], Appleton, Naismith, and Ingram [1], Piddington [2], Pierce [70], Eckersley [14], and Farmer [15]. In 1945 Hey and Stewart [28] began their observations with modified radar equipment operating at a wave length of 4–5 m and established the existence of extensive daytime meteor activity. They recorded range versus time to measure the velocity of the 1946 Giacobinids by the fast-moving echo preceding the main, more persistant, echo.

Lovell [45] and his colleagues at Manchester have since made more extensive meteor observations with similar basic equipment, operating for the most part near a wave length of 4 m but also near 6 and 8 m. As a part of this research project Clegg [11, 12] developed a novel and effective method for measuring the radiants of meteor showers by a single radar. A narrow beam antenna is used to record the frequency and ranges of meteors at various orientations of the radiant (time) and the antenna. The spacial direction of the radiant can then be deduced on the assumption, proven by Lovell, Banwell, and Clegg [47], that an ion column gives the strongest echo when the radar beam is directed perpendicularly to it. By this technique, Clegg, Hughes, and Lovell [13] observed a number of daylight radiants from May to August 1947 and, with Aspinall [4], confirmed their reoccurrence in 1948.

In close coordination with Clegg's work, Ellyett and Davies [19] developed an ingenious method for measur-

ing the velocities of meteor streams. The strengths of individual pulses are recorded on a rapid time base triggered by early echoes from the meteoric ion trail. At nearly perpendicular incidence the echoes from the lengthening ion column vary in strength according to the Fresnel pattern presented from instant to instant. The time rate of the variations measures the angular velocity, which, combined with range data, gives the meteoric velocity. The method was proven on the Geminid shower of 1947 and has been used by Ellyett [18] to demonstrate the heliocentric character of a number of the daylight radiants. The radar technique is being used for radio meteors by Bateman, McNish, and Pineo [5] at the U. S. Bureau of Standards.

The study of meteors by the continuous-wave, Doppler, or "whistling meteor" method was initiated by Chamanlal and Venkatamaran [10] in 1941. Hey, Parson, and Stewart [27] showed that the rapid changes in pitch must arise from changing range rather than deceleration of the meteoroid. A number of investigators listened to the whistles during the Giacobinid shower in 1946. Manning and Villard, heading a group at Stanford University, developed, with Peterson [49], the technique of recording "whistle" oscillations in 1948, a technique that was proven by excellent velocities derived for the 1948 Perseid shower. An important meteorological development by Manning and Villard [48] is that of determining wind velocities and directions in the upper atmosphere by the "body doppler" motion of the meteor ion columns. In eighteen observations from May to September 1949 wind velocities have varied from 47 to 210 km hr^{-1} (mean = 105) with a predominant direction towards SSE, but with a number of motions oppositely directed. This method of observing high-altitude winds shows exceptional promise.

The electronic group of the Canadian Research Council, under the direction of McKinley and cooperating with Millman of the Dominion Observatory, have done some excellent work on meteors, combining radar techniques [53, 54] with visual and photographic techniques (Millman, McKinley, and Burland [58]) and have made most remarkable progress in measuring meteoric velocities by the Doppler method. McKinley [52] has reported on some 3000 meteor velocities measured by this method; the data show no indication of extra-solar-system bodies. The group have succeeded in measuring the deceleration of a meteor by radio techniques alone,[2] a significant step forward in meteoric and upper-atmospheric research. Since the Canadian-group plan to combine radio techniques with visual and photographic meteor observations, using Super-Schmidt cameras when available, their further activities should be scrutinized frequently by those interested in the upper atmosphere.

Theoretical progress in the radio meteor field is still somewhat behind the technological progress; several of the basic processes have not been clearly demonstrated. It is probable that the radio meteor echo arises from the coherent scattering by electrons in a long column of diameter *small* compared to the wave length, as suggested by Blackett and Lovell [7] and investigated more fully by Lovell [45] and Lovell and Clegg [46]. At shorter wave lengths this theory appears to be in better agreement with the observations than Pierce's theory involving a column diameter finite compared to the wave length. Added confidence in the former theory is provided by Herlofson [25], who has semiquantitatively determined the fraction of the meteoroid energy devoted to electron production as about 10^{-4}.

The marked increase in the number of meteors observed at wave length 8 m as compared to 4 m is partially explained. The lower limit of wave length useful for observing meteors, at roughly 3 m, is set by the high electron densities required and by the infrequency of meteors with sufficient energy. The facts of an upper useful limit in wave length, around 100 m, and the time delay (several seconds) in the formation of the observed echoes indicate, however, that an ion column of finite width is required at longer wave lengths. The upper limit is also dependent upon the effects of other atmospheric ionization and the lack of resolving power.

Evidence that the electron diffusion in the ion trail is influenced by the earth's magnetic field has been presented by Lovell [45], following a suggestion by Herlofson. The detailed processes of dissociation, ionization, detachment, diffusion, turbulence, recombination, and attachment of electrons in meteor ion columns, however, require much more theoretical elaboration, since only a modest advance has been made beyond Öpik's meteor theory. Even so, the present theory [45] shows clearly that an important fraction of sporadic-E ionization must arise from some source other than meteors as we now understand them and that the ionization normally produced by meteors is trivial compared to the normal daytime E-layer ionization. During the 1946 Giacobinid Shower, however, meteors produced an appreciable E-layer; Pierce [71] calculated an energy flow of 3 watts km^{-2} over a 4-hour period.

Also, there still is needed a detailed theoretical mechanism for the formation of the ion cap of short duration that follows the motion of the meteoroid. It is difficult to choose among the hypotheses of electron production (1) ultraviolet radiation from the meteoroid, (2) encounters arising from excessive mean free paths of the high-velocity atoms from the meteoroid, or (3) encounters arising from crumbling or spraying effects from finite particles of the meteoroid.

Nevertheless, these present uncertainties only serve to stress the enormous technological progress that has been made in the field of radio meteors. The future possibilities of upper-atmospheric research by radio meteors appear brilliant.

2. Private communication.

REFERENCES

1. Appleton, E. V., Naismith, R., and Ingram, L. J., "British Radio Observations During the Second International Polar Year 1932–33." *Phil. Trans. roy. Soc., London*, (A) 236: 191–259 (1937).

2. Appleton, E. V., and Piddington, J. H., "The Reflexion

Coefficients of Ionospheric Regions." *Proc. roy. Soc.*, (A) 164: 467–476 (1938).

3. Arend, S., et Hunaerts, J., "Courbure des trajectoires d'étoiles filantes." *Bull. astr. Obs. Uccle*, 3: 39–42 (1939).
4. Aspinall, A., Clegg, J. A., and Lovell, A. C. B., "The Daytime Meteor Streams of 1948—I. Measurement of the Activity and Radiant Positions." *Mon. Not. R. astr. Soc.*, 109: 352–358 (1949).
5. Bateman, R., McNish, A. G., and Pineo, V. C., "Radar Observations during Meteor Showers 9 October 1946." *Science*, 104: 434–435 (1946).
6. Bauer, C. A., *On the Age and Origin of Meteorites.* Harvard University Thesis, 1949.
7. Blackett, P. M. S., and Lovell, A. C. B., "Radio Echoes and Cosmic Ray Showers." *Proc. roy. Soc.*, (A) 177: 183–186 (1941).
8. Breit, G., and Tuve, M. A., "A Test of the Existence of the Conducting Layer." *Phys. Rev.*, 28: 554–575 (1926).
9. Brown, H., and Patterson, C., "The Composition of Meteoritic Matter. III. Phase Equilibria, Genetic Relationships and Planet Structure." *J. Geol.*, 56: 85–111 (1948).
10. Chamanlal, C., and Venkatamaran, K., "Whistling Meteors—Doppler Effect Produced by Meteors Entering the Ionosphere." *Electrotechnics*, 14: 28–40 (1941).
11. Clegg, J. A., "Determination of Meteor Radiants by Observation of Radio Echoes from Meteor Trails." *Phil. Mag.*, 39: 577–594 (1948).
12. —— "Determination of Meteor Radiants for the Daytime Showers of 1948 May." *J. Brit. astr. Ass.*, 58: 271–279 (1948).
13. —— Hughes, V. A., and Lovell, A. C. B., "The Daylight Meteor Streams of 1947 May–August." *Mon. Not. R. astr. Soc.*, 107: 369–378 (1947).
14. Eckersley, T. L., "Irregular Ionic Clouds in the E Layer of the Ionosphere." *Nature, Lond.*, 140: 846–847 (1937).
15. —— and Farmer, F. T., "Short Period Fluctuations in the Characteristics of Wireless Echoes from the Ionosphere." *Proc. roy. Soc.*, (A) 184: 196–217 (1945).
16. Elkin, W. L., "Instrument for the Photography of Meteors for the Yale Observatory." *Pop. Astr.*, 2: 17–18 (1894).
17. —— "The Velocity of Meteors as Deduced from Photographs at the Yale Observatory." *Astrophys. J.*, 12: 4–7 (1900).
18. Ellyett, C. D., "The Daytime Meteor Streams of 1948—II. Measurement of Velocities." *Mon. Not. R. astr. Soc.*, 109: 359–364 (1949).
19. —— and Davies, J. G., "Velocity of Meteors Measured by Diffraction of Radio Waves from Trails during Formation." *Nature, Lond.*, 161: 596–597 (1948).
20. Fedynsky, V. V., "Results of Observations of Meteor Trains in Jadjikistan (1934–1938)." *Astr. J. Sov. Un.*, 21: 291–306 (1944).
21. —— and Stanjukovitsch, K. P., "Photograph of a Bright Meteor." *Russ. astr. J.*, 12: 440–449 (1935).
22. Fisher, W. J., and Dimitroff, G. Z., "The Curvature of a Meteor Path." *Bull. astr. Obs. Harv.*, No. 877, pp. 1–6 (1930).
23. Foster, J. F., "Notes on Photographic Meteor Showers." *Bull. astr. Obs. Harv.*, No. 916, pp. 17–18 (1942).
24. Heineman, M., "Theory of Drag in Highly Rarefied Gases." *Comm. appl. Math.*, 1: 259–273 (1948).
25. Herlofson, N., "The Theory of Meteor Ionization" in *Reports on Progress in Physics*, 11: 444–454. Phys. Soc., London, 1946–47.
26. Hey, J. S., "Reports on the Progress of Astronomy. Radio Astronomy." *Mon. Not. R. astr. Soc.*, 109: 179–214 (1949).
27. —— Parson, S. J., and Stewart, G. S., "Radar Observations of the Giacobinid Meteor Shower, 1946." *Mon. Not. R. astr. Soc.*, 107: 176–183 (1947).
28. Hey, J. S., and Stewart, G. S., "Derivation of Meteor Stream Radiants by Radio Reflexion Methods." *Nature, Lond.*, 158: 481–482 (1946).
29. Hoffleit, D., "A Study of Meteor Light Curves." *Proc. nat. Acad. Sci., Wash.*, 19: 212–221 (1933).
30. Hoffmeister, C., *Die Meteore; ihre kosmischen und irdischen Beziehungen.* Leipzig, Akad. Verlagsges., 1937.
31. —— "Zur Frage nach der kosmischen Stellung der Sternschnuppen." *Astr. Nachr.*, 221: 353–379 (1924).
32. —— "Über das Leuchten der Meteore." *Astr. Nachr.*, 269: 109–113 (1939).
33. Hoppe, J., "Die physikalischen Vorgänge beim Eindringen meteoritischer Körper in die Erdatmosphäre." *Astr. Nachr.*, 262: 169-198 (1937).
34. Jacchia, L. G., *Ballistics of the Upper Atmosphere.* Harv. Coll. Obs., Center of Analysis, Mass. Inst. Tech., Tech. Rep. No. 2, 30 pp., Cambridge, 1948. (Harvard Repr. Ser. II, 26.)
35. —— *Photographic Meteor Phenomena and Theory.* Harv. Coll. Obs., Center of Analysis, Mass. Inst. Tech., Tech. Rep. No. 3, 36 pp., Cambridge, 1949. (Harvard Repr. Ser. II, 31.)
36. —— *Atmospheric Density Profiles and Gradients from Early Parts of Photographic Meteor Trails.* Harv. Coll. Obs., Center of Analysis, Mass. Inst. Tech., Tech. Rep. No. 4, 12 pp., Cambridge, 1949. (Harvard Repr. Ser. II, 32.)
37. Kopff, A., "Kometen und Meteore," in *Handbuch der Astrophysik.* Berlin, J. Springer, 1929. (See Vol. 4, pp. 477–495)
38. Landsberg, H. E., "A Report on Dust Collections Made at Mount Weather and Arlington, Virginia, 1 October to 20 November, 1946." *Pop. Astr.*, 55: 322–325 (1947).
39. Levin, B. J., "Elements of the Physical Theory of Meteors." *Astr. J. Sov. Un.*, Vol. 17, No. 3, pp. 12–41 (1940).
40. —— "Elements of the Physical Theory of Meteors. II." *Astr. J. Sov. Un.*, 18: 331–342 (1941).
41. Lindemann, F. A., and Dobson, G. M. B., "A Theory of Meteors, and the Density and Temperature of the Outer Atmosphere to Which It Leads." *Proc. roy. Soc.*, (A) 102: 411–437 (1923).
42. —— "Note on the Photography of Meteors." *Mon. Not. R. astr. Soc.*, 83: 163–166 (1923).
43. —— "A Note on the Temperature of the Air at Great Heights." *Proc. roy. Soc.*, (A) 103: 339–342 (1923).
44. Link, F., "Exploration météorique de la haute atmosphère. *Čas. Pěst. Math.*, 71: 79–90 (1936).
45. Lovell, A. C. B., "Meteoric Ionization and Ionospheric Abnormalities" in *Reports on Progress in Physics*, 11: 415–442. Phys. Soc., London, 1946–47.
46. —— and Clegg, J. A., "Characteristics of Radio Echoes from Meteor Trails: I. The Intensity of the Radio Reflections and Electron Density in the Trails." *Proc. phys. Soc. Lond.*, (A) 60: 491–498 (1948).
47. Lovell, A. C. B., Banwell, C. J., and Clegg, J. A., "Radio Echo Observations of the Giacobinid Meteors, 1946." *Mon. Not. R. astr. Soc.*, 107: 164–175 (1947).
48. Manning, L. A., and Villard, O. G., *Meteor Ionization Research.* Quart. Status Rep. No. 8, pp. 23–26, Electr. Res. Lab., Stanford University, 1949.
49. —— and Peterson, A. M., *Radio Doppler Investigation of Meteoric Heights and Velocities.* Tech. Rep. No. 7, pp.

23–29, Electr. Res. Lab., Stanford University, 1949. (Or see *J. appl. Phys.*, 20: 475–479 (1949).)

50. Maris, H. B., "The Theory of Meteors." *Terr. Magn. atmos. Elect.*, 34: 309–316 (1929).
51. McIntosh, R. A., "Seasonal Variation in the Height of Meteors." *Mon. Not. R. astr. Soc.*, 100: 510–528 (1940).
52. McKinley, D. W. R., "Meteor Velocities Determined by Radio Observations." (Abstract) *Astr. J.*, 54: 179 (1949).
53. —— and Millman, P. M., "A Phenomenological Theory of Radar Echoes from Meteors." *Proc. Inst. Radio Engrs., N. Y.*, 37: 364–375 (1949).
54. —— "Determination of the Elements of Meteor Paths from Radar Observations." *Canad. J. Res.*, (A) 27: 53–67 (1949).
55. Millman, P. M., "An Analysis of Meteor Spectra." *Ann. Harv. Coll. Obs.*, 82: 113–147 (1932).
56. —— "An Analysis of Meteor Spectra: Second Paper." *Ann. Harv. Coll. Obs.*, 82: 149–177 (1935).
57. —— "One Hundred Meteor Spectra." (Abstract) *Astr. J.*, 54: 177–178 (1949).
58. —— McKinley, D. W. R., and Burland, M. S., "Combined Radar, Photographic and Visual Observations of the Perseid Meteor Shower of 1947." *Nature, Lond.*, 161: 278–280 (1948).
59. Mitra, S. K., *The Upper Atmosphere*. Calcutta, Royal Asiatic Society of Bengal, 1948. (See pp. 72–90)
60. Newton, H. A., "Altitudes of Shooting Stars." *Amer. J. Sci.*, Ser. II, 38: 135–141 (1864).
61. Olivier, C. P., *Meteors*. Baltimore, Williams & Wilkins, 1925.
62. —— "Long Enduring Meteor Trains." *Proc. Amer. phil. Soc.*, 85: 93–135 (1941–42).
63. —— "Long Enduring Meteor Trains. Second Paper." *Proc. Amer. phil. Soc.*, 91: 315–327 (1947).
64. —— "Results of the Yale Photographic Meteor Work, 1893–1909." *Astr. J.*, 46: 41–57 (1937).
65. Öpik, E. J., "Meteor Heights From the Arizona Expedition." *Proc. nat. Acad. Sci. Wash.*, 22: 525–530 (1936).
66. —— "Results of the Arizona Expedition for the Study of Meteors. VI. Analysis of Meteor Heights." *Ann. Harv. Coll. Obs.*, 105: 549–600 (1936).
67. —— "Research on the Physical Theory of Meteor Phenomena. III. Basis of the Physical Theory of Meteor Phenomena." *Tartu Obs. Pub.*, 29: 1–67 (1937).
68. —— "Results of the Arizona Expedition for the Study of Meteors. III. Velocities of Meteors Observed Visually." *Harv. Obs. Circ.*, No. 389, 9 pp. (1934).
69. —— "Atomic Collisions and Radiation of Meteors." Harv. Repr., No. 100, 39 pp. (1933).
70. Pierce, J. A., "Abnormal Ionization in the E-Region of the Ionosphere." *Proc. Inst. Radio Engrs., N. Y.*, 26: 892–908 (1938).
71. —— "Ionization by Meteoric Bombardment." *Phys. Rev.*, 71: 88–92 (1947).
72. Porter, J. G., "An Analysis of British Meteor Data." *Mon. Not. R. astr. Soc.*, 103: 134–153 (1943).
73. —— "An Analysis of British Meteor Data: Part 2. Analysis." *Mon. Not. R. astr. Soc.*, 104: 257–272 (1944).
74. Schafer, J. P., and Goodall, W. M., "Observations of Kennelly-Heaviside Layer Heights During the Leonid Meteor Shower of Nov., 1931." *Proc. Inst. Radio Engrs., N. Y.*, 20: 1941–1945 (1932).
75. —— "Kennelly-Heaviside Layer Studies Employing Rapid Method of Virtual-Height Determination." *Proc. Inst. Radio Engrs., N. Y.*, 20: 1131–1148 (1932).
76. Schiaparelli, J. V., *Sternschnuppen*. Stettin, Verlag von Th. von der Nahmer, 1871. (A classic, practically unobtainable in America.)
77. Shapley, H., Öpik, E. J., and Boothroyd, S. L., "The Arizona Expedition for the Study of Meteors." *Proc. nat. Acad. Sci. Wash.*, 18: 16–23 (1932).
78. Sitnikov, P. F., "On the Curvature of the Meteor Trail of 1937, August 4th." *Russ. astr. J.*, Vol. 16, No. 1, pp. 47–52 (1939).
79. Skellett, A. M., "The Ionizing Effect of Meteors in Relation to Radio Propagation." *Proc. Inst. Radio Engrs., N. Y.*, 20: 1933 (1932).
80. —— "The Ionizing Effects of Meteors." *Proc. Inst. Radio Engrs, N. Y.*, 23: 132–249 (1935).
81. Sparrow, C. M., "Physical Theory of Meteors." *Astrophys. J.*, 63: 90–110 (1926).
82. Tsien, H. S., "Superaerodynamics. Mechanics of Rarefied Gases." *J. aero. Sci.*, 13: 653–664 (1946).
83. Vyssotsky, A. N., "A Meteor Spectrum of High Excitation." *Astrophys. J.*, 91: 264–266 (1940).
84. Warfield, C. N., "Tentative Tables for the Properties of the Upper Atmosphere." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1200, 50 pp. (1947).
85. Watson, F., *Between the Planets*. Philadelphia, Blakiston, 1941.
86. —— "Daily Variation of Photographed Meteors." *Bull. astr. Obs. Harv.*, No. 910, 30 pp. (1939).
87. Whipple, F. L., "Upper Atmosphere Densities and Temperatures from Meteor Observations." *Pop. Astr.*, 47: 419–424 (1939).
88. —— "Photographic Meteor Studies, I." *Proc. Amer. phil. Soc.*, 79: 499–548 (1938).
89. —— "Photographic Meteor Studies, II. Non-Linear Trails." *Proc. Amer. phil. Soc.*, 82: 275–290 (1940).
90. —— "Meteors and the Earth's Upper Atmosphere." *Rev. mod. Phys.*, 15: 246–264 (1943).
91. —— "The Harvard Photographic Meteor Program." *Sky and Telescope*, 8: 90–93 (1949).
92. —— "The Theory of Micro-meteorites, I." *Proc. nat. Acad. Sci., Wash.*, 36: 687–695 (1950). "II." *Ibid.*, 37: 19–30 (1951).
93. —— Jacchia, L. G., and Kopal, Z., "Seasonal Variations in the Density of the Upper Atmosphere," *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed. Chicago, University of Chicago Press, 1949.
94. Whitney, W. T., "The Determination of Meteor Velocities." *Pop. Astr.*, 45: 162–165 (1937).
95. Wylie, C. C., "Where Do Meteorites Come From?" *Science*, 90: 264–265 (1939).

SOUND PROPAGATION IN THE ATMOSPHERE

By B. GUTENBERG

California Institute of Technology

THEORY OF SOUND WAVES IN GASES

Velocity of Sound. The velocity C with which a change in volume is propagated in a material with Lame's constants λ and μ and with the density ρ is given by

$$C^2 = (\lambda + 2\mu)/\rho. \tag{1}$$

In gases, the rigidity μ is practically zero, and the constant λ is equal to the bulk modulus k. Therefore equation (1) reduces to

$$C^2 = k/\rho \qquad \text{where} \qquad k = -(dp/dv)v, \tag{2}$$

and p = pressure, v = volume. Newton assumed that the propagation of sound is an isothermal process, but disagreement between observations and calculations under this assumption led Laplace to suppose that it is an adiabatic process. Except for instances where the change in pressure is large as compared with the pressure itself, results calculated under this assumption agree with the observations.

If v = volume and p = pressure, $K = c_p/c_v$, where c_p = specific heat at constant pressure, c_v = specific heat at constant volume, the condition for an adiabatic process is

$$pv^K = \text{const} \qquad \text{or} \qquad dp/dv = -Kp/v. \tag{3}$$

Introducing this in equation (2), we obtain

$$k = Kp \qquad \text{and} \qquad C^2 = Kp/\rho. \tag{4}$$

It has been assumed in equation (3) that K does not depend on the pressure. In general, this approximation is sufficient; for a better approximation, see Hardy and others [15]. If T is the absolute temperature and R the gas constant of the specific gas,

$$R = p_0\alpha/\rho_0 = p/\rho T, \tag{5}$$

where $\alpha = 1/273.18$ and ρ_0 = density of the gas at 0C and a pressure of 1 atmosphere; if C_0 is the corresponding sound velocity, equation (4) can be written

$$C^2 = KRT = C_0^2\alpha T \quad \text{where} \quad C_0^2 = KR/\alpha. \tag{6}$$

For dry air $K = 1.403$, $R = 2.87 \times 10^6$ cm^2 sec^{-2} deg^{-1} which gives $C_0 = 331.6$ m sec^{-1}. From a critical discussion of observations, Hardy and others [15] found 331.46 ± 0.05 m sec^{-1}. With this value equation (6) becomes

$$C = 20.06\sqrt{T}\ (\text{m sec}^{-1}). \tag{7}$$

Experiments did not disclose any appreciable difference in the sound velocity for frequencies between 10^3 and over 10^6 cycles. However, at very low and very high temperatures the observed velocity is smaller than the value calculated from equation (7). At -100C the difference is about 2 m sec^{-1}, at $+1000$C about 10 m sec^{-1}. Quigley [25] suggested the following empirical equation from observations at low temperatures:

$$C^2 = 387.62T + 180430T^{-1} - 20364000T^{-2} + 806 + 0.03007T^2, \tag{8}$$

which gives $C = 330.6$ m sec^{-1} for $T = 273.1$K.

In humid air the sound velocity is higher than in dry air by an amount H given approximately [8] by

$$H = 0.14Ch/p, \tag{9}$$

where h = partial pressure (or tension) of the water vapor. For warm, humid air near the ground H may be as high as 5 m sec^{-1}.

Sound Rays in Quiet Air. The ray equation for sound waves in still air follows from Snell's law and equation (7):

$$\frac{\sin i_1}{\sin i_2} = \frac{C_1}{C_2} = \left[\frac{T_1}{T_2}\right]^{1/2}, \tag{10}$$

in which i = angle of incidence (between the ray and a vertical line) at the points 1 and 2, where the wave velocities are C_1 and C_2; C and T are supposed to be a function of height z only.

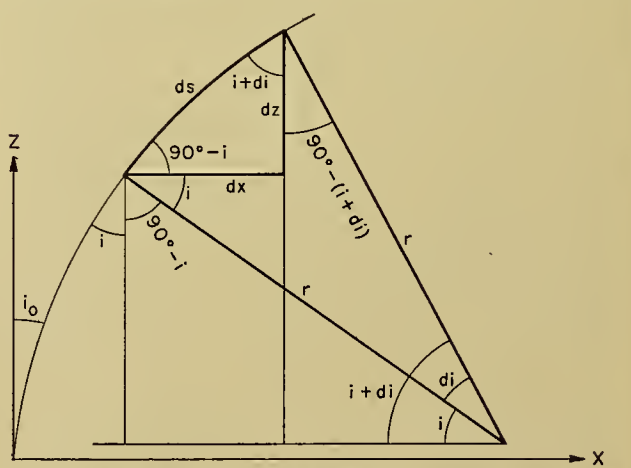


FIG. 1.—Path of sound waves, angles of incidence i, and radius of curvature r.

The radius of curvature r of the ray can be found from Fig. 1 and equation (10):

$$r = ds/di = dz/(\cos i)\, di = dz/d(\sin i),$$

$$\frac{1}{r} = \frac{d(\sin i)}{dz} = \frac{\sin i}{C}\frac{dC}{dz} = \frac{dT}{dz}\frac{\sin i}{2T}. \tag{11}$$

In a layer with a constant value of

$$\gamma = -dT/dz, \tag{12}$$

$$r = \frac{-2T}{\gamma \sin i}. \tag{13}$$

Since the lapse rate γ is usually several degrees per kilometer, the radius of curvature of sound rays is in most instances of the order of 100 km. In the troposphere, positive values of γ (*i.e.*, negative values of dT/dz) prevail, and the sound rays are turning upward; in layers of inversions and in the lower stratosphere γ is negative, and the sound rays are turning towards the earth's surface in still air.

The ray equation in Cartesian coordinates follows from Fig. 1 and equation (10):

$$x = \int (\tan i)\, dz = \int \left[\left(\frac{C_0}{C \sin i_0} \right)^2 - 1 \right]^{-\frac{1}{2}} dz = \int \left[\frac{T_0}{T \sin^2 i_0} - 1 \right]^{-\frac{1}{2}} dz. \tag{14}$$

Even in the simple case of a constant lapse rate the solution becomes rather complicated:

$$x = (AY - Y^2)^{\frac{1}{2}} + \tfrac{1}{2} A \sin^{-1}[(2AY/A) - 1], \tag{15}$$

where

$$A = \frac{-T_0}{\gamma \sin^2 i_0}, \qquad Y = z - T_0/\gamma.$$

For this reason, the radius of curvature is frequently used to find approximately the form of the ray, or $\tan i$ is calculated from equation (10) and graphically integrated to find the horizontal distance x corresponding to a certain part of the ray (using the first part of equation (14)). The corresponding travel time t follows from

$$dt = \frac{ds}{c} = \frac{dz}{C \cos i} \quad \text{or} \quad t = \int \frac{dz}{C \cos i}. \tag{16}$$

The angle of incidence i at a given level, especially at the ground, can be found from Fig. 2,

$$\sin i = dw/dx = C\,(di/dx) = C/V, \tag{17}$$

where V is the "apparent velocity" of the sound observed in the horizontal direction (along x) and given

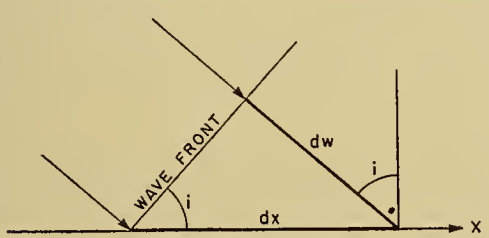


FIG. 2.—Sound waves arriving at the surface of the earth and angle of incidence i.

by the "travel-time curve" (time of arrival of a given impulse plotted against the distance x).

The apparent velocity V is equal to the true velocity C^* at the highest point of the ray (if $i_2 = 90°$ and $C_2 = C^*$, in equation (10), we find $\sin i = C/C^*$; equation (17) then gives $V = C^*$). The corresponding height $z = H$ can be found from

$$H = \frac{1}{\pi} \int q\, dx \quad \text{where} \quad \cosh q = \frac{V_\Delta}{V_x} = \frac{\sin i_\Delta}{\sin i_x}. \tag{18}$$

This equation is based on a solution by Herglotz [17] which, in simplified form, can be found in any book on theoretical seismology. The quantities V_Δ and i_Δ are the apparent velocity and angle of incidence respectively at a given level for the ray which arrives at the distance $x = \Delta$ and has its highest point at $z = H$; V_x and i_x are the corresponding values for all points between $x = 0$ and $x = \Delta$. Equation (18) can be used only if the travel-time curve is continuous in this interval. In this case it permits the calculation of H as a function of the distance x, while the velocities at the height H (and from them T) are given by the value of V at the corresponding distances x. Since in most practical cases the temperature T and the velocity C decrease with height in the troposphere, the method can not be applied there. In this case a reference height (the tropopause or a higher surface) must be selected above which the temperature increases with height, and the horizontal component for the path and the corresponding travel time t for the parts of the ray below the reference height must be calculated from equations (14) and (16) and deducted from the total distance and travel time. Then equation (18) can be applied to the part of the ray which is above the reference height.

Sound Rays in Moving Air. Equations for the paths of sound waves, in the case where the wind direction forms an angle with the horizontal component of the sound path, are too complicated to be useful [6] (even if it is assumed that the wind has no vertical component and that its direction does not change with height) since the sound path is no longer a plane curve. Usually the wind component in the plane of the sound wave is considered and the component perpendicular to it is neglected.

If W = wind component in the direction of the sound wave, and C = sound velocity given by equation (7), then the ray equation is

$$(C/\sin i) + W = \text{const}, \qquad \sin i = C/(C_0/\sin i_0) + W_0 - W. \tag{19}$$

Neglecting squares and higher terms in W/C, the radius of curvature r of the ray is now given [8] by

$$r = \frac{C + 3W \sin i}{\frac{10}{\sqrt{T}} \left(\sin i - \frac{W}{C} \cos 2i \right) \frac{dT}{dz} + \left(1 + \frac{W}{C} \sin^3 i \right) \frac{dW}{dz}}, \tag{20}$$

where C and W are measured in meters per second and i is given by equation (19). If $-dT/dz = \gamma$ = lapse rate, and the change of wind velocity with height,

$dW/dz = a$, the condition for a curvature of the ray towards the ground is given by

$$a\,[1 + (W/C)\sin^3 i] > 10\gamma T^{-\frac{1}{2}}\,[\sin i - (W/C)\cos^2 i]. \quad (21)$$

Equation (21) shows that for our problem the lapse rate γ and the change a of wind velocity with elevation are more important than the ratio of the wind velocity to the sound velocity (W/C).

For very flat rays (i near 90°) equation (21) leads to

$$a > 10\gamma T^{-\frac{1}{2}}, \quad (22)$$

from which it follows that under average conditions in the lower troposphere (γ about 6C km^{-1}, T = 290K) the wind must increase with elevation by more than about 3½ m sec^{-1} km^{-1} in the direction of the sound propagation to bring the sound rays, which leave the source nearly horizontally, back to the surface. In the opposite direction the sound rays are curved upward more strongly than in still air (r is smaller).

The following are a few critical values of a for straight rays under the same assumptions for γ and T and supposing that W/C is 0.1:

i	80°	45°	5.7°	0°
$a = dW/dz$ (m sec^{-1} km^{-1})	3.1	2.5	0	−0.4

If a exceeds these critical values, the ray curves downward in the wind direction; if a is smaller, the rays turn upward. It is noteworthy that in general the wave front is no longer perpendicular to the rays if sound is propagated in air in which the wind velocity changes with elevation.

If the paths of the rays are close to a part of a circle, use of trigonometry shows that the maximum height H^* which is reached by a ray at the distance Δ is given approximately by

$$H^* = r\{1 - [1 - (\Delta/2r)^2]^{\frac{1}{2}}\} \approx \Delta^2/8r. \quad (23)$$

If, for example, a = 5 m sec^{-1} km^{-1}, γ = 6C km^{-1}, T = 300K, W_0 = 5 m sec^{-1} (which gives C = 348 m sec^{-1}, r = 235 km), the following values result:

Δ	1 km	10 km	100 km
i	89.9°	88.8°	77.6°
H^*	½ m	54 m	5½ km

The values in the last column would normally be useless, since the assumption of circular rays is not fulfilled between the ground and an elevation of 5½ km. On the other hand, the results show that conditions favorable to good audibility of sound waves to distances of many kilometers at the ground need not extend to great heights.

The preceding equations can also be used to answer the question, How close must a sound source at an elevation h be in order that the sound can be heard at the ground? If, for example, the lapse rate is γ and the wind is negligible, the maximum horizontal distance Δ^* at which direct rays arrive at the ground is given approximately (assuming circular rays) by

$$\Delta^* = a\,(Th/\gamma)^{\frac{1}{2}},$$

which for γ = 6C km^{-1} and T = 289K gives $\Delta^* = 14h^{\frac{1}{2}}$.

All preceding results are based on the assumption that the change in pressure during the process is small compared with the pressure itself. For larger changes in pressure, the sound velocity is greater than given by the preceding equations. Since the theory is very complicated and has rarely been needed in problems of sound propagation in the atmosphere, it is not considered here.

Another assumption is that the wave length is small as compared with the height A of the homogeneous atmosphere, since otherwise dispersion results as an effect of gravity. Assuming that the gravitational constant g does not change with altitude (which does not hold for the higher parts of the stratosphere), Schrödinger [26] found that the velocity v of sound waves moving vertically upward is given approximately by

$$v/C = 1 + (1/32\pi^2)\,(L/A)^2 = 1 + 0.003166\,(L/A)^2, \quad (24)$$

where L = wave length, C = sound velocity in quiet air and for waves moving horizontally; C is given by equation (7). For waves with periods of less than 1 sec, the last term in equation (24) is smaller than 10^{-6}, and the effect of the dispersion (long waves travel faster) is negligible. In addition, in most practical cases the sound waves travel much closer to the horizontal than the vertical direction, which decreases the effect of dispersion. The group velocity for waves travelling upward is given by an equation similar to (24) except that the last term is negative.

For long waves, equation (24) cannot be used. In addition to the effect of decrease in gravity with elevation, free vibrations of the atmosphere affect the wave propagation. The theory for various models representing the atmosphere has been developed by Pekeris [23]. In addition to the body waves discussed above, waves corresponding to free oscillations may be excited within certain ranges of frequencies, depending on the assumed model.

Energy of Sound Waves in the Atmosphere. The energy and amplitudes of sound waves in the atmosphere with periods not exceeding a few seconds depend on two major effects: (1) absorption, and (2) changes due to the increase (or occasionally decrease) in the distance between rays (which controls the energy flux).

The absorption of the energy is usually introduced by a factor e^{-kD}, where k is the coefficient of absorption, supposed to be constant over the distance D along the ray. If k changes along the ray, kD has to be replaced by $\int k\,dD$.

The propagation of sound is a molecular process. The velocity C of sound and the molecular velocity c are connected by the equation (see, for example, Schrödinger [26]):

$$(C/c)^2 = K/3; \quad (25)$$

for air $K = c_p/c_v$ = 1.403, and

$$C = 0.6840c.$$

As long as the molecules are relatively close together, absorption of sound remains small. It increases rapidly

if the wave length L approaches the mean free path l of the molecules. Schrödinger [26] writes

$$k = Fl/L^2, \tag{26}$$

where

$$F = 12\pi^2\sqrt{3}[(K-1)AK^{-3/2} + 4BK^{-1/2}].$$

$K = c_p/c_v$; A and B are constants depending on the heat conduction q of the gas, its coefficient of internal friction u, the molecular velocity c, and the mean free path l of the molecules:

$$q = Ac, \qquad u/c = Bcl. \tag{27}$$

Schrödinger calculated from laboratory experiments $F = 30.1$; Kölzer [18] later used $F = 33.0$. Consequently, loss of energy E from absorption between two points close enough to each other so that the absorption can be assumed to be constant is given by

$$\ln(E_2/E_1) = -0.4343\,Fl\,D/L^2 = -0.0013\,e^{h/8}/L^2. \tag{28}$$

The factor $e^{h/8}$ is based on the assumption of constant temperature throughout the atmosphere, but the resulting mean free paths are within the limits given by the N.A.C.A. tables for the stratosphere (see Table I).

TABLE I. MEAN FREE PATH ($l \times 10^{-5}$ cm) OF MOLECULES AT DIFFERENT LEVELS

	Height above sea level (h km)						
	0	20	40	60	80	100	120
(a)	1.0	12	150	1,800	22,000	270,000	3,300,000
(b)	0.6	15	580	14,000	200,000	6,160,000	66,000,000
(c)	0.7	10	259	2,600	20,000	260,000	1,700,000
(d)	0.8	8	120	830	4,400	36,000	140,000

(a) $l = 10^{-5}\,e^{h/8}$.
(b) N.A.C.A. [21] tentative minimum temperatures.
(c) N.A.C.A. tentative standard temperatures.
(d) N.A.C.A. tentative maximum temperatures (for $h = 100$ and 120 km during the day).

If it is assumed that the distance D is 1 km, that the free path of the molecules decreases exponentially with altitude h (in km) from 10^{-5} cm at sea level, that the height of the homogeneous atmosphere is 8.0 km, and that the wave length L is measured in meters,

$$\ln L = 0.27h - 1.443 - \tfrac{1}{2}\ln[-\ln(E_2/E_1)]. \tag{29}$$

Equation (29) enables the calculation of L as a function of h for a given absorption. Results are plotted in Fig. 3, which can be used to find the height h above which waves of a given length L lose their energy by absorption too quickly to be observed. Figure 14 shows that sound waves travel relatively long distances near the top level of their path. If it is assumed that a loss of 10 per cent of the energy per kilometer over a distance of 100 km (which results in a reduction of the energy to less than 10^{-4} and of the amplitude to less than 0.01) corresponds to the normal limit for the observation of such waves, it follows from Fig. 3 that, even in the lower stratosphere, tones with frequencies above that of a soprano cannot travel very far without being absorbed almost completely. For waves with a frequency near that of the middle a (435 cps, wave length at 0C about ¾ m) the critical level is about 25 km; for the lowest audible tones (16 cps, L near 20 m) the critical level is below 80 km; and for waves with periods of 0.3 sec it is about 100 km.

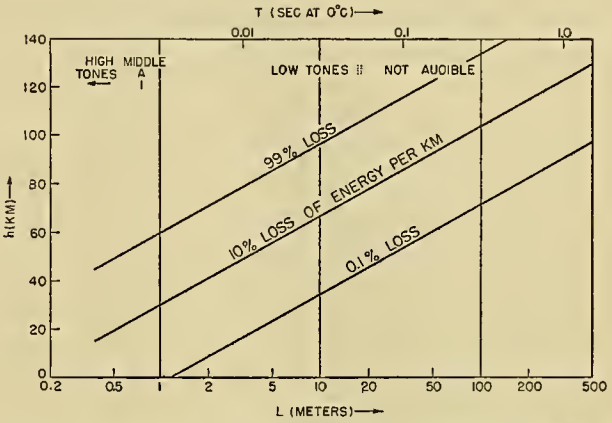


FIG. 3.—Fractions of sound energy which are lost by absorption when the sound waves travel a distance of 1 km (assuming a temperature of 0C throughout the atmosphere) based on the research of Schrödinger [26]; h = height in km, L = wave length in meters, T = period in seconds.

The preceding equations give only the order of magnitude of the absorption. Fog, smoke, water droplets, etc., affect the absorption, and the equations do not hold for waves with very short wave lengths (less than one meter), for which the absorption increases faster than given by the equations, nor for waves with lengths over a few hundred meters, for which the wave length is an appreciable fraction of the height of the homogeneous atmosphere.

The amplitudes of recorded sound waves through the troposphere and the lower part of the stratosphere depend less on absorption than on the change in energy flux due to the change in size of the wave front. To find these effects [8] we suppose that there is no wind,

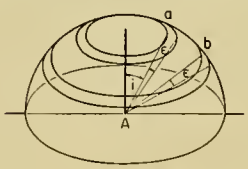


FIG. 4.

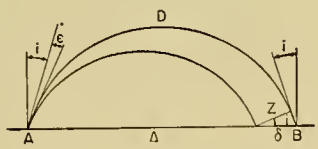


FIG. 5.

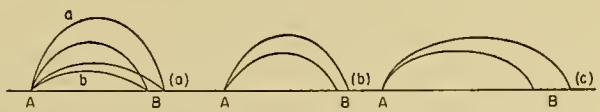


FIG. 6.

and that sound waves produced at A in Fig. 4 start with the same energy in all directions. The energy flux E_0 through a zone z between two cones formed by rays with angles of incidence i and $i + \epsilon$ respectively is given (see Figs. 4 and 5) by

$$E_0 = A[\cos i - \cos(i + \epsilon)] = -B\delta\,\frac{d\cos i}{d\Delta}, \tag{30}$$

where A and B depend on the energy at the source. Neglecting the absorption along the path D (Fig. 5) the energy arrives on a zone of area Z perpendicular to the rays,

$$Z = \pi[\Delta^2 - (\Delta - \delta)^2] \cos i = 2\pi\delta\Delta \cos i. \quad (31)$$

The energy flux E at the distance Δ is then given by

$$E = \frac{E_0 z}{Z} = \frac{C\delta\left(\frac{d \cos i}{d\Delta}\right)}{\delta\Delta \cos i} = \frac{C(\tan i)\frac{di}{d\Delta}}{\Delta}. \quad (32)$$

C is a constant depending on the energy at the source. The factor $\tan i$ is partly a consequence of the fact that near the source less energy passes through the zone a (Fig. 4) than through the larger zone b (see also Fig. 6a). The greatest changes are produced by $di/d\Delta$. If i changes slowly with distance (Fig. 6c), the energy flux near B is relatively small.

To simplify the picture, we assume that rays are parts of circles (r = radius). Then $\cos i = \Delta/2r$, and equation (32) becomes in this special case [8]

$$E^* = (C/\Delta^2)\,[1 - (\Delta/r)\,(dr/d\Delta)], \quad (33)$$

which shows the expected decrease in energy with the square of the distance. It is evident that in general the energy at a given point is especially large if the (mean) radius of curvature r decreases rapidly with distance Δ (or with the maximum height reached by the ray). According to equation (20), this requires that either the wind or the temperature or both increase rapidly with increasing height in the region of the highest point reached by the sound wave. Focal points (caustics) will result if $di/d\Delta$ becomes infinite; this would correspond to Fig. 6b, if two rays with small differences in i arrive at the same point B. On the other hand, the energy decreases considerably with distance if the highest parts of the rays enter a region where the radius of curvature increases rapidly, until the limit for straight rays is reached (equation (21)) and no energy arrives at the ground. For rays through the stratosphere, the absorption must be considered, and

$$E = Ce^{\int k d\nu}\,\frac{(\tan i)}{\Delta}\,\frac{di}{d\Delta}. \quad (34)$$

INSTRUMENTS FOR THE RECORDING OF SOUND WAVES

Most instruments used in recording sound waves through the atmosphere react to the change in pressure produced by the sound. In addition, records of distant explosions are sometimes obtained through seismographs; in such instances the vibrations of the atmosphere are transmitted by buildings or otherwise to the ground near the instrument. For sound-recording instruments, membranes or pistons which form a part of an airtight container of a given volume of air are frequently used; their movements are magnified either by levers which operate a recording pen, by the mirror of an optical recording system, or through electromagnetic systems which record by means of galvanometers. Even rather simple devices, such as a wire attached to the center of a windowpane and wound under tension around a thin needle carrying a mirror, have been used successfully for the recording of sound waves from distant explosions. The magnification of these types of instruments for long-period movements equals their magnification for a constant continuous pressure; it decreases in the case of high damping to about 25 per cent for waves with half the free period of the vibrating system and exponentially toward zero for waves with still higher frequencies. If the damping ratio is less than 23:1, resonance increases the magnification near the free period of the instrument [9] and produces a maximum there of about twice the static magnification for a damping ratio of about 2½:1.

Many types of microbarographs with galvanometric recording have been developed in recent years. In Benioff's microbarograph [13] a permanent-magnet moving-conductor type loud-speaker mounted in one of the sides of a sealed container is used as the responding element. Output currents are recorded on standard seismograph galvanometric recorders. With a 1.2 sec galvanometer the maximum sensitivity is approximately 1 mm deflection for 0.001 mb. In the microbarograph of Baird and Banwell [2] a diaphragm of "dulalium," 0.005 mm thick, separates the body of the microphone into two compartments, one of which is closed to short-period pressure changes. Parallel to the diaphragm at a distance of about 0.015 mm is a brass disk which together with the diaphragm acts as a condenser whose capacity changes are magnified by electronic means and recorded by a galvanometer. The maximum sensitivity is about 1-mm deflection for less than 0.0001 mb. In the instruments designed and built at the Naval Ordnance Laboratory, Washington, D. C. [1] to record the subsonic waves from the Helgoland explosion in 1947, the flexing of a diaphragm in a microphone alters one of two matched inductive circuits; the unbalance produces a signal voltage. This signal, the low-frequency modulation of an audiofrequency carrier, is amplified and used to drive a modified Esterline-Angus graphic recorder for which a response curve has been given by Cox [4]. The maximum sensitivity is about 0.8-mm deflection for 0.001 mb. Other microbarographs, such as Macelwane's, have their maximum response for longer waves.

Unfortunately, all these instruments respond not only to pressure waves, but also to pressure changes caused by air currents. If three instruments, forming a triangle with sides of somewhat less than one half the wave length of the sound waves, are used, the time differences between the instruments indicate whether the disturbance has travelled with the velocity of sound or with the much smaller velocity of air currents [3]. Simultaneously, the time differences between the passing of a given point of the same wave at two pairs of stations makes it possible to calculate the direction from which the waves arrive and their angle of incidence [19]; the sound velocity must be calculated from the temperature.

OBSERVATIONS AND THEIR INTERPRETATION

Natural Sound. Microbarographs of sufficient sensitivity record a background which consists of the effect of air currents and of natural pressure waves which are sometimes especially clear on calm winter days. There are several types of natural sound waves. At Pasadena, California, the most frequent are sinusoidal waves with variable amplitudes and periods of about ½ to 5 sec [3, 13] (Fig. 7), corresponding to wave lengths of about

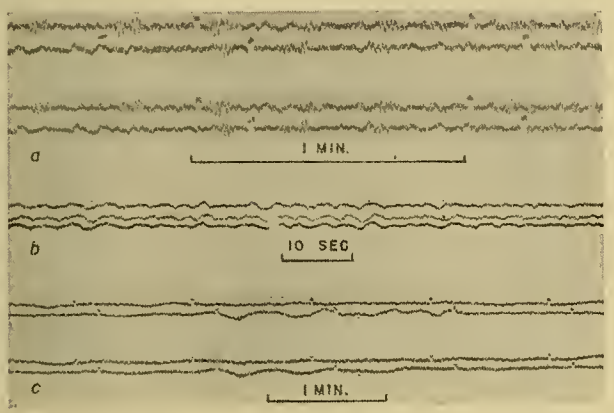


FIG. 7.—Pressure waves recorded by Benioff microbarographs at Pasadena. (*a*) January 6, 1939; these coincided with the largest surf waves in several years near Scripps Institute of Oceanography at La Jolla; no meteorological element had unusually great values within several hundred miles; the two records correspond to two instruments about 30 m apart. (*b*) Pressure waves recorded on December 27, 1940, with three microbarographs forming a triangle with sides of 290, 320, and 264 m, respectively; these waves are frequent during the winter and arrive usually from the SW. (*c*) Pressure waves with longer periods, recorded on January 8, 1939, by two instruments 120 m apart. (*After Benioff and Gutenberg* [3].)

100 m to 1500 m. In general, they are largest in winter. During the three winter months of 1940/41, three microbarographs were operating (Fig. 7*b*) which permitted the calculation of the direction from which the waves arrived; all came from southwest to west (azimuths between west and 40° south of west). In all instances of large amplitudes a low-pressure area was situated off the coast of southern California; as soon as the low-pressure area passed the coast, the amplitudes decreased rather rapidly. The waves arrived almost horizontally, but this is to be expected (even if the source is rather high in the atmosphere) due to the curvature of the rays. No connection with microseisms could be found. Similar air-pressure oscillations have been recorded in Christchurch, New Zealand [2]. Their periods were between 4 and 10 sec, and they showed similarity to microseisms, but a time lag of pressure oscillations, up to 200 sec, was suspected. Baird and Banwell believed that the two types of waves are independent phenomena but possibly have the same cause. Polli (1949) found similar results in Trieste. Observations of these pressure waves at other locations are very desirable.

Occasionally, short groups of pressure oscillations were recorded at Pasadena with periods of about ½ to 1 sec and occurring at intervals of about 20 sec (Fig. 7*a*). Their direction (from SW), their regular coincidence with high ocean waves at the coast, and the lack of other unusual phenomena make it likely that they are "sound" from surf. The causes of longer pressure waves which were recorded occasionally (Fig. 7*c*) could not be found due to the scarcity of the observations.

Sound Waves through the Troposphere. When unusually strong sound waves from distant explosions were first heard, attempts were made to explain the observed zones of audibility and of silence by combinations of temperature, lapse rates, and change of wind with elevation [20]. This possibility was disproved by the occasional occurrence of instances in which the abnormal zone formed a full ring around the source of sound, separated from the normal zone of sound by a zone of silence. However, there are instances, not infrequent, of strong sound in relatively small areas, which are due to meteorological conditions in the troposphere. It follows from equations (20) to (22) that either a temperature inversion, or a certain increase of wind with elevation (depending on the lapse rate), or both jointly, may produce zones of strong sound. It is of interest to note that in the attempted explanations mentioned above an increase of wind with elevation by 4 or 5 m sec^{-1} km^{-1} was usually assumed together with the typical temperature lapse rate, after considering a variety of conditions; such requirements follow now from equation (22). Figure 11*a* shows records with clear dispersion. Schulze [27] explained this dispersion as a consequence of the fact that the velocity changes with elevation, and because the energy of longer waves is propagated in a thicker layer than that of shorter waves.

Strong audibility of certain sounds (*e.g.*, of trains) is occasionally used by laymen for short-range weather forecasting. This possibility is based on the fact that the repeated occurrence of a certain unusually strong sound from the same source will frequently be a consequence of similar meteorological conditions at a given place.

Sound Waves Through the Stratosphere and Abnormal Audibility Zones. In 1903 an accidental explosion of dynamite in Westphalia resulted in sound waves which were heard far away, beyond a zone of silence. The observations were investigated by von dem Borne who believed that an increase in the percentage of light gases with elevation in the atmosphere causes an increase in sound velocity. (For historical data and bibliographies see [28, 10].) The details of these abnormal zones were studied in instances of artificial explosions, especially in Germany [16], of gun fire, and of accidental explosions, using ear observations (Fig. 8) as well as instrumental records. There are instances where the abnormal zone forms a complete ring around the source, and others in which only parts of a ring with audible sound were developed. Frequently, parts of a second or third ring of abnormal audibility were found (Figs. 8 and 9).

In Europe and in Japan, the radius of the ring as well as the distance of the largest sound intensity shows a yearly period (Fig. 10) with a minimum distance of about 100 km late in winter or early in spring,

and a maximum (about 200 km) late in summer [28; 29, pp. 184–187]. In central Europe, the ring is much better developed to the east of the source during winter,

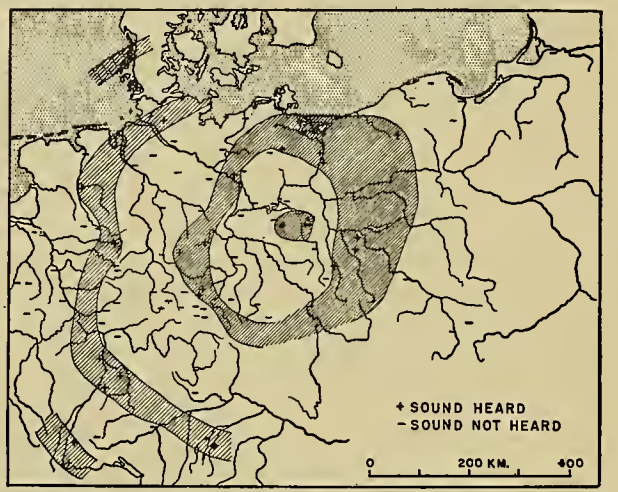


FIG. 8.—Observations of sound waves through the stratosphere after explosion of 5000 kg of ammunition, December 18, 1925. (*After Hergesell and Duckert* [16].)

to the west during summer. This is due to the fact that the rays have rather large angles of incidence (normally between 75° and 90°) so that at the source sin i is usually larger than 0.98. If the temperature at the point of observation is more than about 10C higher than at the source (continent in summer, ocean in winter), it becomes increasingly probable that this difference makes the return of the ray to the ground impossible; the rays turn upward again at a level with a temperature given by equation (10). Similar effects can be produced by prevailing winds (equation (19)).

Records show relatively small differences in travel time of rays to the abnormal zone between day and night [19], although details of the records change gradually (Fig. 11*b*). Wide variations in the amount of explosive do not result in different travel times. This proves that the return of the rays to the ground is not affected by an increase in the ratio of the pressure change to the pressure itself and that, even at the highest levels reached by the rays, equation (7) holds. This conclusion has been confirmed by the agreement between the temperatures calculated from equation (7) and those found from records made with V-2 experiments (Fig. 12). Finally, records taken at equidistant points in opposite directions from the source did not give appreciably different travel times. Therefore, the conclusion can be drawn that relatively high temperatures in the stratosphere account for the sound in the outer zones, as suggested by Whipple [30].

FIG. 9.—Sound intensities observed after an explosion near Vergiate, Italy. (*After Oddone.*)

Annual changes in the travel-time curves (Fig. 12), and the annual period of the radius of the zone of abnormal audibility, are caused by (1) annual changes

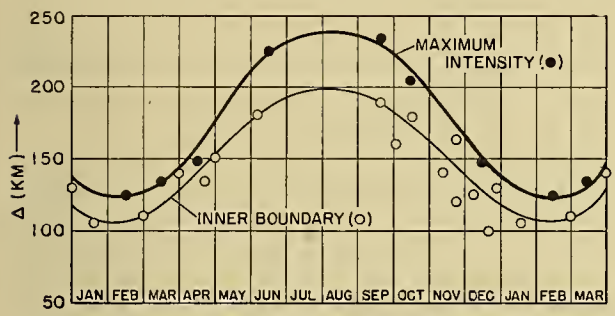


FIG. 10.—Yearly period of the distance from the source at which the first ring of audibility begins (circles and thin curve) and of the distance of the maximum number of reported observations (dots and thick curve) based on data collected by Wegener [28, 29].

in the speed and direction of the wind, or (2) an increase in temperature above the value near the ground at elevations varying, in central Europe, between about 30 km in late winter and 40 km in late summer. Either (1)

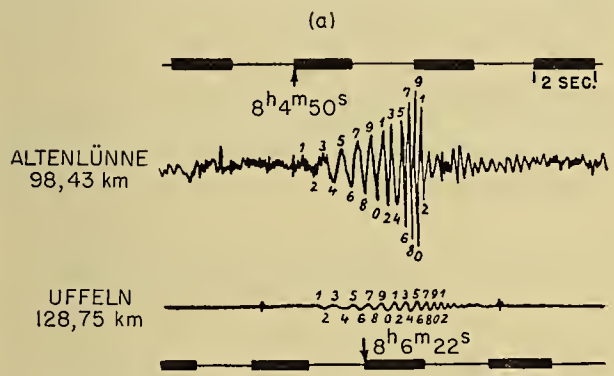


FIG. 11*a*.—Direct sound waves, recorded by Kühl's undograph, showing dispersion. (*After Schulze* [27].)

or (2) or both operating jointly may produce this effect. Since the travel times for the rays arriving in the second abnormal zone were always close to twice the travel time at half the distance in the first zone (Fig. 12) it was concluded that the second ring is produced by rays reflected at the surface of the earth [11, 29]. Similarly, the following zone (travel-time curve *c* in Fig. 12) is probably due to twice-reflected waves. Still later phases may be due to waves which left the ground under too small an angle of incidence to be turned back in the warm layer near 55 km, passed upward into the colder layer above, and finally were turned back (Fig. 14) in the layers at an elevation about 100 km [5], where the temperature increases beyond that near 55 km.

Calculations of the temperature in the part of the stratosphere in which the sound waves are turned down to produce the first abnormal zone can be made as follows [10, 11]. First, equation (7) is used to calculate the sound velocity in the troposphere, and possibly in the lower part of the stratosphere. Equation (17) gives the angles of incidence corresponding to the travel-time curve (*b* in Fig. 12); usually they change little with distance. Equation (14) gives the horizontal distance

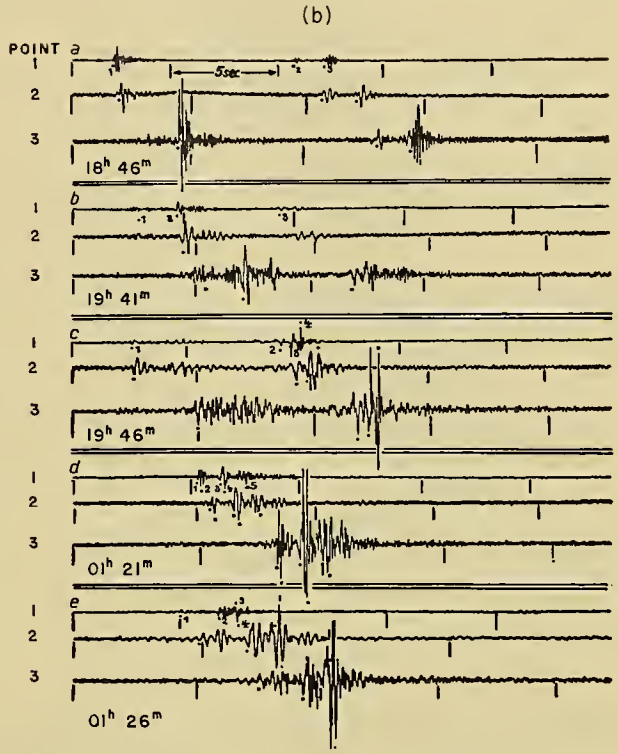


FIG. 11*b*.—Records of sound waves at a distance of about 200 km from the source at three points (distances from each other about 450, 910, and 845 m, respectively) at different times between July 21, 1927, 6:46 P.M. and July 22, 1:26 A.M.; source near Jüterbog, recorded near Wurzbach, Thüringen, in Germany. (*After Meisser* [19].)

corresponding to the ray section between the ground and the level for which the sound velocity has been calculated, and equation (16) the corresponding travel time. Twice these distances and times are subtracted

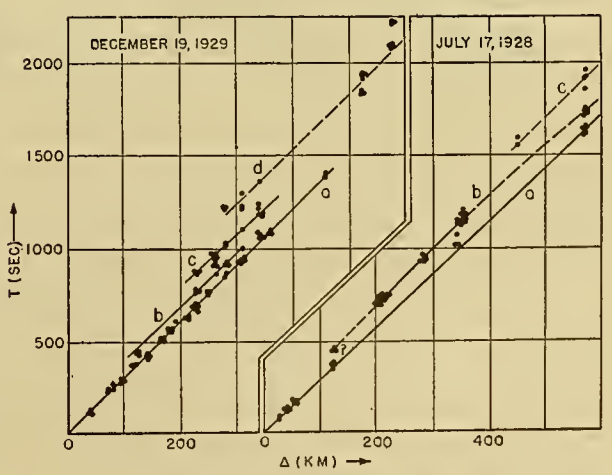


FIG. 12.—Travel-time curves from explosions near Jüterbog, Germany, in summer (July 17, 1928) and in winter (December 19, 1929). Curves *a*, *b*, *c*, and *d* apply to the first, second, third, and fourth zones of abnormal audibility, respectively. (*After Gutenberg* [10].)

from selected observed values of the travel-time curve considering the calculated angles of incidence at the ground. This results in a travel-time curve for rays

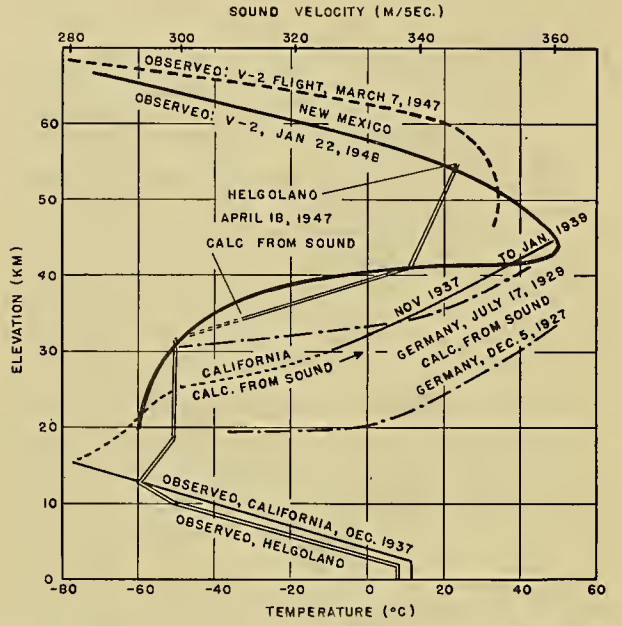


FIG. 13.—Temperature and sound velocity in the atmosphere from various sources.

at the level where the temperature observations end. Equation (18) then permits the calculation of the highest point H^* above the level of reference reached by a given ray; the wave velocity at this point is equal to the apparent velocity at the distance where the ray arrives. In this way the velocity (and consequently the temperature) for a number of points in the stratosphere can be found as far as the temperature increases with height. If necessary and possible, corrections for the wind should be made. The upper limit of height for which results can be found is given either by the increasing absorption or by a decrease in wave velocity with height (usually as a consequence of a decrease in temperature); both may be involved. Figure 13 gives a few results. The temperatures, calculated from the sound observations, agree with the observations from V-2 data [22] within the limits of error.

The yearly period of the radius of the abnormal zone (Fig. 10) and of the travel times (Fig. 12) is affected by the yearly period of the wind, but it is caused mainly by the annual period of the temperature (and thus by the similar period of the ozone content) at elevations between about 25 and 60 km. Annual changes in the direction of the wind near and above the tropopause shift the whole zone in the direction of the prevailing wind, but cannot explain the annual expansion and contraction of the zones which has been observed.

Observations of the amplitudes of sound waves through the stratosphere are very scarce. The maximum near the inner boundary of the abnormal zone results from the concentration of energy connected with the cusp of the travel-time curve at the minimum distance (Fig. 14) reached by the rays [11], where the very large values of $di/d\Delta$ produce a focal point according to equation (32). The intensity of the sound waves there may be so large that windowpanes are broken [14]. The records of the Helgoland explosion [4] showed good agreement between the observed sound-intensity and the energy calculated from equation (32). However, Cox [4] has pointed out that the decrease in short waves relative to the long waves with increasing distance (to be expected from the increase in absorption, see Fig. 3) is not confirmed by these records. The differences in recorded periods are relatively small and may be a consequence of local effects.

In very large explosions, another group of waves is recorded. The best example is furnished by barograph records following the explosion of Krakatao in 1883; the pressure waves circled the earth several times, with

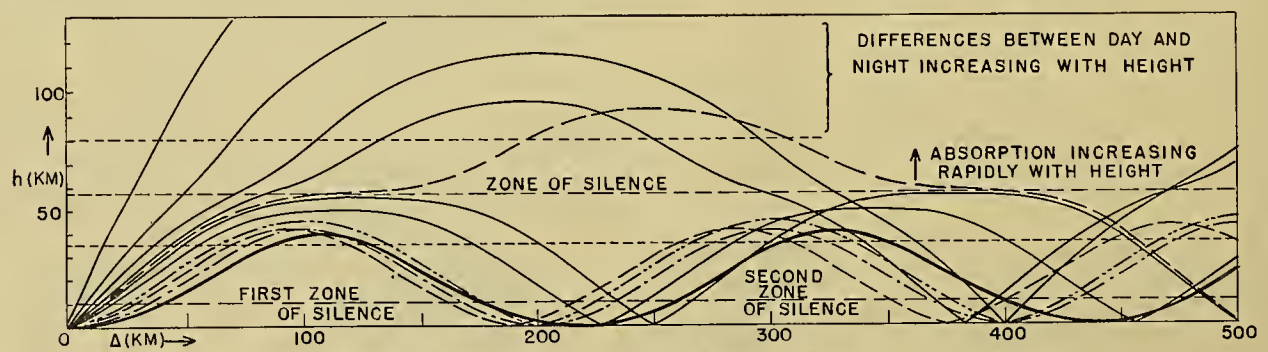


FIG. 14.—Typical paths of sound waves in the atmosphere. (*Extended from Gutenberg* [11].)

a mean velocity [24] of 314.1 m sec^{-1} (measured along the surface of the earth); the mean value from recorded sound waves travelling three times around the earth eastward was about 320 m sec^{-1}, westward about 304 m sec^{-1}. The velocity of 314 m sec^{-1} corresponds to a temperature of −28C. Similar waves with a velocity of 301 m sec^{-1} (not corrected for wind effect) were recorded at Tucson, Arizona, in southern California, and in Nevada [12] after the explosion of the atomic bomb in New Mexico in 1945. This group of waves followed, after many minutes, the waves discussed previously, and carried the largest amplitudes. At Pasadena, it was strong enough to record on the strain seismograph as a single wave having a period of about 12 sec; apparently, the arriving pressure wave com-

pressed the hill on which the Seismological Laboratory is located. The waves are probably of the type discussed theoretically by Pekeris [23].

If the source of the sound is not at the surface of the earth, sound paths can easily be constructed by using equations (10) to (17). For a source in the tropopause, the rays follow patterns similar to those constructed by Ewing [7] for the "tropopause" in the ocean.

PROBLEMS FOR FURTHER RESEARCH

Results concerning the following problems would aid most in the interpretation and use of recorded sound waves passing through the atmosphere:

1. Determination of the velocity and absorption of sound waves in rarified air at pressures down to 0.01 mb.

2. Effect of wind at various levels up to 100 km on sound propagation. (Some authors probably overestimate such effects, others underestimate them.)

3. Effects of the wind component perpendicular to the direction of the sound propagation.

4. Determination of additional travel-time curves for sound waves refracted in the stratosphere (*a*) in various latitudes, (*b*) their annual period, (*c*) their diurnal period (no clear period has been found), (*d*) correlation of results under (*a*) to (*c*) with periodicities of ozone content in the "ozonosphere."

5. Change in frequencies prevailing in sound waves with distance from the source; effects of selective absorption.

6. Indications of dispersion of sound waves under various conditions. (Present experiments do not indicate dispersion.)

7. Theory of free pressure waves in the atmosphere; extension of the theory of Pekeris [23] to other models, considering the most recent data on temperature in the stratosphere and in the ionosphere.

8. Properties of sound waves refracted at levels near 100 km [4] and of free pressure waves [12, p. 329].

9. Natural pressure waves in the atmosphere in various latitudes, on islands, near coasts, and far inland (including use of tripartite stations with base lengths of about ¼ of the wave length of the waves to be studied, see Fig. 7); causes of such waves [13] and their possible use in weather forecasting.

REFERENCES

1. Atanasoff, J. V., Snavely, B. L., and Brown, J., "A New Instrument for Subsonic Frequency Measurements" (abstract). *J. acoust. Soc. Amer.*, 20: 222–223 (1948).
2. Baird, H. F., and Banwell, C. J., "Recording of Air-Pressure Oscillations Associated with Microseisms at Christchurch." *N. Z. J. Sci. Tech.*, 21: 314B–329B (1940).
3. Benioff, H., and Gutenberg, B., "Waves and Currents Recorded by Electromagnetic Barographs." *Bull. Amer. meteor. Soc.*, 20: 421–426 (1939).
4. Cox, E. F., "Abnormal Audibility Zones in Long Distance Propagation through the Atmosphere." *J. acoust. Soc. Amer.*, 21: 6–16 (1949). Correction, *ibid.*, p. 501.
5. —— and others, "Upper-Atmosphere Temperatures from Helgoland Big Bang." *J. Meteor.*, 6: 300–311 (1949).
6. Emden, R., "Beiträge zur Thermodynamik der Atmosphäre." *Meteor. Z.*, 35: 13–29, 74–81, 114–123 (1918).
7. Ewing, M., and Worzel, J. L., "Long Range Sound Transmission." *Geol. Soc. Amer., Mem.* No. 27 (1948). (See p. 19)
8. Gutenberg, B., "Propagation of Sound Waves in the Atmosphere." *J. acoust. Soc. Amer.*, 14: 151–155 (1942).
9. —— "Die dynamische Vergrösserung von Schallregistrierinstrumenten für andauernde Sinuswellen." *Beitr. Geophys.*, 26: 34–36 (1930).
10. —— "Die Schallausbreitung in der Atmosphäre," *Handbuch der Geophysik*, Bd. 9, SS. 89–145. Berlin, G. Borntträger, 1932.
11. —— "Die Schallgeschwindigkeit in den untersten Schichten der Atmosphäre." *Z. Geophys.*, 2: 101–106 (1926).
12. —— "Interpretation of Records Obtained from the New Mexico Atomic Bomb Test, July 16, 1945." *Bull. seism. Soc. Amer.*, 36: 327–330 (1946).
13. —— and Benioff, H., "Atmospheric-Pressure Waves near Pasadena." *Trans. Amer. geophys. Un.*, 22: 424–426 (1941).
14. Gutenberg, B., and Richter, C., "Pseudoseisms Caused by Abnormal Audibility of Gunfire in California." *Beitr. Geophys.*, 31: 155–157 (1931).
15. Hardy, H. C., Telfair, D., and Pielemeier, W. H., "The Velocity of Sound in Air." *J. acoust. Soc. Amer.*, 13: 226–233 (1942).
16. Hergesell, H., und Duckert, P., "Die Ergebnisse der Sprengungen zu Forschungszwecken in Deutschland vom 1. April 1923 bis zum 30. September 1926." *Arb. preuss. aero. Obs., Wiss. Abh.*, Bd. 16 (B), 55 SS. (1927).
17. Herglotz, G., "Über das Benndorfsche Problem der Fortpflanzungsgeschwindigkeit der Erdbebenstrahlen." *Phys. Z.*, 8: 145–147 (1907).
18. Kölzer, J., "Die Schallausbreitung in der Atmosphäre und die äussere Hörbarkeitszone." *Meteor. Z.*, 42: 457–463 (1925).
19. Meisser, O., "Der Einfallswinkel des anormalen Luftschalls." *Z. Geophys.*, 3: 285–292 (1927).
20. Morf, H., "Ueber den Einfluss der meteorologischen Zustände der Troposphäre auf die Ausbildung der anormalen Schallzone." *Ann. schweiz. meteor. Zent-Anst.*, Anhang, 36 SS. (1918). (With bibliography)
21. National Advisory Committee for Aeronautics, *Tentative Tables for the Properties of the Upper Atmosphere*, prepared by C. N. Warfield, Washington, D. C., 1946.
22. Naval Research Laboratory, Rocket-Sonde Research Section, *Curves of Temperature vs. Altitude over White Sands, New Mexico.* 1948.
23. Pekeris, C. L., "The Propagation of a Pulse in the Atmosphere. Part II." *Phys. Rev.*, 73: 145–154 (1948).
24. Pernter, J. M., "Der Krakatau-Ausbruch und seine Folge-Erscheinungen." *Meteor. Z.*, 6: 329–339, 409–418, 447–466 (1889).
25. Quigley, T. H., "An Experimental Determination of the Velocity of Sound in Dry CO_2-Free Air and Methane at Temperatures below the Ice Point." *Phys. Rev.*, 67: 298–303 (1945).
26. Schrödinger, E., "Zur Akustik der Atmosphäre." *Phys. Z.*, 18: 445–453 (1917).
27. Schulze, G.-A., "Luftseismik" in *Naturforschung und Medizin in Deutschland 1939–1946* (*FIAT Rev.*). Wiesbaden, Dieterich, 1948. (See Vol. 18, Pt. 2, pp. 66–71)
28. Wegener, A., "Akustik der Atmosphäre" in *Müller-Pouillet's Lehrbuch der Physik*, Bd. 11. Braunschweig, Vieweg, 1928. (See pp. 171–198)
29. —— "Die äussere Hörbarkeitszone." *Z. Geophys.*, 1: 297–314 (1925).
30. Whipple, F. J. W., "The High Temperature of the Upper Atmosphere as an Explanation of Zones of Audibility." *Nature*, 111: 187 (1923).

COSMICAL METEOROLOGY

Solar Energy Variations as a Possible Cause of Anomalous Weather Changes *by Richard A. Craig and H. C. Willett* 379

The Atmospheres of the Other Planets *by S. L. Hess and H. A. Panofsky* 391

SOLAR ENERGY VARIATIONS AS A POSSIBLE CAUSE OF ANOMALOUS WEATHER CHANGES

By RICHARD A. CRAIG

Harvard College Observatory

and H. C. WILLETT

Massachusetts Institute of Technology

INTRODUCTION

This article is concerned only with the possible relationships existing between anomalous changes of weather and the variations, in one form or another, of the output of solar energy. There can be no question that the sequence of weather change which is represented by the normal diurnal and seasonal sequences of weather over the globe is to be explained entirely by the corresponding diurnal and seasonal variation of the sun's distance and position in the sky, whether or not the quantitative explanation of this normal variation is entirely satisfactory. There is no suggestion of any normal diurnal or seasonal variation of the output of solar energy which would account for any part of the normal variation of global weather. This is entirely a question of the regularly variable distribution of the solar energy.

As soon as a comparison is made between the irregular, anomalous fluctuations of world weather patterns and the irregular variability of the output of solar energy, the evidence of definite relationship between the highly complex variations of these two patterns of change becomes at many points contradictory, indecisive, and highly controversial. In our present state of ignorance as to (1) the specific nature and amount of the solar energy changes, (2) the specific effects which this irregular solar activity produces in the higher atmosphere, and (3) the mechanism of interaction between the higher and the lower atmosphere, it is quite impossible conclusively to prove or disprove the effective influence of irregular solar energy variations on the anomalous fluctuations of the global weather patterns. The evidence for this influence must inevitably depend to a large extent on inference and on the elimination of possible alternative explanations of observed effects. Such evidence is of necessity somewhat subjective in character, and liable to some stretch or strain in its interpretation. In the following discussion an effort will be made to maintain objectivity in the presentation and judgment of the evidence, but the interpretation is intended to be sympathetic to the view that irregular solar activity is the primary factor in the control of anomalous weather fluctuations.

To facilitate the presentation of irregular solar activity as the predominant factor in controlling the anomalous fluctuations of weather and climate, the following discussion is divided into three sections, as follows:

1. The nature, periodic character, and observed effects in the higher atmosphere of irregularly variable solar activity, which presumably is causally related to anomalous weather fluctuations.
2. The periodic character and geographical pattern of anomalous weather fluctuations, which are presumably related to variable solar activity.
3. The applicability of variable solar activity as the primary explanation of anomalous weather fluctuations.

THE NATURE AND PERIODIC CHARACTER OF VARIABLE SOLAR ACTIVITY

Systematic studies of the sun have revealed several characteristic features that vary semicyclically. It is the purpose of this section to discuss briefly these features, their variations, and the observed effects of the variations on terrestrial phenomena. The only book devoted exclusively to a technical discussion of the sun is that of Abetti [1]. A more recent semipopular book is that of Menzel [15]. Various astronomical texts, for example [2] and [19], give brief discussions of the sun.

The sun's visible surface, or *photosphere*, presents a mottled appearance on telescopic inspection. Many small brilliant *granules* can be detected against a darker background. *Sunspots* and *faculae* are much larger and more stable features of the solar surface. The sunspots are relatively dark areas, roughly circular in form. They usually occur in groups, dominated by two large spots. The individual spots vary in diameter from a few hundred miles to tens of thousands of miles. Faculae, on the other hand, are relatively bright areas, usually occurring in the vicinity of sunspot groups.

The spectroheliograph photographs the sun in the selective radiation of one element, and hence enables us to study the sun at levels above the photosphere. In both calcium and hydrogen light, large bright (relatively hot) or dark (relatively cool) patches are observed, especially in the vicinity of the sunspots. These are called *flocculi*. When the hydrogen flocculi are carried by the sun's rotation to the solar limb, they appear as bright *prominences*, or projections of gas from the sun. At times very intense outbursts, called *flares*, occur. These flares are accompanied by hydrogen emissions detectable at the wave length of the $H\alpha$ line in the visible part of the solar spectrum. Another solar phenomenon, recently noted by Roberts, is the occurrence of *spicules*. These are relatively small and relatively short-lived extensions of gas from the chromosphere, and are most noticeable at high solar latitudes.

Even farther out from the photosphere is a very

tenuous envelope of gas called the *corona*. Recent evidence shows that the corona is extremely hot, having a temperature of about 1,000,000K compared with the photospheric temperature of about 6000K. The cause of this high temperature is still a mystery.

All of the solar phenomena described above vary with time in intensity or character. The variations of all the features seem to be in phase. Only the sunspots have been observed for a long enough period of time to include a large number of cycles of variation. Accordingly, sunspot variation is discussed here as an indicator of solar variation.

The *relative sunspot number* is an index of the number of spots and groups of spots visible on the solar surface. Reasonably reliable records of sunspot numbers extend back to 1749, although the earlier years of this record are much less reliable than the later ones. These records show that the sunspot number has alternately increased and decreased with a period that has averaged about eleven years. The length of this period, however, has varied from seven to seventeen years and the sunspot number has also varied from maximum to maximum.

The beginning of a solar cycle is characterized by the appearance of a few spots near 30°N and S heliographic latitudes. As the cycle progresses, the number of spots increases and the latitudes of most frequent occurrence move toward the solar equator. The largest number of spots occurs when the spot zones are near 16°N or S. After sunspot maximum, the spot zones proceed equatorward and at sunspot minimum there are only a few spots near the equator and a few spots in higher latitudes marking the beginning of a new cycle.

Sunspots possess intense magnetic fields. The polarities of these fields vary characteristically, with a period twice that of the ordinary sunspot cycle. The two large spots in a group, which are called the *leader* and *follower* spots, have opposite polarities. The leader spots in the Northern and Southern Hemispheres also have opposite polarities. Moreover, spots in the same relative position in their groups and in the same hemisphere have opposite polarities from one sunspot cycle to the next.

Along with sunspot variations, other solar phenomena show characteristic changes. Faculae, flocculi, flares, and prominences are more numerous and intense at sunspot maximum, although they are by no means absent at sunspot minimum. The corona is approximately circular at the time of sunspot maximum, but is flattened at the poles near sunspot minimum. The intensity of the coronal emission lines is also greater at sunspot maximum than at minimum, according to recent observations.

The energy output of the sun is difficult to measure at the earth's surface, because of absorption and scattering of sunlight by the earth's atmosphere. Nevertheless, the Smithsonian Institution has for the past forty years undertaken the routine measurement of the *solar constant*. This is the amount of solar energy incident at the outer edge of the earth's atmosphere per unit area and time reduced to mean solar distance. These measurements, of course, do not include solar energy to the violet of about 2900 A, which is completely absorbed in the earth's upper atmosphere. This energy is estimated and included in the solar constant.

The value of the solar constant is near 1.94 cal cm^{-2} min^{-1}. C. G. Abbot, who has supervised most of the measurements, claims that the solar constant varies from time to time by a few per cent of the mean value. This claim has been disputed by many other scientists, who feel that the uncertainties involved in the measurements are at least as great as the suspected range of variation. In any case, Abbot's work gives an upper limit to the amount of solar variability during the past forty years in the part of the spectrum available for routine measurement. The measuring program of the Smithsonian Institution is described fully in the *Annals* of the Smithsonian Institution and in numerous papers of the *Smithsonian Miscellaneous Collections*.

Abbot has claimed that the variations in the solar constant primarily reflect large variations in the violet and ultraviolet part of the spectrum. This claim seems to be verified by some measurements of Pettit [16]. Pettit measured the ratio of solar radiation at 3200 A to that at 5000 A. Over a period of years from 1924 to 1931, this measured ratio varied by about 40 per cent of its mean value, presumably reflecting variability in the ultraviolet. Unfortunately, observing conditions were not ideal and atmospheric factors were suspected to be the cause of at least part of this variation. On the other hand, none of the specific tests applied by Pettit showed instrumental or atmospheric effects of this magnitude. To some extent, at least, the ratio varied in a manner similar to that of the sunspot number.

Farther into the ultraviolet, below about 2900 A, solar energy is absorbed in the upper atmosphere and cannot be observed at the earth's surface. In the ultraviolet, however, there is abundant indirect evidence of solar variability because of effects on the upper atmosphere. The frequency of occurrence of magnetic storms and of aurorae, as well as the ionization of the upper atmosphere, clearly varies over the solar cycle. The correlation between upper-atmospheric phenomena and sunspot number becomes progressively poorer as the parameters involved are averaged over shorter and shorter periods of time. Thus, there is little or no correlation between daily sunspot numbers and daily values of the various indices of upper-atmospheric conditions. This would make it appear that all spots do not affect the earth directly but rather are correlated with the occurrence of the solar phenomena that do affect the atmosphere. Effects in the upper atmosphere indicate clearly solar variations only in the short wave lengths of hydrogen and helium emission ($\lambda < 1216$ A) and variations in particle emission of the sun. For the wavelength interval 1216–3200 A, there is apparently no evidence to indicate whether the sun's energy is or is not variable.

The present state of knowledge of solar variability, then, reveals little or no variability in the infrared and visible spectrum, perhaps some moderate variability in the violet and near ultraviolet, and almost certainly rather large variability in the far ultraviolet. The solar

variability that exists is related to the sunspot number, at least in a general way. The record of sunspot number in the past 200 years, therefore, gives some idea of the character of solar fluctuations. A study of the record shows great irregularity. Although the average time from maximum to maximum has been eleven years, the interval between maxima has been exactly eleven years only three times in eighteen cycles and has varied from seven to seventeen years. The annual mean sunspot number at maximum has varied from 46 (1816) to 152 (1947). The maxima appear to have run in cycles of four large values and three small values. This longer-period variation can also be expressed in terms of an 80-year cycle, that is, in successive forty-year periods the average number of sunspots has been alternately large and small.

There is some evidence that during the historical past solar activity has varied considerably more than during the past 200 years. Fragmentary observations and records indicate that during parts of the 13th and 14th centuries (a period of considerable climatic stress in Europe and Asia) solar activity as evidenced by very large and numerous sunspots was probably greater than it has been since, although observations were so few and primitive that no real comparison can be made. On the other hand, between 1672 and 1704 (the 17th and 18th centuries were notably lacking in climatic stress) not a single sunspot was observed on the solar northern hemisphere, so that in 1705 when the observation of such a spot was announced it was considered a surprising fact of extreme scientific interest.

THE PERIODIC CHARACTER AND GEOGRAPHICAL PATTERN OF THE ANOMALOUS WEATHER FLUCTUATIONS

The flow pattern of the general circulation of the earth's atmosphere, by which the world weather pattern is determined, is in a perpetual state of anomalous fluctuation comparable in amplitude to and much more rapid than the normal slow seasonal fluctuation. The periodic character of these anomalous fluctuations of the general circulation has been the subject of extensive statistical analysis. Components of periods ranging from a few days to millenia or geological epochs in length can be detected, but it has not been possible to demonstrate any regular periodicity of real statistical significance in these fluctuations. To all practical purposes, there are no established periods of anomalous fluctuation of the general circulation.

The fluctuations occur in an almost continuous spectrum of periods from the very short to the very long. For purposes of discussion it is convenient to group the anomalous fluctuations of the general circulation (*i.e.*, of the world weather pattern) by length of period of the fluctuation concerned, in five general classes. A discussion of each of these classes follows.

Geological Fluctuations. These refer to the major glacial and interglacial periods during geological time. According to the best geological evidence [7], major glacial epochs, at least during the last billion years, have occurred approximately at quarter-billion-year intervals, but no over-all climatic trend towards more mild or more severe conditions is indicated. Each of these glacial epochs (with the exception of the Pleistocene, which is not yet terminated) continued for many, perhaps as much as 50 million years, being separated by approximately 200 million years of relatively mild interglacial climate. The geological evidence indicates that each major glacial epoch was by no means continuous, but consisted of an extended sequence of periods of maximum glaciation (ice-sheet development) separated by periods of ice-free interglacial conditions. These glacial-interglacial sequences apparently run in cycles of from one hundred thousand to five hundred thousand years duration.

The present Pleistocene Epoch, which apparently has not lasted more than one or two million years, has experienced four such glacial maxima, separated by three interglacial periods, of which the middle and longest one lasted for at least two hundred thousand years [7]. Since the Pleistocene sequence of glacial and interglacial climates is known in much greater completeness and detail than that of the earlier glacial epochs, and since it occurred under essentially the present condition of topography (land and water distribution) which determines our climatic patterns today, only this epoch is referred to in the following discussion of glacial and interglacial climates. It is assumed that any deviation of the earlier glacial-interglacial patterns from those of the Pleistocene period are to be explained by the extensively different terrestrial topography of the earlier periods.

The climatic conditions of a period of maximum glaciation may be characterized essentially as follows:

1. Ice sheets covering up to 30 per cent of the continental area of the Northern Hemisphere, with principal ice sheet centers between latitudes 60° and 65° over northeastern America and over Scandinavia, extending southward over favorably located continental areas to the 40th parallel of latitude.

2. Greatly expanded anticyclonic polar-cap circulations, with a corresponding equatorward displacement of the climatic belts and zonal wind systems, notably of the prevailing storm tracks of middle latitudes.

3. Increased poleward temperature gradient in middle latitudes, with a correspondingly marked intensification of the general circulation, notably of storminess in middle latitudes and of the effective operation of the evaporation-precipitation cycle.

4. Predominance of excessively cool and wet conditions in the middle and lower middle latitudes of maximum storminess; of excessive rainfall in the intertropical convergence zones, and probably of hot dry conditions in a narrow intense subtropical high pressure belt on either side of the equator.

In contrast the climatic conditions of a typical interglacial period may be characterized essentially as follows:

1. Complete disappearance of permanent ice from the face of the globe on land and sea, with only tempo-

rary freezing of shallow portions of the polar seas during the winter season.

2. Complete disappearance of the anticyclonic polar-cap circulations with a corresponding poleward displacement of the climatic belts and zonal wind systems, notably with the contraction of the circumpolar storm tracks of either hemisphere into permanent cyclonic centers over the poles.

3. Great warming of the polar regions, with a correspondingly weakened poleward temperature gradient, decreased circulation, and absence of storminess in middle latitudes.

4. Expansion of the subtropical high-pressure belts with relatively storm-free, dry, and mild conditions prevailing through most of the middle latitudes. Precipitation restricted to the local convective rather than cyclonic type, with a marked weakening of the intertropical convergence, hence with relatively dry conditions in the tropics as well as in the middle latitudes.

This contrast between the typical glacial and interglacial climate is summarized here in its essential detail because it is quite characteristic of, though somewhat more extreme in degree than, the shorter-period climatic, secular, and anomalous fluctuations of the general circulation.

Climatic Fluctuations. These include the considerable fluctuations of climate during postglacial time. The last glacial maximum of the Pleistocene Epoch terminated officially some 8000 years ago, at approximately 6500 B.C. Geological evidence indicates that at that date the recession of both the Scandinavian and the North American ice sheets, and the corresponding amelioration of climate, had reached the point at which conditions stand today. However, during these 8000 years there have occurred very substantial fluctuations of climate [4, 7, 12, 27], running in irregular cycles of from three to five thousand years. In amplitude these fluctuations probably have amounted to at least half of a glacial-interglacial cycle, and according to all evidence they have followed almost identically the glacial-interglacial pattern of climatic change. The evidence is quite clear that during the preceding twenty thousand years of the recession of the last ice sheet, and also during the preceding glacial maxima of the Pleistocene Epoch, a similar considerable secondary fluctuation of climate was superposed on the primary glacial-interglacial cycle, to such a degree that each glacial stage was marked by temporary peaks and recessions.

During postglacial time, in the period of the Climatic Optimum, which lasted from approximately 4000 to 2000 B. C., the world climate very closely approximated the interglacial type. Warm, dry, storm-free conditions prevailed in middle latitudes so that hardwood forests flourished in Scandinavia, in the British Isles, and in much of North America in regions where the climate is far too severe (cold and stormy) for such growth today. The polar regions were warm and stormy, the polar seas free of permanent ice, and the Greenland and antarctic icecaps probably several hundred feet lower than today. Lake levels in Asia, in northern and central Africa, and in the western United States reached their lowest levels of postglacial time, many of them drying up entirely, while most glaciers in middle latitudes completely disappeared. Most of these glaciers today represent new formations, not remnants preserved from the last glacial maximum. If this condition had persisted for fifty thousand instead of only two thousand years, it probably would have constituted a true interglacial period.

At the start of the sub-Atlantic period, which was well established from about 1000 B.C. to A.D. 300, world climate took a strong turn towards the glacial type. Storminess and cold in Europe became extreme, the extensive forests of the preceding Climatic Optimum were replaced by peatbogs, while glaciers in all parts of the world advanced well beyond preceding limits. The Caspian Sea, the nonoutlet lakes of the western United States, and the lakes of northern and equatorial Africa reached their highest postglacial levels, while relatively cool, moist conditions in the lower middle latitudes permitted the development of extensive civilizations in the Mediterranean and on the steppes of central and western Asia in regions which subsequently became too arid for agricultural pursuits. During the period, severe arctic conditions were re-established in the polar regions.

From approximately A.D. 400–1000 the world climate returned to a minor and abbreviated edition of the Climatic Optimum, mild extremes being reached in the higher latitudes during the eighth and tenth centuries. This period marked the peak of the Viking explorations and colonization in Iceland and Greenland. In their small boats the Vikings regularly traversed seas which today would be impassable for them by reason of ice and storms, and their colonies throve by agricultural pursuits in areas of Greenland which are now covered by glaciers. Dryness in the Mediterranean and central Asia led to mass migrations and to the crumbling of many civilizations. The degree of dryness was attested to by tree-ring records and by the lowest lake levels in Asia, Africa, and the western United States since the Climatic Optimum. Relative dryness in tropical regions which at present have heavy rainfall was evidenced by the development of the Mayan civilization of Central America and the great city of Angkor in French Cambodia. The sites of both of these civilizations were reclaimed by the jungle during subsequent centuries of wet climate.

The eleventh and twelfth centuries saw a return of the glacial type of climate, which during the 13th and 14th centuries reached a peak of great storminess and climatic stress in Europe and other regions in higher middle latitudes. This period repeated all of the characteristics of the sub-Atlantic period to a moderate degree. Since the 14th century there has been some amelioration of conditions, particularly during the first half of the 17th and last half of the 18th centuries. This was followed by increased severity of climate during the early nineteenth century, and a turn for the better from 1880 to the present. Although the climatic fluctuations of the past 500 years have been relatively slight, the

prevailing type of climate has been that of postglacial severity rather than mildness. The important fact about the postglacial period is that in the relatively short time of 8000 years the world climate has swung through two cycles of change which in amplitude are a substantial part of the much longer typical glacial-interglacial cycle of the Pleistocene Epoch.

Secular Fluctuations. These refer to cycles of change or trends of weather and climate which are completed within a single century. Examples are the so-called *Brückner Cycle*, the single or double (Hale) sunspot cycle, and the world-wide trend since 1880 towards higher temperatures which has progressed poleward and reached a crest in the higher latitudes during the past two decades. Many investigations of these secular fluctuations of the world weather pattern have been undertaken. Here it need only be mentioned that these fluctuations are small in amplitude compared with both the climatic and the geological fluctuations, but they do show notably similar characteristics as to latitudinal displacement of the zonal wind systems and prevailing storm tracks, and as to the corresponding severity of cyclonic storm activity in middle latitudes. This latitudinal shift of storm tracks and belt of maximum cyclonic precipitation is shown quite clearly by the last, and statistically best-established, of Brückner's rainfall cycles (1855–1885), by a number of Tannehill's secular trend graphs of pressure, temperature, and rainfall [24], by the trend since 1880 towards warmer and drier conditions in progressively higher latitudes, and by many of the double sunspot-cycle phase-change patterns of pressure, temperature, and rainfall [27, 28]. The basic similarity of the geographical patterns of secular changes and trends of weather anomalies to the patterns of climatic and geological changes is considered to be highly significant for any physical interpretation of these fluctuations.

Anomalous Fluctuations. This class is made up of the anomalous irregularly cyclical fluctuations of the general circulation from week to week, from month to month, or from season to season within the year. In recent years considerable synoptic and statistical analysis has been directed toward the determination of the essential character and the most significant parameters for the diagnosis and prognosis of these fluctuating patterns of the general circulation. As a result of these studies there was developed the concept of the high- as opposed to the low-index pattern of the general circulation as an expression of the essential character of the opposite extremes between which the pattern typically fluctuates. These fluctuations usually run in cycles of from three to seven or eight weeks. During some periods the opposite type patterns reach more extreme development than during others, and during some seasons or years or even longer periods (secular, climatic, and geological) one type or the other is largely predominant. The most noteworthy feature of these two contrasting type patterns is that they are world-wide in character, so that they appear to express the operation of a mechanism of basic significance in all of the longer-period fluctuations of the general circulation [17, 23].

Originally the single parameter by which the high- and low-index patterns of the general circulation were defined was that of the strength of the sea-level zonal westerly winds between latitudes 35° and 55°, but much statistical and synoptic analysis [18, 28] has indicated that there are other parameters which are probably of more significance.

In the light of these statistical studies the change from a high- to a low-index pattern (the reverse of a change from low to high index) is best expressed by the significant parameters of the state of the general circulation, other than by a weakening of the sea-level zonal westerlies, in the following terms:

1. An intensification and expansion of the polar anticyclonic circulations, with a corresponding equatorward displacement of the zonal wind systems and the related climatic zones, notably of the zonal westerlies and the prevailing storm tracks.

2. An intensification of the cellular (as opposed to the zonal) pressure and wind pattern. At sea level this trend is indicated by a splitting of the major cyclonic and anticyclonic centers of action, with a north-south rather than east-west orientation of the major axes. At upper levels the trend is indicated by an increased amplitude and shortened wave length of the trough-ridge pattern, with a tendency toward the formation of closed anticyclonic centers in the higher latitudes and closed cyclonic centers in the lower. The result is an increase of the latitudinal exchange of air masses, and of storminess and extreme air mass contrasts in lower middle latitudes.

3. Initially an increased poleward temperature gradient in the lower troposphere in middle latitudes, hence, with the increased austausch, a marked increase of poleward transport of heat and energy.

It is immediately apparent, from the above summary of the high-low index contrast, how strikingly this contrast of the general circulation pattern resembles the interglacial-glacial, and less extreme climatic and secular, contrasting world weather patterns.

Daily Fluctuations. These are essentially the day-to-day progression of pressure centers, with related air mass and frontal phenomena, as observed on the daily weather maps. Essentially the daily fluctuations of the weather pattern are local rather than global in character, for only relatively small-scale circulation phenomena can run through a cycle of change in a day or two. However, it must be recognized that the sum total of these small-scale fluctuations integrated over a period of days or weeks constitute the world-wide patterns of weather change discussed above as the anomalous changes. When the anomalous changes are progressing rapidly, the trend of this progression, and the contribution to it of the daily local changes, becomes quite evident. In particular, the progressive effects of any sudden impulses (solar) to the anomalous fluctuations of the world weather pattern must be traced in daily pattern changes. Such effects are noted below in the reference to Duell and Duell's study [6] of the effects

of sudden solar outbursts on the large-scale pressure changes.

HYPOTHESES OTHER THAN SOLAR VARIABILITY AS PRIMARY FACTORS CONTROLLING THE ENTIRE SPECTRUM OF ANOMALOUS WEATHER CHANGES

The irregular weather changes described above have many characteristics that seem to indicate solar variability as a primary causative factor. These arguments cannot be considered as definite proof of the existence of an atmospheric reaction to solar variability. At least, however, they point out the importance of further studies of the question.

Perhaps the most striking feature of these changes is the fact that the entire spectrum of weather variations reveals a similar oscillation between two definite types of weather pattern. The severe, stormy weather of the glacial epochs, glacial stages, glacial substages, peat-bog period, and stormy centuries could be satisfactorily explained in terms of the predominance over varying periods of time of the low-index patterns that are apparent in our weather today. Thus, all these weather patterns indicate an expansion of the circumpolar vortex, with extension into middle latitudes of polar conditions, a retreat southward of the zone of westerlies and the storm tracks, a contraction and equatorward movement of the subtropical high-pressure systems, and probably a contraction and intensification of the intertropical convergence zone. On the other hand, the relatively mild and dry weather which has predominated at various times in the earth's history is similar in nature to our mild seasons characterized by high-index patterns.

This similarity of pattern is strongly indicative of similarity of the disturbing impulses for all fluctuations. The necessity of some *universal* disturbing impulse which varies irregularly in long as well as short cycles is indicated by the fact that the amplitude of the fluctuations increases with the period. No random combination of localized centers of disturbance of the cellular patterns of the circulation could be expected to produce such globally consistent and integrated patterns of change.

Unfortunately, the physical basis for present-day changes between high-index and low-index weather patterns is not understood. Rossby [17] has shown that relatively high pressure near the poles (anticyclonic polar vortex) is dynamically unstable and that cold anticyclonic domes tend to move southward with a resulting increase in the meridional austausch.

Starr [23] considers the latitudinal transport of *momentum* and *heat* and arrives at the conclusion that relatively high pressure in higher latitudes is a necessary condition for, but does not require, poleward transport of heat and kinetic energy. Thus Rossby's interpretation is essentially dynamic and Starr's is thermodynamic. Correlation studies based on Starr's hypothesis have yielded coefficients considerably higher than those based on any other hypothesis of the mechanics of the general circulation.

Irregular weather changes also seem to be world-wide in extent. Certainly the glacial epochs and probably the glacial stages have been essentially simultaneous throughout the Northern Hemisphere. This also applies to the climatic fluctuations. For the secular and week-to-week changes, we cannot be so certain, but even there evidence indicates the same tendency. The difficulties of calibrating the geological time-scale in the Southern Hemisphere with that in the Northern Hemisphere are great; however, the evidence now available indicates the coincidence of the gross weather variations in the two hemispheres.

If the two premises are accepted that irregular weather variations of all periods have a common cause and tend to be world-wide in extent, it is natural to look to solar variability as the controlling factor. All that we know about solar variability indicates a striking similarity between the patterns of solar changes and of weather changes. The sun is capable of short-term changes in emission of ultraviolet light and charged particles. It also shows a secular change in these emissions that is only roughly periodic. The meagre evidence available from sunspot observations indicates also changes extending over centuries that might correspond to climatic weather changes. Quite probably the sun has varied even more in the geological past than in the immediate past where records are available.

In spite of these considerations, many previous investigators have suggested other causes for various types of weather changes. Each of these suggestions has encountered serious objections. None, except solar variability, is capable of explaining *all* irregular weather variations. Nevertheless it is of interest to consider the various weather changes in turn and to discuss the alternative suggestions that have been offered.

Geological Weather Changes. The cause of the ice ages has been and still is a subject of lively debate. Of the many theories that have been propounded, three have seemed particularly attractive and have received the most attention. These are (1) the *theory of distribution of insolational heating*, (2) the *theory of mountain or continent building*, and (3) the *theory of solar variability*.

1. The theory of distribution of insolational heating holds that variations in the earth's orbital elements (eccentricity, inclination of the earth's axis, and precession) affect the distribution of solar radiation over the earth and thus cause cycles of weather. The cyclic variation of these elements is such as to predict the pattern of Pleistocene glaciation (four glacial stages and three interglacial stages), and at one time was used to establish a Pleistocene chronology. The arguments against this theory are:

a. These same cyclic changes carried back beyond the Pleistocene Epoch predict similar patterns of glaciation, whereas pre-Pleistocene time was relatively free of glaciation for some 250 million years.

b. These solar effects require that changes in the Northern and Southern Hemisphere be out of phase, contrary to present opinion and available evidence.

c. This solar variation does not affect the total

amount of radiation received in either hemisphere in a year, and scarcely affects the total annual insolation at a given latitude.

d. The effects are extremely small according to recent calculations by Simpson [22].

2. The theory of mountain or continent building is based on the idea that glaciation is caused by periods of continental uplift and mountain building. Continentality presumably leads to cold winters, and mountains focus the precipitation orographically and give cool summers. Even today, evidence is abundant that these factors are favorable for glaciation. However, it seems likely that these factors, while they probably determine the patterns of glaciation and perhaps are necessary conditions for glaciation, are not in themselves sufficient. Simpson [22] has very neatly pointed out that the Southern Hemisphere, with only about one-half as much land as the Northern Hemisphere, actually averages 2C colder in its annual mean, a maximum effect of 3C being reached at latitude 60° where the Northern Hemisphere is 60 per cent land, the Southern 0 per cent. Moreover, geological evidence indicates that periods of continental uplift have occurred in some cases without subsequent glaciation, and that in other cases the associated glaciation has followed only after millions of years. Furthermore, the glacial and interglacial stages and substages of the Pleistocene ice age have no associated geological changes.

Climatic Weather Changes. Climatic changes, with their smaller amplitudes, have received less attention than geological changes. Of the theories discussed above for the geological changes, only the theory of solar variability could carry over to explain climatic changes. For example, C. E. P. Brooks, a proponent of the theory of continental uplift as a cause of geological variations, has turned to solar variability as a cause of climatic changes [4]. As was pointed out above, climatic oscillations have been similar in character and intermediate in degree to geological variations, on the one hand, and to secular or week-to-week changes on the other. The inability of a given theory to explain all the changes is therefore a weakness in that theory.

Secular Weather Changes. No cause of the secular weather changes, other than possible solar effects, has been seriously offered. Some of the statistical studies attempting to relate secular weather changes to the sunspot cycle are reviewed below (see pp. 386–387).

Week-to-Week Weather Changes. In recent years week-to-week changes in the general circulation have been studied extensively by many meteorologists. In particular the Weather Bureau–M.I.T. Extended Forecasting Project under Willett's direction has conducted an exhaustive statistical study of the week-to-week pressure changes. Using a linear-correlational technique, Willett [26] has compared the zonal index with parameters that describe other segments of the general circulation, such as the zonal easterlies, the subtropical easterlies, and the intensity of meridional interchange. These studies yield a number of consistent and statistically significant contemporaneous correlations between the various parts of the general circulation. However, they fail to yield any consistently significant lag correlations.

The implication of this result is that the causes of large-scale weather changes are not to be found in the pre-existing state of the general circulation. Possible explanations are that:

1. The changes are purely chance fluctuations.
2. The crudeness of the statistical techniques has masked some real relationships.
3. The control mechanism is in the earth's atmosphere, but is outside the region thus far studied—the Northern Hemisphere between the surface and 3 km.
4. The control mechanism is outside the earth's atmosphere, presumably in solar variability.

The first explanation, that of chance fluctuations, is not only unlikely in view of the world-wide and persistent character of the fluctuations, but also represents an extremely undesirable conclusion which could not be accepted until all other possibilities were definitely ruled out. With regard to (2), the statistical methods, while undoubtedly crude, have served to show up certain contemporaneous relationships and could hardly fail at least to indicate the existence of significant lag relationships. Explanation (3) cannot be investigated further at the present time because of lack of data. The external control mechanism, here, as in the cases of other types of irregular fluctuations, remains as a distinct possibility, in that adequate causes of any other nature have not been established.

THE SOLAR HYPOTHESIS AS THE PRIMARY FACTOR CONTROLLING THE ENTIRE SPECTRUM OF ANOMALOUS WEATHER CHANGES

The discussion of the preceding pages indicates in certain respects the general over-all aptness of the solar explanation for the entire range of irregular weather fluctuations, and certain objections to possible alternative explanations of the individual categories of climatic change. It remains only to present any available evidence favoring the solar explanation of the specific categories of the fluctuation of world weather or climate.

Geological (Glacial-Interglacial) Fluctuations. As indicated above, the work of Simpson and others renders relatively untenable both continentality (terrestrial topography) and the long-period geometrical variation of the distribution of insolation as the primary cause of the fluctuations of climate during geological time. Of the current widely accepted hypotheses by which to account for the glacial-interglacial climatic cycles, only the variable output of solar energy remains.

There are two schools of thought as to the probable solar energy change required to produce an ice age, that of most geologists, as formulated by Flint [7], which calls for a substantial decrease of the solar constant, and that of Simpson [20, 21, 22], which requires a substantial increase of the solar constant. Neither hypothesis visualizes any important selective variation of the energy distribution in the solar spectrum.

It is inevitable that a sufficient decrease of the solar constant would lead to a corresponding lowering of the

mean temperature of the earth's surface by 11C, the approximate difference of mean temperature which most geologists accept as differentiating between a glacial and an interglacial climate. However, this temperature differential is computed on the assumption that total precipitation remains unchanged during the temperature cycle, a totally unjustified assumption.

Simpson [20, 21] pointed out that a general lowering of the earth's mean temperature by a decrease of the solar constant would inevitably decrease both the moisture content of the air and, by reducing the latitudinal temperature contrast, the intensity of the general circulation. These two effects would cause a drastic worldwide reduction of precipitation, which precludes the possibility of an ice age, rather than favoring it. No adequate answer to this objection has been offered by the proponents of a reduction of the solar constant as the primary cause of an ice age.

Simpson's conception of glaciation by increased solar heating rests on the assumption that the effect of this heating is reflected primarily by an intensification of atmospheric circulation, cloudiness, and precipitation. The mean temperature of the entire earth's surface is not decreased, but heating in certain regions (subtropical high-pressure belts) and cooling in others (zones of convergence, storm tracks) is a necessary part of the intensification of the general circulation. Glaciation occurs in regions of increased storminess and localized cooling. Simpson recognizes that if the increased solar heating proceeds far enough, the temperature of the entire atmosphere must rise enough to reverse the trend towards glaciation abruptly, with a quick reaction to a warm, very wet (pluvial) interglacial period. He considers that the Pleistocene Epoch consisted of two solar maxima, each of which corresponded to a double glacial maximum interrupted in the middle by a short pluvial interglacial. The long second interglacial presumably represented a dry period of reduced solar constant.

The principal objection to Simpson's hypothesis of glaciation lies in the paradox of warmer sun and cooler earth, at least locally. Most geologists find it very difficult to accept the possibility of an ice age without a substantial lowering of the earth's mean temperature such as that caused by a decreased solar constant. However, there are several considerations that definitely favor Simpson's hypothesis over the opposite point of view, in particular the following:

1. The increased intensity of the general circulation, with the great increase of precipitation not only in glaciated areas, but also in the lower middle latitudes and in the tropics, which is noted to characterize a glacial (in contrast to an interglacial) climate, certainly fits Simpson's conception of the effect of an increased (as opposed to a decreased) solar constant.

2. The statistical verification of Starr's concept [23] of the low-index pattern (relatively high pressure in the polar latitudes) as a necessary condition for increased poleward heat transport definitely conforms to the low-index character of the glacial as opposed to the interglacial weather pattern. By Starr's hypothesis the glacial low-index pattern manifests the necessity of an increased poleward transport of heat, hence presumably also the occurrence of increased solar heating.

3. Two advantages for Simpson's Ice Age theory, including a partial removal of the primary objection to it, ensue from the assumption that the necessary variation of the solar constant occurs primarily in the ultraviolet, increasing with an increase of the other solar phenomena discussed above (see pp. 379–381). These advantages are:

a. Contrary to the lack of any observational evidence of significant variation in the visible spectrum, there is much direct and indirect evidence of highly erratic fluctuation in the ultraviolet roughly paralleling the sunspot cycle. This erratic sunspot fluctuation apparently runs in irregular cycles which increase in amplitude with the time range and which to some extent parallel the correspondingly irregular weather cycles (see below).

b. The increase of solar ultraviolet does not entail the same increased heating of the earth's surface and entire atmosphere as must a significant increase in the entire visible spectrum. The total absorption of the *increased* ultraviolet in the higher atmosphere may alter and intensify the pattern of the general circulation without sufficient heating of the troposphere to interfere with glaciation as Simpson conceives its occurrence.

Climatic (Postglacial) Fluctuations. There has been little effort made to explain the very considerable postglacial fluctuations of climate. Of the currently accepted explanations of the glacial-interglacial fluctuations, only irregular solar variability can conceivably be applied to the similar postglacial fluctuations of relatively short period. Furthermore, there are a few scattered observations since the twelfth century which indicate that solar disturbances (sunspots) have been exceptionally active during periods of exceptional climatic stress (13th and 14th centuries), and exceptionally inactive during the period of minimum climatic stress (later 16th and earlier 17th centuries). However, in general, solar observations previous to 1750 were entirely inadequate to make possible any satisfactory correlation with climatic conditions. On the other hand, the marked similarity of the postglacial to the geological climatic fluctuations suggests the operation of the same factor of climatic control (probably variation of solar ultraviolet radiation), and there is no alternative explanation which has been suggested.

Secular (Intra-Century) Fluctuations. Only the secular fluctuations of the climatic changes are short enough in period so that extensive comparison with observations of variable solar activity is possible. Many studies of the secular variations and trends of climate have been undertaken, but only solar activity survives as a plausible primary factor of control, and that control is far from proved. Certainly it can be said that the secular fluctuations of the general circulation unmistakably show the influence of the sunspot cycle, but it is equally certain that there exists no demonstrable strict correspondence between solar variability (sunspots) and

the secular weather cycles. The connection between sunspots and weather is so complex and indirect that it remains completely unexplained physically, but unfortunately sunspots are the only index of irregular solar activity for which the record of reliable observation is sufficiently long to be useful for extended correlation. Among the many studies of this kind which have been made, the following might be mentioned briefly as among the more significant:

1. Köppen [13], Walker [25], and others have demonstrated beyond question the reality of an eleven-year sunspot cycle of temperature in many regions, primarily in the tropics, such that lower temperature occurs with the sunspot maximum.

2. Helland-Hansen [11], and others, have indicated that at least across the North Atlantic Ocean prevailing storm tracks tend to be displaced equatorward during periods of high sunspots. Some of Tannehill's graphs [24] indicate a similar latitudinal displacement of storm tracks on the west coast of the United States.

3. Hanzlík [8, 9] finds quite impressive changes of the world-wide distribution of pressure, particularly during the winter season, from the three years centered at sunspot minimum to the three years at the *following* sunspot maximum. The most striking feature of Hanzlík's pattern studies is that, in high latitudes, the pressure changes in effect give a decreasing zonal index as a *major* sunspot maximum commences, and an increasing zonal index as a *minor* sunspot maximum commences. This double sunspot cycle of pressure rise in the higher latitudes is confirmed by Clayton [5], and by Wexler (unpublished manuscript) from the Weather Bureau Northern Hemisphere historical maps. Both of these investigations indicate, on alternating successive minor and major sunspot maxima, anomalous zonal pressure rises which reach their peaks at about 50°N and north of 60°N, respectively. Willett [28] has confirmed Hanzlík's alternating high- and low-index patterns of pressure anomaly on successive sunspot maxima, and has found further that the anomaly patterns of temperature and precipitation manifest corresponding characteristics. In fact the contrasting anomaly trends of pressure, temperature, and rainfall going into successive sunspot maxima correspond in a small degree over both North America and northern Europe to the contrast between the Pleistocene glacial and the interglacial pattern of climate. The occurrence of the high- versus low-index pattern contrast in connection with the double sunspot cycle is particularly suggestive of a primary solar role in this basic change of the general circulation pattern.

4. Baur [3] has produced some impressive statistical evidence, from temperature records of 160 years in the northeastern United States and of 180 years in north-central Europe, that the occurrence of severe winters in both localities shows a marked preference for certain phases of the sunspot cycle, notably just preceding and shortly following the sunspot maximum. According to Baur the years of sunspot minimum and maximum are times of the least probable occurrence of a severe winter in these localities. Other miscellaneous relationships of this type have been found by a number of investigators for specific localities, but they contribute little to any improved understanding of the physical control of the secular weather fluctuations. However, it certainly can be stated that the solar factor is the one which is usually implicated in most of the secular fluctuations of climate.

Anomalous (Weekly, Monthly, or Seasonal) Fluctuations of the World Weather Pattern. The week-to-week and month-to-month anomalous fluctuations of the general circulation pattern manifest quite clearly essentially the same type of high- versus low-index contrast that is evidenced by the secular, postglacial, and glacial-interglacial fluctuations. In fact, it is primarily from these relatively short-period fluctuations that the index types were recognized, and their essential characteristics identified. The recognition of essentially identical characteristics in the longer-period anomalous fluctuations of the general circulation followed. Consequently the question naturally arises as to the extent to which variable solar activity, for which there is so much evidence as the controlling factor in the longer period fluctuations, also exerts primary control on the anomalous fluctuations. This question is particularly pertinent because the major single or double sunspot cycle is very long in comparison to the weekly, monthly, and seasonal anomalous fluctuations. There are a number of observational and deduced facts which have a direct bearing on the probability of such solar control, notably as follows:

1. Besides the longer periods of variable solar activity as evidenced by the so-called eleven-year sunspot cycle and multiples thereof, the sun is almost continuously undergoing a great variety of short-period sudden disturbances and almost eruptive outbursts, as evidenced by faculae, flocculi, flares, prominences, etc., and by a two or three fold or larger variation of sunspot numbers from week to week and month to month (see pp. 379–381). Hence there occur outbursts of solar activity of variable frequency and intensity corresponding to the large-scale variable weather activity. Maris [14], Haurwitz [10], and others have indicated that the sudden emission of short-wave radiation which apparently accompanies this eruptive activity of the sun is capable of producing considerable heating of the higher atmosphere in only a few hours, and thus leads indirectly to substantial poleward displacement of atmosphere in the higher levels, that is, to a rise of sea-level pressure poleward from the latitudinal zone of maximum heating. This reasoning suggests one possibility of a mechanism by which sudden solar outbursts can change the index character (zonal pressure distribution) of the general circulation.

2. As mentioned earlier, Willett's results [26, 27, 28] may be interpreted as an indication of the existence of some primary external control of the anomalous fluctuations of the general circulation, for example, variable solar activity. To test this hypothesis statistically, Willett [28] correlated daily and five-day mean values of a number of indices of solar activity with corresponding indices of the general circulation, but no significant

correlation was found. This negative result constitutes statistical (but by no means conclusive) evidence against an important role of variable solar activity in the anomalous fluctuations of the general circulation. Our qualitative and quantitative knowledge of the physical characteristics of the variable solar emissions, of their direct effects in the higher atmosphere, and of their possible indirect effects in the lower atmosphere, is utterly inadequate either for the certain designation of suitable indices of solar activity, or for any estimate of the probable manifestation of their complex and indirect effects on the circulation of the earth's troposphere. One rather surprising bit of evidence for a direct solar influence on the zonal distribution of pressure on both hemispheres is furnished by highly consistent, but not significantly high, negative correlation between monthly mean zonal pressure anomalies in the lower latitudes, and the contemporary monthly mean solar pyrheliometric values as determined by all recording stations in Europe and North America [28]. In the polar latitudes of the Northern Hemisphere there is a seasonal reversal of the sign of this correlation from summer to winter, a reversal which cannot be checked on the Southern Hemisphere for lack of pressure data.

3. Without comparison the highest correlations which have been found between basic parameters of the general circulation pattern are those based on Starr's concept [23] of the thermodynamics of the general circulation. By this concept the index character of the circulation pattern limits the effective poleward transport of heat. The fact that these correlations pertain to the anomalous fluctuations of the general circulation, with specific implications for the tropical-polar heat balance, is cited as further possible evidence for the direct role of variable solar energy in the anomalous fluctuations.

4. One final item in evidence of direct solar influence on the general circulation pattern must be mentioned as of particular interest in that it indicates a clear day-to-day progression of the disturbing effect. Duell and Duell [6] show that during the winter months (November–February) of years of low sunspot activity (relative sunspot number less than 40), following geomagnetically disturbed days (presumably days of strong particle emission by the sun), pressure falls on the average by about 2 mb at European stations within two to three days, while at stations in the Greenland-Iceland area the pressure rises by an equal or larger amount. Hence by the second day following the disturbance, the pressure gradient from Greenland to northwestern Europe is increased by 5 mb on the average, representing definitely a trend towards a low-index or glacial weather pattern. On the other hand, Duell and Duell found that during the same period of time the pressure variation in northwest Europe after quiet geomagnetic days is in the opposite direction, that is, toward *higher* pressure a few days after the quiet conditions. Craig has carried out similar investigations for many other geographical points in the Northern Hemisphere. He has found that the pressure variations after disturbed geomagnetic days are, to a highly significant degree, negatively correlated with the variations at the same location and during the same period of time after quiet days. However, the patterns of change after disturbed days are not always the same and seem to depend markedly on initial conditions in the atmosphere.[1]

A less extensive investigation of the pressure changes following days of strong solar flares (presumably representing strong outbursts of solar ultraviolet radiation) gives indications of a reverse trend of the pressure pattern, towards a higher index condition. This trend is noted irrespective of season and sunspot number.

Duell and Duell's results need extension and further verification. Taken at their face value, their results appear to be highly significant, in the first place as a clear indication of direct solar influence on the day-to-day weather changes of the basic index type, and in the second place as a possible physical clue to the opposite effects of major and of minor sunspot-maxima on the world weather patterns as first determined by Hanzlík.

SUGGESTIONS FOR FUTURE RESEARCH

As long as there remains a reasonable possibility that irregular weather changes may be linked with solar variations, the subject deserves expanded and careful study. The possibility of long-range weather forecasting as a result of such studies holds the promise of great returns for the efforts involved. The suggestions for future research that follow are based on the conviction that such a reasonable possibility exists.

First of all, an attempt should be made to investigate, independently of the Smithsonian Institution, the variations of the solar constant. It seems rather incongruous to discuss the effects of solar variability, when variations in solar energy over the entire spectral range above 3000 A are not definitely established. Despite the fine efforts of the Smithsonian investigators, probably the present controversy can best be resolved by an independent study. In particular, there should be additional direct determinations of solar variability in the ultraviolet, from the earth's surface in the case of the 3000–3500 A spectral region and from rockets or balloons at shorter wave lengths.

Secondly, the theoretical and physical meteorologist must give attention to the question of how impulses received in the upper atmosphere can affect the troposphere. Duell and Duell's results have no explanation at present. If they stand up under future investigation, they become an important clue to the whole problem of solar-weather relationships.

Thirdly, if the effects of ultraviolet solar variability on the weather become established, there is room for much synoptic and statistical study of interrelationships between the troposphere and the stratosphere. The question is of great interest, in any case, as a strictly meteorological problem.

Finally, and most obviously, the statistical study of day-to-day, week-to-week, and secular weather varia-

1. (*Added in press.*) Consult R. A. Craig, "Atmospheric Pressure Changes and Solar Activity." *Trans. N. Y. Acad. Sci.*, Ser. 2, Vol. 13, No. 7 (1951).

tions as related to solar variability should continue. Certainly, this type of approach is the least economical in that much time has been and will be wasted in correlations of parameters that are actually not related. From such work, however, may come occasional important clues to the understanding and explanation of solar effects. In any case, it is the most direct method of investigation and, in the absence of basic physical understanding of the sun and of our atmosphere, may turn out to yield the most important information.

SUMMARY

The discussion and description of known irregular weather changes leads to some important conclusions:

1. Changes ranging in time scale from weeks to epochs apparently involve oscillations between two extreme types of weather pattern that are observed today.

2. The changes tend to be world-wide in extent.

3. The amplitude of the changes does not decrease as the time scale increases.

4. The variability of the world weather pattern seems to be similar in its quasi-periodic character and variable period to the known variability of the sun.

The specific investigations of solar-weather relationships outlined herein are only a small fraction of those that have been accomplished. Perhaps others equally significant are in existence, but the ones cited are illustrative of techniques and results. Walker's studies [25] show that no significant results can be expected from a simple correlation of annual means of sunspot numbers and weather elements. The work of Hanzlík shows further that atmospheric changes may vary significantly from season to season and even from one sunspot maximum to the next. Tannehill has furnished some striking examples of weather changes that parallel the sunspot curve. Duell and Duell have given the first suggestion of relationships between day-to-day weather changes and specific solar anomalies. Their work also suggests the possibility that more than one solar phenomenon (particle emission versus ultraviolet emission) may affect the atmosphere.

These considerations suggest that glacial epochs, stages, and substages, periods of climatic stress such as the sub-Atlantic period, and stormy centuries, probably result from the predominance over various periods of time of low-index conditions such as are apparent in the weather at the present time. Further, these conditions may result from comparatively great solar activity, particularly with relation to particle emission that affects the upper atmosphere. These tentative suggestions are the best that can be made on the basis of present knowledge, but, unfortunately, they are vague and uncertain. They merely suggest the direction of future research.

The writers feel that the present state of the problem justifies much additional research of observational, theoretical, synoptic, and statistical nature. At present, there is no definite agreement as to the extent of solar ultraviolet variability or even as to the existence of variability in the visible part of the spectrum. These questions should be cleared up as soon as possible, by observation in the latter case and by observation if possible, or by astrophysical reasoning, in the former case. The meteorologists themselves have many problems to investigate of a theoretical, synoptic, and statistical nature.

Intensified research should at least settle the questions of the extent of solar variability and whether it significantly affects the weather. Even definite information that no such relationships exist would be extremely valuable in planning the direction and emphasis of future meteorological research.

REFERENCES

1. Abetti, G., *The Sun*, trans. by A. Zimmerman and F. Borghouts. New York, Van Nostrand, 1938.
2. Baker, R. H., *Astronomy*, 3rd ed. New York, Van Nostrand, 1938.
3. Baur, F., "Zurüekführung des Grosswetters auf Solare Erscheinungen." *Arch. Meteor. Geophys. Biokl.*, (A) 1: 358–374 (1949).
4. Brooks, C. E. P., *Climate Through the Ages*. London, Benn, 1926.
5. Clayton, H. H., "World Weather and Solar Activity." *Smithson. misc. Coll.*, Vol. 89, No. 15 (1934).
6. Duell, B., and Duell, G., "The Behavior of Barometrie Pressure during and after Solar Particle Invasions and Solar Ultraviolet Invasions." *Smithson. misc. Coll.*, Vol. 110, No. 8, 34 pp. (1948).
7. Flint, R. F., *Glacial Geology and the Pleistocene Epoch*. New York, Wiley, 1947.
8. Hanzlík, S., "Der Luftdruckeffekt der Sonnenfleckenperiode. I. Mitteilung: Jahresmittel." *Beitr. Geophys.*, 28: 114–125 (1930).
9. —— "Der Luftdruckeffekt der Sonnenfleckenperiode für die Monate Dezember, Januar, Februar und Juni, Juli, August. II. Mitteilung: I. Dezember, Januar, und Februar." *Beitr. Geophys.*, 29: 138–155 (1931).
10. Haurwitz, B., "Relations between Solar Activity and the Lower Atmosphere." *Trans. Amer. geophys. Un.*, 27: 161–163 (1946).
11. Helland-Hansen, B., and Nansen, F., "Temperature Variations in the North Atlantic Ocean and in the Atmosphere." *Smithson. misc. Coll.*, Vol. 70, No. 4 (1920).
12. Huntington, E., and Visher, S. S., *Climatic Changes*. New Haven, Yale University Press, 1922.
13. Köppen, W., "Über mehrjährige Perioden der Witterung, insbesondere über die 11-jährige Periode der Temperatur." *Z. öst. Ges. Meteor.*, 8: 241–248 (1873).
14. Maris, H. B., and Hulburt, E. O., "A Theory of Auroras and Magnetic Storms." *Phys. Rev.*, 33: 412–431 (1929).
15. Menzel, D. H., *Our Sun*. Philadelphia, Blakiston, 1949.
16. Pettit, E., "Measurements of Ultra-Violet Solar Radiation." *Astrophys. J.*, 75: 185–221 (1932).
17. Rossby, C.-G., "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6: 163–180 (1949).
18. —— and Willett, H. C., "The Circulation of the Upper Troposphere and Lower Stratosphere." *Science*, 108: 643–652 (1948).
19. Russell, H. N., Dugan, R. S., and Stewart, J. Q., *Astronomy*, 2 Vols. Boston, Ginn, 1926–27.
20. Simpson, G. C., "Further Studies in Terrestrial Radiation." *Mem. R. meteor. Soc.*, 3: 1–26 (1928).
21. —— "World Climate during the Quaternary Period." *Quart. J. R. meteor. Soc.*, 60: 425–478 (1934).

22. —— "Possible Causes of Change of Climate and Their Limitations." *Proc. Linn. Soc. Lond.*, 152: 190–219 (1940).
23. Starr, V. P., *A Physical Characterization of the General Circulation*. Dept. of Meteorology, Mass. Inst. of Tech., Report No. 1, General Circulation Project No. AF 19-122–153. Geophys. Res. Lab., Cambridge, Mass., 1949.
24. Tannehill, I. R., *Draft Notes on Weather of Future Years, Part II. Magnitude of Weather Changes Associated with Solar Variations*. U. S. Weather Bureau, Washington, 1945.
25. Walker, G. T., "Correlation in Seasonal Weather. V, Sunspots and Temperature." *Mem. Indian meteor. Dept.*, 21: 61–90 (1915).
26. Willett, H. C., *Final Report of the Weather Bureau—M.I.T. Extended Forecasting Project for the Fiscal Year July 1, 1946–July 1, 1947*. 110 pp., Cambridge, Mass., 1947.
27. —— "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6: 34–50 (1949).
28. —— *Final Report of the Weather Bureau—M.I.T. Extended Forecasting Project for the Fiscal Year July 1, 1948–June 30, 1949*. 109 pp., Cambridge, Mass, 1949.

THE ATMOSPHERES OF THE OTHER PLANETS

By S. L. HESS

Lowell Observatory and Florida State University

and H. A. PANOFSKY

New York University

Introduction

The study of planetary atmospheres is a relatively new field of meteorology. Its main impetus comes from the likelihood that the behavior of the several planetary envelopes, with their varying masses, rotations, constituents, and other physical parameters, may yield important evidence bearing on the general laws which govern our own atmosphere. Much of the information sought through a study of the other planets cannot be obtained by investigation of the terrestrial atmosphere simply because its various parameters of interest are fixed and the effects of their variation cannot be determined.

Beyond the possibility of gaining information that may add to our understanding of Earth's atmosphere lies the hitherto somewhat neglected fact that the behavior of planetary atmospheres, per se, is a proper field of study for meteorologists. One really needs no further justification than the knowledge that there exist several atmospheres, besides Earth's, whose behavior may be observed and interpreted.

In Table I various meteorological parameters are listed for all the planets known to possess atmospheres. They are divisible into two distinct groups. The first three planets listed are relatively small and have atmospheric constituents of moderately high molecular weight. The last four are much larger and have predominantly light-weight constituents. This difference in composition is a natural consequence of the greater masses and far lower temperatures of the outer planets, since these two factors enable them to retain the light gases which the smaller, warmer planets have lost to space.

The most interesting planets in this list are Mars and Jupiter. This is because both planets present, at times, sufficiently large discs to enable us to examine and photograph the atmospheric phenomena which are present. Venus, too, has a large enough disc for this purpose but it is characteristic of her that little visible atmospheric or surface detail exists. Thus our knowledge of Venus' atmosphere is rather small. Saturn, Uranus, and Neptune have successively smaller discs and, in addition, probably have progressively less detail.

Venus

The planet Venus is the closest in size and mass to Earth; she is also our nearest planetary neighbor. Despite this we know relatively little about her atmosphere and surface. Little detail can be discerned in visible light, which leads one to believe that we may be seeing not the actual solid surface but a continuous cloud sheet. This is consistent with the planet's high albedo (0.59). The lack of surface markings prevents a direct determination of the rotation period. Spectroscopic investigations of the rotation indicate that the day on Venus is probably more than three of our weeks [9, p. 317]. It may even be that Venus always presents the same face to the sun, in which case her day would be equal to her year (225 terrestrial days). In any event, we may be certain that the rotation is so slow that atmospheric motions are not subject to significant Coriolis deflections.

There is no question that Venus possesses an atmosphere. This is demonstrated by various physical phenomena [9, p. 318] and by the positive, spectroscopic identification of carbon dioxide as one of its constituents. The amount of CO_2 is estimated to be some 500 times that contained above a unit area on Earth [3, p. 351]. No appreciable amounts of oxygen or water vapor are present. Consequently the basic cloud layer cannot be aqueous; it may be dust. When Venus is photographed in ultraviolet light, large dark and light bands may appear. The nature of these markings is unknown, except that they are undoubtedly manifestations of some sort of cloud because they vary in shape and size from day to day.

Despite our lack of observational knowledge of the circulation of Venus' atmosphere we may be confident, in view of the long period of rotation, that the major feature of the circulation is a rapid exchange of air between the warm, sunlit hemisphere and the cool, dark hemisphere. This would be a direct circulation, and the direction of motion ought to be close to the direction of the pressure-gradient force. This circulation apparently succeeds in transferring considerable quantities of heat to the dark side of Venus because infrared radiation can be detected from the night side. Also important in keeping the dark side warm is the large greenhouse effect supplied by such an enormous quantity of CO_2.

On occasion the far infrared radiation (10 μ) emitted by the warm CO_2 fails to appear [4]. This is certainly not due to a failure of the gas to emit such radiation, but must be due to a transient layer of high cloud which prevents the energy from escaping. The nature of this high cloud layer (as distinct from the lower cloud surface) is unknown, but it may well be condensed carbon dioxide itself. Even on such a warm planet as Venus the saturation vapor pressure of CO_2 can be reached in the

vicinity of only 20 km above the surface if an adiabatic lapse rate is established. This suggests that these clouds and their movements could be detected if one could examine the planet's image in the far infrared. The relationship between these clouds and the ultraviolet markings should also be investigated.

Table I. Values of Meteorological Parameters for Certain Planets

Planet	Mean distance from sun (⊕ = 1)	Length of year (⊕ = 1)	Coriolis parameter at pole ($sec^{-1} \times 10^4$)	Mean radius (⊕ = 1)	Inclination of equator to orbit	Albedo	Surface gravity (⊕ = 1)
Venus	0.72	0.6152	<0.07	0.97	?	0.59	0.86
Earth	1.00	1.0000	1.46	1.00	23° 27′	0.39	1.00
Mars	1.52	1.8808	1.42	0.53	25° 12′	0.15	0.37
Jupiter	5.20	11.862	3.5	11.0	3° 07′	0.44	2.64
Saturn	9.54	29.457	3.3	9.0	26° 45′	0.42	1.17
Uranus	19.19	84.013	3.3	4.0	98°	0.45	0.91
Neptune	30.07	164.783	2.2	3.9	29°	0.52	1.12

Planet	Most important probable atmospheric constituents in order of abundance	Probable value of adiabatic lapse rate (°C km^{-1})	Approximate pressure at "visible" surface (mb)	Approximate temperature at "visible" surface in sunlight (°C)
Venus	N_2?, CO_2	10.0*	>160	50–100
Earth	N_2, O_2, H_2O, A, CO_2	9.8	1000	10
Mars	N_2?, A?, CO_2, H_2O	3.7†	50–100	0
Jupiter	H_2, He, CH_4, NH_3	4.4‡	>50‡	−120
Saturn	H_2, He, CH_4, NH_3	2.0‡	>50‡	−150
Uranus	H_2, He, CH_4	1.5‡	>170‡	
Neptune	H_2, He, CH_4	1.9‡	>350‡	

* For a pure CO_2 atmosphere; in a predominantly N_2 atmosphere the value is 8.0.

† For a predominantly N_2 atmosphere.

‡ Assuming a mixture of six molecules of H_2 to one molecule of CH_4. In such a mixture the amount of He affects neither the adiabatic lapse rate nor the molecular weight.

Mars

The planet Mars is the one which, in meteorological matters, most closely resembles Earth. The Martian day is almost the same length as ours, and the inclination of the polar axis to the plane of rotation about the sun is approximately the same (thus his seasons are much like ours), but the year is almost two of Earth's. Mars is distinguished from Venus and the major planets by the fact that his solid surface is usually clearly visible. Indeed, so little does the atmosphere interfere (except in blue light where a high haze layer is evident) that we suffer from a relative lack of data on such factors as cloud motions, upon which to base a discussion of Martian meteorology.

That Mars possesses an atmosphere has been known for many years, mainly because the polar caps, which wax and wane with the seasons, must have a vapor phase. Various optical scattering phenomena support this conclusion. Recently the first definitely identified constituent of the Martian atmosphere, carbon dioxide, was found by Kuiper [3, p. 335]. The planet has about twice as much CO_2 per unit area as Earth. Water vapor has not been definitely identified, but this is because the amount present is usually too small for spectroscopic detection [1]. The presence of clouds and polar caps points to the presence of water, and the similarity of the reflection spectrum of the caps to that of snow makes the conclusion inescapable. Little can be said about other constituents, except that considerable quantities of nitrogen are to be expected because of this element's universal abundance, its resistance to chemical combination with the planet's crust, and the sufficiency of Martian gravity to prevent its escape.

Radiometric measurements permit estimation of the surface temperature over areas as small as 200 miles in radius. This allows delineation of the general temperature field [2]. The analysis of an extensive sequence of measurements made during 1926 is presented in Fig. 1. The salient features are a belt of high temperature at about lat. 20° in the summer hemisphere, markedly lower temperatures and larger gradients in the winter hemisphere, and an anomalously warm region near lat. 30°S, long. 350°. The first two features are in qualitative agreement with the situation on Earth, while the last feature, which is probably real, is associated with an anomaly in the circulation.

The vertical shear of the zonal winds computed from these data and the thermal wind equation are given in Table II. The rates of increase of west wind with elevation are somewhat smaller than corresponding values on Earth. Since the wind velocity next to the ground must be quite small due to friction, the magnitude of the wind U at some elevation z must be roughly $\frac{\partial U}{\partial z} z$. On

this basis the data of Table II indicate somewhat lower wind velocities than on Earth. This is consistent with the results of observation of the motion of clouds on Mars, which also indicate relatively low wind-speeds.

TABLE II. MEAN VALUES OF THE VERTICAL WIND-SHEAR ON MARS COMPUTED FROM THE DATA OF FIG. 1

Latitude	$\partial U/\partial z$ (m sec^{-1} km^{-1})
15°N–50°N	4.0
15°S–47°S	0.8
47°S–78°S	1.6

The last value in Table II, however, is appreciably larger than the corresponding value on Earth. Nevertheless, this does not mean higher wind-speeds than on Earth, because this value is from the most poleward zone considered in the summer hemisphere, and the temperature at the southernmost boundary (lat. 78°S) is probably influenced by the proximity of Mars' polar icecap. This depresses the temperature at the surface and makes the gradient seem quite high. If one could measure the temperature at, let us say, one kilometer above the surface, the effect of the polar ice would be diminished and the north-south temperature gradient reduced. This is to say that more stable lapse rates predominate near the icecap than elsewhere, a common situation on Earth.

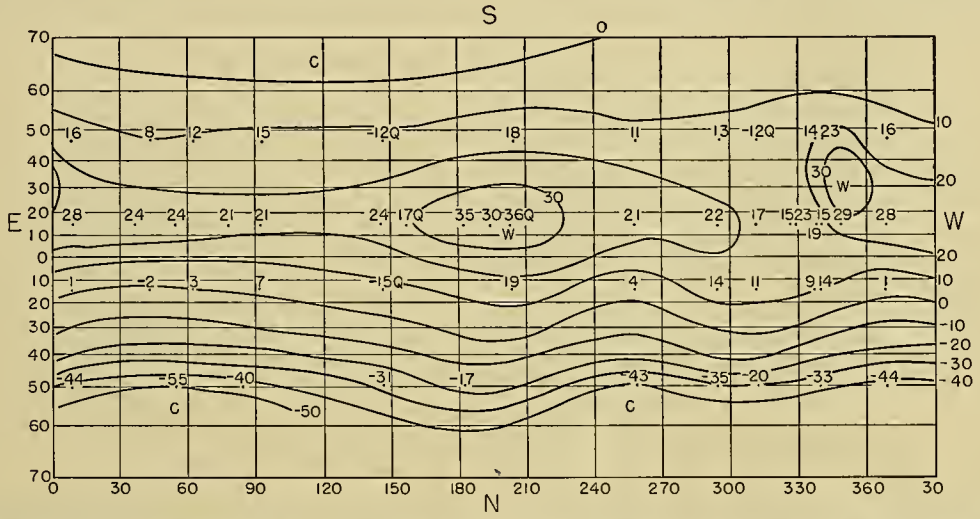


FIG. 1.—The distribution of temperature (°C) on Mars in Northern Hemisphere winter. This is the normal telescopic view with south at the top. To obtain the usual meteorological view, merely invert the page. The observations marked "*Q*" are questionable either because they are interpolated values or because of the presence of clouds. A fifth row of data at 78°S was available for the analysis but does not appear on this projection. Taken from [2].

Figure 2 is a streamline map for Mars drawn in a fashion as consistent as possible with the temperature distribution of Fig. 1, eighteen wind directions deduced from the observed drift of clouds, and general meteorological principles. While these did not suffice to define the flow pattern uniquely, the amount of imagination required was quite small.

The most striking aspect of Fig. 2 is its marked resemblance to terrestrial weather maps. *Mars has an atmospheric circulation very similar to Earth's.* This is by no means due solely to a tendency on the analyst's part to force the pattern into a familiar shape. An examination of the winds will show that the broader aspects of this map could not be altered very much by another independent analysis. The existence of a more or less broad belt of westerlies in both hemispheres is clearly indicated, and there are observations to support the drawing of a pair of subtropical high-pressure belts which are broken up into individual cells. The pairs of winds which form the basis of the two cyclonic storms indicated in the Southern Hemisphere are synchronous in one case, and in the other case were observed on successive nights. Thus there is no way of avoiding an analysis which involves sharp cyclonic shears, that is, fronts, without ignoring the data.

The cyclonic circulation at lat. 35°S, long. 345° is indicated by an easterly cloud drift in a latitude where one would expect to find west winds. But this is the same region in which the temperature distribution indicated an abnormally warm spot. It is probable that we have here an example of a "heat low." This is all the more remarkable because the two pieces of evidence are a cloud-drift direction observed in 1894 and a temperature field determined in 1926. This persistence points to a local peculiarity of the surface which is conducive to high summer temperatures and the formation of a heat low. Examination of a map of the normal surface markings of Mars reveals no visible feature that can explain this phenomenon, but there is a large dark area which forms seasonally in this region. It is present in Southern Hemisphere summer (when these measurements were made) and disappears in winter. It would be valuable to obtain radiometric measures of this area at other Martian seasons to see if the high temperatures vanish with the dark coloration.

Despite these similarities the circulation on Mars differs in important respects from that on Earth. Factors contributing to such differentiation are: (1) the absence of oceans and sharp mountain ranges, (2) the lesser water-vapor content, and (3) the longer year on

Mars. The greater uniformity of the Martian surface should lead to more pronounced regularities in weather; thus one may be justified in combining observations from different years and in accepting a very short climatological record with more confidence than would be possible for Earth. The severe Martian aridity, as manifested by the scarcity of clouds, should also contribute to this greater uniformity of weather from year to year and day to day, since insolation and outgoing radiation are not subject to the erratic and complex interference experienced on Earth.

The fact that the Martian year is almost twice ours means that the temperature contrast between the summer and winter hemispheres ought to be larger on Mars. Here on Earth one can deal with the circulation qualitatively by assuming it to be driven by the temperature difference between equator and poles. While this would not be completely invalid on Mars it would seem that the temperature contrast between the hemispheres should also be an important factor. That is,

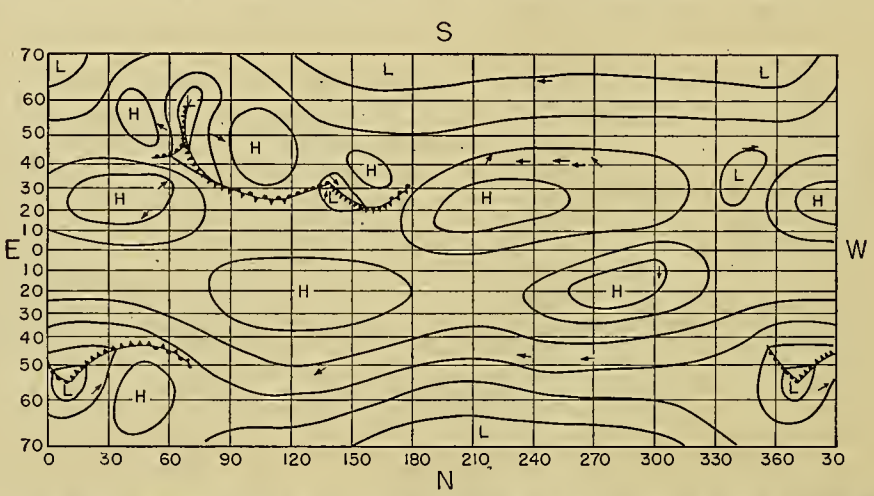


Fig. 2.—A schematic streamline map for Mars in Northern Hemisphere winter. This is the normal telescopic view with south at the top. To obtain the usual meteorological view, merely invert the page. The arrows represent observed cloud-drift directions. Taken from [2].

the long Martian seasons would seem to promote an exchange of air between hemispheres as a major feature of the circulation. One can check this conclusion by computing the magnitude of the interhemispheric exchange from the known rate of transfer of water vapor from the dissipating icecap to the forming one [2]. This calculation indicates the necessity for a wind of the order of 10–15 m sec^{-1} blowing from the warm to the cold hemisphere all around the equator, at low levels. This constitutes an appreciable interhemispheric flow.

The Characteristics of the Atmospheres of the Major Planets

The four major planets are, in order of distance from the sun, Jupiter, Saturn, Uranus, and Neptune. Jupiter's linear dimensions are 11 times larger than those of the Earth; Neptune's, 3.9 times larger. The other two planets are intermediate. All four planets are massive enough to retain hydrogen and, because of the cosmic preponderance of hydrogen, it is very likely the most important constituent of the major planets. Similarly, helium almost certainly is abundant [3, p. 316]. Unfortunately, neither hydrogen nor helium show strong spectral lines in the accessible regions of the spectrum at the temperatures of the visible surfaces of these planets (below 155K) and therefore estimates of the amounts of these molecules in the atmospheres of the outer planets must be extremely crude. Hydrogen and helium differ greatly from most other gases in some of their physical constants such as specific heat and, of course, molecular weight, so that, for example, the adiabatic lapse rate and the relation between pressure, temperature, and density are not well known.

Methane and ammonia are the only constituents of the major planets which have been identified spectroscopically. In the atmospheres of all four planets, methane is more abundant than ammonia; but the relative abundance of the two gases is not constant. The absorption of light by methane becomes more pronounced in the spectra of the more distant planets, whereas the ammonia absorption almost disappears. The surfaces of all these planets are sufficiently cold so that ammonia can exist in its frozen state. The vapor pressure of the gas in equilibrium with this frozen ammonia becomes smaller for the more distant planets, so that hardly any gaseous ammonia can exist in the atmosphere of Neptune. Ammonia is practically "frozen out" on this planet.

The spectrograph not only indicates the presence of ammonia, but also permits an estimate of the total amount of ammonia in the line of sight between us and the visible "surface." If we then assume that the atmosphere is thoroughly mixed, and if we know the approximate mean molecular weight of the other gases, we can calculate the partial pressure of ammonia at the "surface," the level of the visible markings. Since the vapor is presumably saturated or nearly saturated with respect to frozen ammonia, measurements of the strength of absorption by ammonia can be used to determine the planet's temperature. Such considerations lead to a temperature of 150K for Jupiter, about 45 degrees warmer than the temperature expected if

Jupiter were heated by the sun only and his atmosphere were transparent to radiation of all wave lengths. Therefore two possibilities suggest themselves: either Jupiter is heated from the inside as well as from the outside, or Jupiter has an appreciable greenhouse effect which is relatively much more pronounced than the Earth's.

The other giant planets also are considerably warmer than would be expected from their distances from the sun. This can be seen from the tremendous absorption by methane. If Uranus and Neptune were black bodies heated by solar radiation only with no greenhouse effect, their temperatures would be 70K and 50K respectively, low enough to "freeze out" methane. Since methane is not frozen out the actual surface temperatures must be considerably higher [8, p. 89].

In summary, the outer atmosphere of all the major planets consists mostly of hydrogen and helium, with methane being the next most common gas, at least on Jupiter and Saturn. In the atmospheres of Uranus and Neptune, inert gases, such as neon, which have lower boiling points, may be more abundant than methane.

The relative brightness of Jupiter's limb as compared to the central portion, the rate of disappearance of a satellite behind the planet's disc, and other optical phenomena indicate that comparatively little atmosphere can exist above the "visible" surface of Jupiter, at least near the equator. The pressures at the surfaces of the major planets can be estimated from the amount of methane observed spectroscopically and an added amount of hydrogen [7].

The largest portion of the atmospheres of Jupiter and Saturn must lie below the visible surface. The variable speed of the "surface" features on these planets shows that the surface is not part of any solid core. The exact thickness of the gaseous portion of the atmospheres is difficult to estimate because both the lapse rate and the exact composition are unknown. It is possible to show from the oblateness, the moment of inertia, and the mean density of these planets that the density of their outer 10 per cent must be considerably less than 0.50 g cm^{-3} [10]. Even gaseous hydrogen will reach densities greater than that at depths of the order of 1000 km [8]. For this reason, the gaseous portion of the atmospheres is estimated to extend down to a depth of several hundred kilometers. Apparently, then, the visible features in the atmospheres of Jupiter and Saturn are high-level phenomena; in contrast to conditions on Mars, little is known from observations concerning the solid core of these planets.

The appearance of Jupiter and Saturn in the telescope is characterized by a system of belts (dark or reddish color) and zones (light color), parallel to the latitude circles. Belts have also occasionally been seen on Uranus. Neptune is too far from us for any such detail to be recognized. Jupiter and Saturn differ from each other in that Jupiter's belts are more pronounced near the equator, whereas Saturn's are distributed more uniformly. Jupiter shows a "zone" centered at the equator (equatorial zone), generally bordered by the two equatorial belts. These are followed by the "tropical" zones and further by a series of "temperate" belts and zones. North of 45°N and south of 45°S, one finds generally greyish regions, which, due to the foreshortening, have been called the "polar" regions. The aspect of these zones and belts, in color and width, may change considerably from year to year. At irregular intervals a belt disappears altogether, apparently covered by a white haze. Other zones and belts change their latitudinal boundaries and get narrower or wider as time progresses.

Superimposed on these general markings is a wealth of detailed structure. White and dark spots may form almost anywhere, lasting normally from a week to several months; rifts appear in the belts, or wavelike patterns form along the boundaries between belts and zones. These patterns may have the peaked appearance of unstable waves yet persist for months. Only the *Great Red Spot*, an oval 22,000 miles long zonally and 7000 miles wide meridionally, seems to show a high degree of permanence, having been observed in the south tropical zone of Jupiter with substantially the same shape for at least seventy years, possibly for several centuries. But even the aspect of the Red Spot changes from year to year; for a few years it may be very red, then turn grey or white and be almost invisible. Its shape, also, undergoes changes; one or the other extremity may develop a peak, or the whole spot may become a little smaller or larger. There is no indication, however, that the Red Spot shows any systematic increase in its dimensions, which would be observed if it were composed of volcanic dust, a hypothesis accepted by many. Another marking, the *South Tropical Disturbance*, which has persisted since 1901, spread almost all around the planet and disappeared entirely in 1949. This marking was apparently destroyed by diffusion.

Spots on Saturn are relatively rare occurrences. Several times, large white spots have been observed for a number of weeks. No details of this type have been seen directly on the discs of Uranus or Neptune; however, Uranus occasionally shows small changes in brightness of the same period as the period of rotation. These changes are of rather an ephemeral nature and soon disappear, making it likely that they are due to fairly short-lived spots.

The inclination of Jupiter's orbit to its equator is only three degrees, so that seasonal effects, if any, should be small. Jupiter's "year" is 12 Earth-years, and color changes in the belts with that period have been reported [11]. However, this result seems to be based on a good deal of uncertain interpretation.

Saturn's seasonal effects should be considerable because of the 27-degree inclination between its orbit and equator; however, as seen from Saturn, an observer on Earth is always in nearly the same direction as the sun; hence only the summer hemisphere of Saturn can be studied in detail. The same is true to an even larger extent for Uranus, the equator of which is almost at right angles to its orbit.

Circulation of the Atmospheres of the Major Planets

All four major planets rotate about their axes rapidly, with periods between 9 and 15 hours. The super-

ficial axial symmetry and the oblateness are related to this fact. The periods can be measured by two independent methods:

1. Extremely accurate average periods have been obtained from definite markings,[1] these averages extending over weeks or months. The probable error of some such determinations is only a fraction of a second. This corresponds to errors of speed of rotation of a few tenths of a meter per second. Related to this method is the use of the periodic light variation of the more distant planets.

2. When the spectroscope slit is set parallel to a latitude circle of a planet, the planet's rotation causes the spectral lines to be inclined to the vertical. The tangent of the angle of inclination is proportional to the speed of rotation. This method gives the instantaneous speed of the layers of the atmosphere which reflect the light. The probable error of this type of measurement is of the order of 100 m sec^{-1}, a quantity which is small from an astronomical point of view, but large in comparison with normal meteorological experience.

For Jupiter, the spectral speed of rotation agrees with the speeds of the spots. In view of the large error of spectral velocities, this does not exclude the possibility that the spots may move with speeds differing from those of the general current by 50 m sec^{-1} or more.

On Saturn, the period of rotation determined by spots near the equator is about 10 hr 14 min. A very careful spectral investigation resulted in a period of 10 hr 2 min [5]. These values differ from each other by three times more than the probable error of the spectroscopic value. If real, the discrepancy probably means that the general currents on Saturn move faster than the markings. This interpretation is tenable if, for example, the markings were of the nature of storms (which usually do not move with the wind speeds).

Even though the nature of the "spots" is not known—they appear to be large cloud systems—the spot periods are usually interpreted to approximate the period of the "air" in the neighborhood of the spots. The heavy crosses in Fig. 3 show the normal distribution of the periods of spots and markings on Jupiter as function of Jovicentric latitude. The same figure, with the sign reversed, would also show the relative velocity distribution. In the bright equatorial zone the periods are short, corresponding to fast westerly winds. A transition from short to long periods takes place in the dark equatorial belts. Spots of nearly the same relatively long period occur in the tropical zones, temperate belts, temperate zones, and even the so-called polar regions near latitude 45°.

The short periods in the equatorial zone might lead to the interpretation that the equatorial spots are situated at a higher level than the markings in the other zones and belts. However, spots on Jupiter occasionally move around each other but never cover each other up as might be expected if they were at different levels.

1. Jupiter data are collected in various Memoirs of the British Astronomical Association, Jupiter Section. For Saturn, see [10] for example.

It is interesting to make an attempt to estimate the period of the core of Jupiter from the given distribution of the spot periods with latitude. One might assume, for example, that the net torque of the currents about the planet should vanish. In that case, the period of the planet would be near 9 hr 52 min, with westerlies near the equator and easterlies between latitudes 20° and 60° in both hemispheres. Such speculation, however, is misleading. After all, we are looking at a high level in Jupiter's atmosphere. If we observed only high levels of Earth's atmosphere, we would find westerlies at almost all latitudes. If, as is likely, the meridional temperature gradient on Jupiter is in the same direction as that on Earth, the high-level winds on Jupiter should also nearly all be westerlies. In that case the period of the core may be longer than that of any of the spots, that is, about 10 hours.

A wealth of information is available concerning details in the spot periods of Jupiter. Occasionally, spots near the boundaries of the currents are found, with periods differing from those of the main currents by as much as 5 min (100 m sec^{-1}). These spots are indicated by light circles in Fig. 3. Again, one might expect them to be located at levels different from those used in the determination of the periods of the principal currents, yet these unusual spots often move around the normal spots. An analysis of these unusual spots was made under the assumption that they were situated at approximately the same level as the normal markings. This analysis indicated systematically that the bright zones are regions of anticyclonic shear, and the dark belts are regions of cyclonic shear (see the line in Fig. 3). This idea is substantiated by the behavior of the

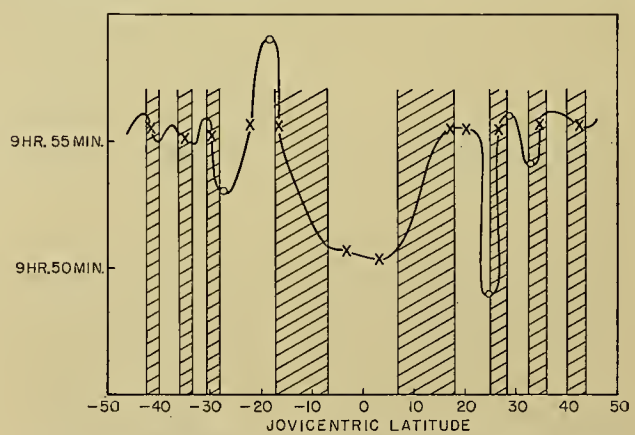


Fig. 3.—Distribution of spot periods on Jupiter with latitude. Heavy crosses indicate average periods for the latitude band; light circles indicate occasional periods. The hypothetical current distribution is represented by a line. Hatched areas are dark belts; clear areas are light zones.

"reversing spots." These spots are some of the few markings on Jupiter which show meridional movement in addition to zonal motion; they actually describe an anticyclonic orbit relative to the bright tropical zone in the Southern Hemisphere. First they move more slowly than the main current, at its northern edge, then they traverse the main current, and finally they move faster than the main current at its southern edge.

If the cyclonic shear in the dark belts and the anticyclonic shear in the light zones are permanent features, it is possible to speculate further. In that case, the bright zones would be high-pressure regions; furthermore, due to lateral friction, they would be regions of horizontal divergence. Since we are looking at a high level in the atmosphere, regions of divergence would be regions of upward vertical motion. It is then possible to explain the light color of these zones by the ammonia crystals formed in rising currents.

Jupiter is observable for about eight out of every thirteen months. In the remaining five months it is too near the sun. For each of these eight-month periods, average speeds of the different currents are available. Usually, the periods in these different observing "seasons" are based on different markings, and represent averages over a number of spots; variations from year to year may not be due to real variations of the currents but to more or less random variation between spots, since the difference of the periods of individual spots in given currents is usually much larger than the change of the average speed from "season" to "season." Occasionally variations in speed of as much as 10 m sec^{-1} occur from one season to the next, especially in the south equatorial current. These jumps affect all visible spots and are certainly not statistical accidents. Also, gradual changes which are real occur over periods of the order of 40 years.

Whatever the physical nature of the Red Spot, the variations of its speed of rotation from year to year are certainly real. The period of rotation has varied from 9 hr 55 min 33 sec to 9 hr 55 min 42 sec, with changes from year to year never exceeding 3 sec (1 m sec^{-1}). Two different, but not independent, methods showed correlations of 0.61 and 0.62, respectively, between the speed of the Red Spot and the speed of the surface westerlies on Earth at moderate north latitudes, based on 37 pairs of observations [6]. Little is known concerning the significance of correlation coefficients between time series. Perhaps a correlation of this type might be accounted for if variations in the radiation from the sun would affect the intensities of circulation both on Jupiter and on Earth. Several objections may be raised against such a hypothesis. The distance of Jupiter from the sun varies by 10 per cent during its orbital period of 12 years. Yet there is no correlation between the period of the Red Spot and the distance of Jupiter from the sun. This means that changes of total radiation of 20 per cent have no effect on the general circulation of Jupiter. The possibility remains that the ultraviolet radiation of the sun may vary by 100 per cent or more, and this large change may produce variations of the speed of air currents both on Jupiter and on Earth.

Another objection against the reality of the correlation is the fact that there is a negligible correlation between Red-Spot period and easterly index on Earth. It is hard to see why the Red Spot, which is located in the south tropical regions of Jupiter, should exhibit variations in motion similar to those in moderate terrestrial latitudes in the Northern Hemisphere.

Little information is available concerning the general circulation of the other major planets. The few spots observed on Saturn indicate the shortest periods near the equator with a *gradual* increase toward higher latitudes. The distribution of rotational velocities determined spectroscopically agrees with this picture [5].

Suggestions for Future Work

For Venus the outstanding meteorological problems are (1) determination of the nature of the variable light and dark bands observed in the ultraviolet, and (2) determination of the cause of the variability of the emission from the planet's CO_2 at 10 μ. These two phenomena may well be related, and a concerted attack should be planned, involving simultaneous observations with the spectrograph, the ultraviolet camera, and the radiometer. Once such basic data have been gathered, further interpretation may prove possible.

For Mars one would most like to see a verification and extension of the work on the temperature field and general circulation. This could best be done at the good oppositions coming in 1954, 1956, and 1958. It would require a coordinated program of radiometric observations of the temperature distribution combined with careful visual and photographic determination of the drift of clouds. This is an observing program of considerable magnitude, but is worth while in view of the basic nature of the result to be derived. We also need further observations of such fundamental quantities as the atmospheric pressure, the amount of water vapor, the height and composition of the blue haze layer, and the value of the nocturnal temperatures.

Additional information concerning Jupiter can be obtained from a study of the latitudinal variation of ammonia and methane absorption. If ammonia is saturated near the cloud surfaces, such a study should indicate the latitudinal variation of temperature and of cloud level. Studies of limb darkening in the zones and belts have so far been carried out without consideration of the extremely important selective absorption of methane and ammonia. Careful investigations of this type are necessary to determine the relative optical depths of light and dark areas. Finally, the correlation between the motion of the Red Spot and the zonal index on the Earth should be verified on observations since 1939, since standard tests of significance of correlation coefficients do not apply here.

For Saturn, similar investigations might be suggested, especially as far as the evaluation of the limb darkening is concerned.

REFERENCES

1. Hess, S. L., "A Meteorological Approach to the Question of Water Vapor on Mars and the Mass of the Martian Atmosphere." *Publ. astr. Soc. Pacif.*, 60: 289–302 (1948).
2. —— "Some Aspects of the Meteorology of Mars." *J. Meteor.*, 7: 1–13 (1950).
3. Kuiper, G. P., ed., *The Atmospheres of the Earth and Planets*. Chicago, University of Chicago Press, 1949.
4. Lampland, C. O., "On the Observable Radiation from the Carbon Dioxide in the Atmosphere of Venus." *Astrophys. J.*, 93: 401–402 (1941).

5. Moore, J. H., "Spectroscopic Observations on the Rotation of Saturn." *Publ. astr. Soc. Pacif.*, 51: 274 (1939).
6. Panofsky, H. A., and Hess, S. L., "Zonal Index and the Motion of the Great Red Spot on Jupiter." *Bull. Amer. meteor. Soc.*, 29: 426–428 (1948).
7. Peek, B. M., "The Physical State of Jupiter's Atmosphere." *Mon. Not. R. astr. Soc.*, 97: 574–582 (1937).
8. Russell, H. N., *The Solar System and Its Origin*. New York, Macmillan, 1935.
9. —— Dugan, R. S., and Stewart, J. Q., *Astronomy; A Revision of Young's Manual of Astronomy, The Solar System*, Vol. 1, 2nd rev. ed. Boston, Ginn, 1945.
10. Wildt, R., "Über den inneren Aufbau der grossen Planeten." *Veröff. UnivSternw. Göttingen*, Bd. 3, Nr. 40, S. 225 (1934).
11. Williams, A. S., "On the Periodic Variation in the Colours of the Two Equatorial Belts of Jupiter." *Mon. Not. R. astr. Soc.*, 97: 105–108 (1936–1937).

DYNAMICS OF THE ATMOSPHERE

The Perturbation Equations in Meteorology *by B. Haurwitz* 401

The Solution of Nonlinear Meteorological Problems by the Method of Characteristics *by John C. Freeman* 421

Hydrodynamic Instability *by Jacques M. Van Mieghem* 434

Stability Properties of Large-Scale Atmospheric Disturbances *by R. Fjørtoft* 454

The Quantitative Theory of Cyclone Development *by E. T. Eady* 464

Dynamic Forecasting by Numerical Process *by J. G. Charney* 470

Energy Equations *by James E. Miller* 483

Atmospheric Turbulence and Diffusion *by O. G. Sutton* 492

Atmospheric Tides and Oscillations *by Sydney Chapman* 510

Application of the Thermodynamics of Open Systems to Meteorology *by Jacques M. Van Mieghem* 531

THE PERTURBATION EQUATIONS IN METEOROLOGY

By B. HAURWITZ

New York University

INTRODUCTION

Before entering into a discussion of the systems of hydrodynamic equations suitable for the investigation of atmospheric dynamics, it is appropriate to make some general remarks on the typical difficulties of investigations in theoretical meteorology and on the general principles on which the formulation of the perturbation equations is based. Such a discussion naturally includes an enumeration of the types of problems where the application of perturbation methods is particularly suitable.

The Problem of Dynamic Meteorology. Meteorology concerns itself with the exploration and the study of the gaseous envelope of the earth. Dynamic meteorology in particular aims at a quantitative explanation of the observed variations and motions of the atmosphere on the basis of physical "laws." If a complete quantitative explanation has been accomplished, a quantitative weather forecast will be a by-product. Conversely, quantitative forecasts of the future state of the atmosphere from a given set of initial conditions may be regarded as the ultimate goal of dynamic meteorology because such forecasts require a thorough understanding of the laws governing the atmosphere.

The peculiar difficulties in meteorology, as in the other earth sciences, in comparison to physics, are that very many different factors act simultaneously on each phenomenon and that it is impossible to perform controlled experiments. In order to study the atmosphere the meteorologist has to consider a fluid shell surrounding a sphere. This fluid is neither homogeneous nor incompressible, a circumstance which complicates the purely hydrodynamical problem. Moreover, one of the constituents which make up the fluid may change from the gaseous to the liquid or solid state at temperatures and pressures regularly occurring in the atmosphere. This versatility of water vapor would be bad enough by itself, but the complications increase further because the fluid shell receives radiant energy in amounts which vary with time and location on the sphere and with the state of the water in the atmosphere. Also, the properties of the lower boundary of the fluid shell are by no means uniform, the greatest differences being those between water and land which react differently to the incoming solar radiation and which have a different frictional effect on the motions of the gaseous layer. Further, the properties of the earth's surface are by no means independent of the atmosphere, but depend rather strongly on its state; the ocean surface, for instance, may be made rougher by wind-generated waves, and the radiative properties of the land surface may be modified drastically by a snow cover. Nor is the land surface itself of uniform elevation, but rather it presents formidable obstacles to the air motion. It would be easy to continue this account of the complications which the atmosphere presents to the meteorologists, and to dwell on such matters as the fact that one of the gases (oxygen) appears in one layer in the triatomic state, in another in the monatomic state, with importantly different absorptive properties, or to discuss the electromagnetic effects in the high atmosphere. However, it may suffice here to state that finally this complex fluid system, bounded by an inhomogeneous surface and subjected to unequal influx of energy, is surrounding a rotating globe.

At that it must be admitted that the problems of astrophysics are presumably more complicated. Such phenomena as sunspots or chromospheric eruptions are subject to additional influences, for instance, to a strong electromagnetic field, or to radiation pressure, influences which the meteorologist can mercifully neglect, at least in the lower atmosphere. However, the solar physicist has the advantage of a really detached viewpoint. He is not concerned with the detailed behavior of the individual sunspot or the individual chromospheric eruption and their detailed structures. The meteorologist, on the other hand, attempts to explain the detailed structure of a cyclone or the motion of an individual hurricane. It is likely that meteorology would appear to be in a much more satisfactory state if we could apply our knowledge to observations of the earth's atmosphere as viewed from Mars.

As stated before, for the solution of its rather formidable problem, dynamic meteorology uses the laws of physics which in the case of the lower atmosphere means mainly the theorems of thermodynamics and hydrodynamics. Because of the complexity of the problems it is always necessary to make restrictive simplifying assumptions, in other words to consider models which include only some of the features of the real atmosphere. Even with these assumptions the necessary mathematical analysis offers as a rule considerable difficulties. One of the most frequently recurring difficulties arises out of the fact that the hydrodynamic equations which describe the motion of any fluid are nonlinear. Very little is known about the solution of nonlinear differential equations. Progress in this direction may presumably be expected eventually from work with high-speed computers because the study of individual nonlinear problems will point the direction in which a theory of nonlinear differential equations should be developed. In the absence of such a general theory it has long been the practice in hydrodynamics to linearize the basically nonlinear equations of fluid motion. In some types of problems of classical hydro-

dynamics the linearization is achieved directly as a result of the far-reaching simplifying assumptions. For instance, the two-dimensional irrotational motion of a homogeneous and incompressible fluid can be dealt with by the methods of the theory of complex functions. In other cases it is necessary to adopt different procedures. In tidal theory, for instance, it is customary to assume that the motion is sufficiently small so that terms which consist of products of the velocities and their derivatives can be neglected in comparison with terms which are of the first order with regard to the derivatives of the velocity components. In consequence of this assumption the convective terms in the equations of motion can be neglected, and the system of hydrodynamic equations becomes linear, at least as long as an incompressible fluid is considered. The characteristic feature of this approach is that the tidal motion is considered as a small perturbation of an original undisturbed state. This method has been used frequently in hydrodynamics, but V. Bjerknes was the first to systematize this procedure and to apply it to atmospheric problems when the earth's rotation and the compressibility and the inhomogeneity of the fluid system have to be taken into account. The system of hydrodynamic equations obtained by him is referred to as the atmospheric perturbation equations. The present article deals with their derivation and shows how they can be applied to various geophysical problems. It is not the purpose of this article to deal with the mathematical methods of solving the resulting systems of differential equations; that is a problem in mathematical physics and is treated in the appropriate textbooks. Neither is it the aim of this article to discuss some special meteorological problems. It is solely the method of approach which is being stressed here; for the solution of specific problems the reader is referred to the literature. Only some very simple examples are given as illustrations in later sections.

Basic Assumptions of the Atmospheric Perturbation Theory. As stated in the preceding section the perturbation equations are derived under the assumption that the state of the fluid motion can be regarded as made up of two parts, namely an undisturbed motion on which is superimposed a perturbation, the sum of the two representing the total motion. Such a division of an actual state of motion is often directly suggested by observations. The capillary ripples on the free surface of a pond can be regarded as a perturbation produced by a slight gust on a fluid otherwise at rest. On a much larger scale the nascent cyclones have been regarded in the wave theory of cyclones as disturbances in an undisturbed flow which consists of two currents of different densities, flowing side by side. It was, in fact, the wave theory of cyclones which led V. Bjerknes [2, 3] to the formulation of the perturbation equations as the tools by means of which the stability of such a system could be investigated.

Since the total motion thus consists of an undisturbed motion on which a perturbation is superimposed, it follows that not only the total motion, but also the undisturbed motion alone must satisfy the system of hydrodynamic equations, that is, the three equations of motion, the equation of continuity, physical equations, and such initial and boundary conditions as are appropriate for the specific problem under consideration. That the undisturbed motion by itself must satisfy the hydrodynamic equations is evident from the fact that it can exist without a superimposed perturbation. It is possible to select for any specific problem a sufficiently simple fluid state as the undisturbed motion so that little or no difficulty is encountered in satisfying the hydrodynamic equations. For instance, in the two cases mentioned above, the undisturbed state would appropriately be one of rest or geostrophic wind motion, respectively.

For the subsequent discussion, the variables expressing the actual disturbed plus undisturbed state will be denoted by letters with an asterisk, the undisturbed state will be denoted by capital letters, and the effect of the perturbation by the corresponding lower-case letters. Thus we have the following relations:

$$\text{for the position vector} \quad \mathbf{r}^* = \mathbf{R} + \mathbf{r}, \tag{1}$$

$$\text{for the velocity} \quad \mathbf{v}^* = \mathbf{V} + \mathbf{v}, \tag{2}$$

$$\text{for the pressure} \quad p^* = P + p, \tag{3}$$

$$\text{for the density} \quad q^* = Q + q. \tag{4}$$

It is now assumed that the deviation from the undisturbed state produced by the perturbation is so small that those terms in the equations which contain products of the perturbation quantities and their derivatives can be neglected as being small of higher order compared to those terms which are of the first order in the perturbation quantities. It is immediately apparent that with this assumption the hydrodynamic equations become linear with regard to the variables expressing the effects of the perturbation.

This basic assumption about the smallness of the perturbation does not, of course, represent an equally satisfactory approximation to the actual state of affairs in every case. It is particularly appropriate in cases where the stability of a certain state is to be investigated. This state will then be regarded as the undisturbed state and the stability investigation becomes the problem of determining whether or not a superimposed small perturbation will grow with time. In such a case the initial perturbation of the undisturbed motion may be regarded as infinitely small since a deviation from this motion must start gradually. If the perturbation increases with time the motion is unstable, otherwise it is stable. The limitations of this statement will be discussed later. The wave theory of cyclones and many other types of atmospheric and oceanic motions pose such stability problems. In other instances where the perturbation equations are applied it will depend on the particular circumstances how well the resulting solution represents reality, and it is impossible to make general statements about the success or failure of the method. However, it can at least be said that in many cases the linearized perturbation equations give a satisfactory approximation to the observed fluid motions.

The results of hydrodynamic wave theory give numerous examples to support this statement.

It may be remarked here in passing that there are certain problems in fluid motion which reduce to linear equations without the introduction of the assumptions of the perturbation theory. Consider, for instance, the frictionless horizontal motion of an incompressible and homogeneous fluid. Then the continuity equation can be satisfied by assuming a stream function $\psi(x, y, t)$, and the pressure can be eliminated from the two equations of motion by cross differentiation resulting in the so-called vorticity equation which is in this case a nonlinear differential equation for ψ. However, if

$$\nabla^2\psi = \text{const}\,\psi,$$

the two quadratic terms cancel each other (more generally if $\nabla^2\psi = f(\psi)$, but this relation is no longer generally linear). The foregoing equation for ψ comprises a great number of equations arising in mathematical physics. In the meteorological literature, examples for this type of motion will be found, among others, in papers by Craig [7], Neamtan [15], and Rombakis [17].

SYSTEMS OF HYDRODYNAMIC EQUATIONS

Only a very short review of the equations of motion and of continuity in the Eulerian and the Lagrangian systems will be given here since these equations can be found in all treatises on hydrodynamics. But since in classical hydrodynamics the fluid is as a rule regarded as incompressible and homogeneous, some more detailed remarks are in order concerning the modifications which are necessary when such inhomogeneous and compressible fluids as the atmosphere are considered.

Barotropic and Baroclinic Fluids. In fluids which are inhomogeneous and compressible the surfaces of equal density and of equal pressure do not as a rule coincide; the stratification of the fluid is baroclinic. In special cases, however, the density distribution may be given completely by the pressure distribution; then the stratification of the fluid is called barotropic. A barotropic stratification exists, for instance, in an ideal gas whose temperature is the same everywhere or in a fluid where pressure and density are functions only of the elevation. An incompressible and homogeneous fluid is evidently barotropic and will always remain barotropic. A fluid whose stratification always remains barotropic is called autobarotropic. In general, a barotropic state will be destroyed by the motion of the fluid; the fluid will become baroclinic.

The significance of the distinction between barotropy and baroclinity for the variation of the circulation of a fluid need not be discussed here since it is explained in the texts on dynamic meteorology.

It is often convenient to introduce a "coefficient of barotropy" when the stratification is barotropic,

$$q^* = q^*(p^*). \tag{5}$$

The coefficient of barotropy Γ is then defined as follows,

$$\Gamma = \frac{dq^*}{dp^*}. \tag{6}$$

In an atmosphere obeying the ideal-gas law with a constant vertical temperature gradient ϵ, for instance,

$$\Gamma = \left(1 - \frac{R}{g}\epsilon\right)\frac{1}{RT}, \tag{7}$$

where R is the gas constant for this atmosphere, g the acceleration of gravity, assumed to be constant, and T the temperature.

It is possible to generalize this terminology of baroclinity and barotropy to apply to the distribution of other variables besides pressure and specific volume [4, p. 3].

The Physical Equation. Since in meteorological and oceanographic investigations the fluid medium can rarely be considered as homogeneous and incompressible, the density has to be regarded as one of the unknown variables. In this case the four equations used mainly in classical hydrodynamics, namely, the three equations of motion and the equation of continuity, are not sufficient to describe the fluid motion because the density has been added as a fifth variable to the three velocity components and the pressure. It is therefore necessary to have an additional equation describing the behavior of the fluid. Such an equation, in the case of the atmosphere, can be derived from thermodynamic considerations. In the first place for the variables of state, pressure p^*, density q^*, and temperature T^*, an equation of state holds of the form

$$F(p^*, q^*, T^*, A, B, C, \cdots) = 0, \tag{8}$$

where $A, B, C, \cdots$ indicate additional parameters characterizing the state of the fluid. In the case of the atmosphere the parameters A, B, etc., refer to the differences in the composition of the atmosphere. For many problems it is sufficient to use the equation of state for ideal gases:

$$p^* - Rq^*T^* = 0, \tag{9}$$

where R is the gas constant for air, which depends on the composition of the atmosphere. In the lower atmosphere, R is mainly affected by the variable amount of water vapor; at higher levels it depends on the separation of the various atmospheric constituents and dissociation of some of the molecules. Consequently R may vary from parcel to parcel in the atmosphere. While equation (8) or (9) adds an additional equation to the system of hydrodynamic equations it is not sufficient to complete the system since it involves also an additional variable, namely, the temperature T^*.

In order to obtain a more complete system of equations the first law of thermodynamics may be added:

$$dh = de + p^*d(1/q^*), \tag{10}$$

where dh denotes the amount of heat added to the unit mass during an infinitely small change of state, de the change of internal energy per unit mass, and where $p^*d(1/q^*)$ represents the amount of work done because of the expansion of the unit mass. The amount of heat

is not a variable of state. If this quantity is also to be considered as unknown, further additional equations have to be introduced for the transfer of heat by conduction and radiation. In many instances in meteorological problems it may be assumed that the changes of state are adiabatic, so that $dh = 0$. A slight generalization of this procedure is to assume polytropic changes where

$$dh = c\, dT^*. \tag{11}$$

Here c is a constant of the character of a specific heat. In most meteorological problems it can further be assumed that

$$de = c_v\, dT^*,$$

where c_v is the specific heat at constant density, an assumption which holds for an ideal gas and is consequently as a rule a satisfactory assumption for the atmosphere. Then with the aid of (9),

$$c\, dT^* = c_p dT^* - \frac{RT^*}{p^*}\, dp^*,$$

since the specific heat at constant pressure for ideal gases

$$c_p = c_v + R.$$

By integration of the preceding differential relation under the assumption that the specific heats are constant, which applies to ideal gases,

$$\frac{T^*}{T_0^*} = \left(\frac{p^*}{p_0^*}\right)^{\bar{K}}, \tag{12}$$

where $\bar{K} = (c_p - c_v)/(c_p - c)$. The relation between pressure and specific volume consequently becomes

$$p^* q^{*-\bar{\lambda}} = p_0^* q_0^{*-\bar{\lambda}}, \tag{13}$$

where $\bar{\lambda} = (c_p - c)/(c_v - c)$, the modulus of the polytropic curve. We shall confine ourselves in the following discussions to the special case of adiabatic changes when

$$\bar{\lambda} = \lambda = \frac{c_p}{c_v}$$

and

$$p^* q^{*-\lambda} = p_0^* q_0^{*-\lambda}. \tag{13a}$$

Equation (13) or (13a) gives the variation of the specific volume of a fluid particle as a function of the pressure variation. The original pressure and density, p_0^* and q_0^* will, of course, in general vary from one particle to another. Compared to the more general equation (8), equations (13) and (13a) have one variable less since the temperature is eliminated by the additional assumptions about the physical process by which the particle changes its density. Thus, the number of dependent variables is reduced by one. Such an equation for the change of state of a particle which, in addition to the pressure, contains only one more variable of state has been called by V. Bjerknes an equation of piezotropy, and a fluid whose every particle is following an equation of this type is called piezotropic. More generally, a piezotropic fluid satisfies an equation of the form

$$f(p^*, q^*, A, B, \cdots) = 0, \tag{14}$$

where the parameters $A, B, \cdots$ will in general not be the same as in (8). Equation (14) may also be written in the form

$$q^* = g(p^*, A, B, \cdots). \tag{14a}$$

It is often convenient to introduce the coefficient of piezotropy,

$$\gamma = \frac{dq^*}{dp^*}. \tag{15}$$

In the case of adiabatic changes, for instance,

$$\gamma = \frac{1}{\lambda RT^*}. \tag{15a}$$

It should be clearly understood that this coefficient of piezotropy describes the physical changes which a particle undergoes, while the coefficient of barotropy depends on the distribution of the fields of pressure and specific volume.

When an equation of piezotropy exists for the fluid, the system of hydrodynamic equations is complete because it consists now of five equations with five dependent variables.

The properties of barotropy and piezotropy are entirely unconnected. The former refers to the stratification of a fluid, the latter to its changes of state. A fluid whose stratification is barotropic at a given instant will in general not remain so if in motion. However, if the appropriate state of piezotropy prevails, barotropy may be maintained. Such a fluid is called autobarotropic, as mentioned before. The condition for autobarotropy is that the coefficients of barotropy and of piezotropy are identical, provided, of course, that the fluid was barotropic at a fixed instant. Examples of autobarotropic fluids are a homogeneous and incompressible fluid, or an atmosphere with adiabatic changes of temperature and an adiabatic lapse rate of temperature. In such a fluid a displaced particle has always the same density as its environment so that indifferent equilibrium is maintained. Consequently, autobarotropic fluids are only a minor generalization of the homogeneous and incompressible fluids considered in classical hydrodynamics.

The Hydrodynamic Equations. Any fluid, whether compressible or not, is subject to the laws of dynamics. For continua, the equations expressing these laws are more complicated than for the mass points of particle dynamics, because the direct effects of neighboring fluid particles on each other have to be taken into account. This interrelationship leads to the appearance of the terms for the viscous stresses and for the pressure-gradient forces in the hydrodynamic equations of motion. In many instances in classical hydrodynamics and in geophysical applications, however, the effects

of viscosity and friction are disregarded, and an "ideal," nonviscous fluid is considered. In particular, the atmospheric perturbation equations have so far almost exclusively been developed for and applied to such ideal fluids. Very often, far-reaching agreement has been obtained between theoretical models dealing with ideal fluids and observed atmospheric motions, so that there is considerable justification for making use of the substantial simplifications which arise when the effects of viscosity on fluid motion are neglected.

In the case of the atmosphere it is customary to describe the motions relative to the rotating earth since the observing meteorologist participates in the earth's rotation. According to the laws of Newtonian mechanics we may then write that

$$\frac{d^2\mathbf{r}^*}{dt^2} + 2\boldsymbol{\Omega} \times \frac{d\mathbf{r}^*}{dt} = -\left(\frac{1}{q^*}\right)\nabla p^* - \nabla\Phi. \qquad (16)$$

Here $\mathbf{r}^*$ is the radius vector of a fluid particle, $\boldsymbol{\Omega}$ the vector of the earth's rotation, q^* the density, p^* the fluid pressure. It is assumed that the external forces acting on the fluid have a potential Φ, because in the case of the atmosphere this external force is, as a rule, only gravity—composed of the gravitational attraction of the earth and the centrifugal acceleration due to the earth's rotation—which has a potential. In some cases, of course, this gravity potential may have to be replaced or augmented by additional forces, for instance, in the study of tides.

In order to apply equation (16) to problems of fluid dynamics one can either focus one's attention on each individual particle and describe its future position, pressure, density, etc., or describe the distribution of velocity, pressure, density, etc., throughout the space occupied by the fluid, as functions of the time. The first method leads to the so-called Lagrangian equations, the latter to the Eulerian equations, although both systems of equations were originally derived by Euler. The Eulerian system is used much more widely in hydrodynamics and will be described first. We introduce in (16) the velocity

$$\mathbf{v}^* = \frac{d\mathbf{r}^*}{dt}.$$

The operator d/dt represents the time differentiation of an individual particle. In order to have only quantities referring to a fixed point in the coordinate system in the equation, this operator may be written

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v}^*\cdot\nabla. \qquad (17)$$

Thus the Eulerian equation of motion becomes

$$\frac{\partial \mathbf{v}^*}{\partial t} + \mathbf{v}^*\cdot\nabla\mathbf{v}^* + 2\boldsymbol{\Omega} \times \mathbf{v}^* = -\left(\frac{1}{q^*}\right)\nabla p^* - \nabla\Phi. \qquad (18)$$

In addition the fluid must satisfy the equation of continuity,

$$\frac{\partial q^*}{\partial t} + \operatorname{div}\,(q^*\,\mathbf{v}^*) = 0, \qquad (19)$$

according to which the density change anywhere in the fluid must equal the mass convergence there. In order to complete the system the physical equation has to be added as explained in the preceding section. We shall assume that the fluid is piezotropic so that an equation of the form (14) holds. Most investigations so far have been dealing with piezotropic fluids, although in some cases it has been assumed that a certain amount of heat energy, given as a function of space and time, is added to the system (for instance, [14]) or that heat is conducted from the earth's surface into the atmosphere (for instance, [12]). In the latter case the effects of eddy viscosity are taken into account in order to have a consistent fluid model.

Equation (14) does not fit into the Eulerian system because it refers to an individual particle, but upon individual differentiation with respect to time and with (15) and (17) it follows that

$$\frac{\partial q^*}{\partial t} + \mathbf{v}^*\cdot\nabla q^* = \gamma\left(\frac{\partial p^*}{\partial t} + \mathbf{v}^*\cdot\nabla p^*\right). \qquad (20)$$

It should be noted that in general γ will vary in space. Equations (18), (19), and (20) represent a complete system of equations which together with the boundary conditions (to be discussed in the next section) and initial conditions, determine the motion of the fluid.

In order to apply the Lagrangian method to the study of the fluid motion it is necessary to identify each particle individually. This can be done by the use of three suitable parameters, a, b, and c, so that the state of the fluid and each of its individual particles is described by expressing $\mathbf{r}^*$, p^*, and q^* as functions of t, a, b, and c. The choice of these parameters is free provided that they permit one to characterize and identify each particle uniquely. It has, for instance, been suggested that under certain conditions potential temperature, specific humidity, and potential vorticity would be suitable parameters [21]. As a rule the coordinates at a given time t_0 are chosen as the three identification parameters, and this practice will be used here.

In order to write the Lagrangian equations in vector form we shall make use of the operation $\nabla\mathbf{B}\cdot\mathbf{A}$. The meaning of this expression is that after the scalar product $\mathbf{B}\cdot\mathbf{A}$ has been formed the vector operator ∇ operates only on the first vector. Thus

$$(\nabla\mathbf{B}\cdot\mathbf{A})_x = A_x\left(\frac{\partial B_x}{\partial x}\right) + A_y\left(\frac{\partial B_y}{\partial x}\right) + A_z\left(\frac{\partial B_z}{\partial x}\right).$$

It may be noted that

$$\nabla\mathbf{r}\cdot\mathbf{A} = \mathbf{A}. \qquad (21)$$

In the case of the Lagrangian equations the operator ∇ has the components $\partial/\partial a$, $\partial/\partial b$, and $\partial/\partial c$. When the operator with the components $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$ is used in the remainder of this section it will be denoted by $\nabla_{\mathbf{r}}$. It should be noted that instead of the total derivatives with respect to time in (16) we may now write the partial derivatives since a, b, and c are independent of time. Furthermore, according to (21),

$$\nabla\mathbf{r}^*\cdot\nabla_{\mathbf{r}} p = \nabla p,$$

and an analogous relation holds for Φ. Hence, the operation $\nabla \mathbf{r}^* \cdot$ applied to (16) results in the Lagrangian equation of motion

$$\nabla \mathbf{r}^* \cdot \left[\frac{\partial^2 \mathbf{r}^*}{\partial t^2} + 2\boldsymbol{\Omega} \times \left(\frac{\partial \mathbf{r}^*}{\partial t} \right) \right] = -\left(\frac{1}{q^*} \right) \nabla p^* - \nabla \Phi. \quad (22)$$

The equation of continuity may be written in the following form:

$$q^* \left(\frac{\partial \mathbf{r}^*}{\partial a} \right) \times \left(\frac{\partial \mathbf{r}^*}{\partial b} \right) \cdot \left(\frac{\partial \mathbf{r}^*}{\partial c} \right) = q_0^*, \quad (23)$$

where q_0^* represents the density of the fluid particle at the time t_0 when a, b, and c were the coordinates of the particle. It is easily verified that in a rectangular Cartesian coordinate system (23) is identical with

$$q^* \frac{D(x^*, y^*, z^*)}{D(a, b, c)} = q_0^*, \quad (24)$$

where the Jacobian is used to express the volume change of the fluid particle.

To complete the Lagrangian system of equations the physical equation has to be added. While for the Eulerian method this equation had to be differentiated in order to obtain quantities referring to fixed points no such differentiation is required in the present case since both the physical equation and the equations of motion refer to individual fluid particles. Of course, the parameters $A, B, \cdots$ in (14) may differ from particle to particle so that they may have to be considered as functions of a, b, and c, but they will not change with time.

The Boundary Conditions. At rigid or internal boundaries and at free surfaces the fluid has to obey certain conditions which exert considerable influence on the possible motions. We shall again first consider the form of these boundary conditions in the Eulerian system. The boundary of the fluid is given by an equation of the form,

$$f^*(\mathrm{r}, t) = 0. \quad (25)$$

The kinematic boundary condition expresses the fact that a fluid particle which forms part of the boundary must remain on this surface and that a particle which is not at a given time part of the boundary can never become part of it. This condition is expressed by the following relation,

$$\frac{\partial f^*}{\partial t} + \mathbf{v}^* \cdot \nabla f^* = 0. \quad (26)$$

As is known from hydrodynamics this relation can also be interpreted as the condition that a fluid particle at the boundary has a velocity component normal to the boundary which must equal the velocity of the surface itself.

If the boundary is a rigid surface, as for instance the surface of the earth, f^* is independent of the time (neglecting such unpleasant geophysical phenomena as earthquakes) and (26) becomes

$$\mathbf{v}^* \cdot \nabla f^* = 0, \quad (26a)$$

which states that the fluid velocity at the boundary must be parallel to the boundary.

In the case of an internal boundary between two different fluid layers a condition similar to (26) must hold on the other side of the boundary, in the second fluid layer, where the velocity is $\mathbf{v}^{*\prime}$,

$$\frac{\partial f^*}{\partial t} + \mathbf{v}^{*\prime} \cdot \nabla f^* = 0. \quad (27)$$

In addition to the kinematic boundary conditions the dynamic boundary condition has to be satisfied, that is, the pressure across the boundary must be continuous,

$$p^*(\mathrm{r}, t) - p^{*\prime}(\mathrm{r}, t) = 0, \quad (28)$$

where the prime refers again to the second fluid. At a free surface

$$p^*(\mathrm{r}, t) = \text{const.} \quad (28a)$$

This constant may be zero, for instance if the fluid is bounded by empty space. In some cases the constant may have to be replaced by a function of space and time, for instance when the effect of variations in atmospheric pressure on the ocean surface is to be studied. Equation (28) or, in the case of a free surface, (28a) represents the boundary given by (25). In many problems the function appearing in (25) will, in fact, have to be determined by (28) or (28a). Therefore, the pressure difference at the boundary, or in the case of a free surface the pressure there, may be introduced into equations (26) and (27). Thus, the following boundary conditions, which represent a mixture of kinematic and dynamic conditions [19], are obtained:

$$\begin{aligned} \frac{\partial}{\partial t}(p^* - p^{*\prime}) + \mathbf{v}^* \cdot \nabla(p^* - p^{*\prime}) &= 0, \\ \frac{\partial}{\partial t}(p^* - p^{*\prime}) + \mathbf{v}^{*\prime} \cdot \nabla(p^* - p^{*\prime}) &= 0. \end{aligned} \quad (29)$$

It is only at a rigid boundary that the dynamic condition need not be satisfied, and the geometric form of the boundary must here be known from other sources. Then equations (26) and (27) have to be used.

In the Lagrangian system the form of the boundary conditions corresponding to those just given for the Eulerian system is considerably more complicated because attention is now focused no longer on points in space but on individual particles on both sides at the boundary which may have been neighboring at the time $t = 0$, but will in general be at a finite distance from each other at a later time. A complete discussion of these conditions is given by V. Bjerknes and collaborators [4, pp. 63–64, 103–104]. We shall here discuss only some simpler forms, also given by V. Bjerknes [3], which are adequate for many problems of fluid motion.

At the time $t = 0$ fluid particles which are at the boundary must satisfy one of the following two equations,

$$f_0^*(a, b, c) = 0, \qquad f_0^*(a', b', c') = 0, \quad (30)$$

where the primes refer to one fluid layer, the coordinates without primes to the other layer and where, as indicated, the same function applies on both sides of the boundary. When the components of the radius vector $\mathbf{r}^*$, or $\mathbf{r}^{*\prime}$ in the other fluid layer have been determined, a, b, c or a', b', c' have to be expressed by these components and t. The results are two equations,

$$f^*(\mathbf{r}^*, t) = 0, \qquad f^{*\prime}(\mathbf{r}^{*\prime}, t) = 0, \tag{31}$$

where the functions will now be different for the different fluid layers. The kinematic boundary conditions require that both equations (31) represent the same surface at all times t, a requirement which V. Bjerknes writes symbolically

$$[f^*(\mathbf{r}^*, t) = 0] \equiv [f^{*\prime}(\mathbf{r}^{*\prime}, t) = 0]. \tag{32}$$

The functions $\mathbf{r}^*$ and $\mathbf{r}^{*\prime}$ of a, b, c, t and a', b', c', t, respectively, will contain certain arbitrary integration constants which have to be chosen so that (32) holds. At a free surface or a rigid lower boundary only one of the two equations in (30) and (31) exists, and this equation has then to be adjusted in such a manner that any prescribed condition concerning the motion at this boundary is satisfied. For instance at a rigid boundary, $\mathbf{r}^*$ must be such that displacements are parallel to it.

In addition the dynamic boundary condition has to be satisfied, namely, that the pressure at an internal boundary is continuous. The pressure in both layers is given by $p^*(a, b, c, t)$ and $p^{*\prime}(a', b', c', t)$. Then, after $\mathbf{r}^*$ and $\mathbf{r}^{*\prime}$ have been determined as functions of a, b, c, t and a', b', c', t, respectively, p^* can be expressed as a function of t and $\mathbf{r}^*$, or rather of the components of the radius vector, and after obtaining the analogous expression for $p^{*\prime}$ one has to satisfy the condition that

$$p^*(\mathbf{r}^*, t) = p^{*\prime}(\mathbf{r}^{*\prime}, t) \quad \text{for} \quad \mathbf{r}^* = \mathbf{r}^{*\prime}. \tag{33}$$

At a free surface the right-hand side of this equation has to be put equal to zero or to the given external pressure.

THE PERTURBATION EQUATIONS

The two systems of hydrodynamic equations derived in the preceding sections can be linearized by the assumptions set forth on page 402. This linearization will now be performed, first on the Eulerian system, then on the Lagrangian system. From the perturbation equations in vector form it is more or less easy to obtain the equations in the coordinate form which is most suitable for any particular problem. As an example which is important for the study of large-scale motions on the earth, the perturbation equations for spherical polar coordinates in the Eulerian system are derived rather explicitly in a later section.

The Eulerian System of Perturbation Equations. The undisturbed motion will be denoted by capital letters as stated earlier. In order to write the complete system of equations for the undisturbed motion it is therefore merely necessary to replace in the relevant preceding equations the quantities with asterisks by the corresponding capital letters. Thus the equation of motion is obtained from (18), the equation of continuity from (19), the physical equation from (20) (with the previously stated restriction that only piezotropic fluids are considered), the boundary equation from (25), the kinematic boundary condition from (26) (from (27) if an internal boundary exists), the dynamic boundary condition from (28), and the mixed boundary condition from (29) which may be used instead of the kinematic boundary conditions. Thus the equations of the undisturbed state are

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V}\cdot\nabla\mathbf{V} + 2\boldsymbol{\Omega} \times \mathbf{V} = -\left(\frac{1}{Q}\right)\nabla P - \nabla\Phi, \tag{34}$$

$$\frac{\partial Q}{\partial t} + \operatorname{div}(Q\mathbf{V}) = 0, \tag{35}$$

$$\frac{\partial Q}{\partial t} + \mathbf{V}\cdot\nabla Q - \gamma\left(\frac{\partial P}{\partial t} + \mathbf{V}\cdot\nabla P\right) = 0, \tag{36}$$

$$F(\mathbf{r}, t) = 0, \tag{37}$$

$$\frac{\partial F}{\partial t} + \mathbf{V}\cdot\nabla F = 0, \qquad \frac{\partial F}{\partial t} + \mathbf{V}'\cdot\nabla F = 0, \tag{38}$$

$$P(\mathbf{r}, t) - P'(\mathbf{r}, t) = 0, \tag{39}$$

$$\begin{aligned} &\frac{\partial}{\partial t}(P - P') + \mathbf{V}\cdot\nabla(P - P') = 0, \\ &\frac{\partial}{\partial t}(P - P') + \mathbf{V}'\cdot\nabla(P - P') = 0. \end{aligned} \tag{40}$$

It is understood that in the boundary conditions the coordinates have to satisfy (37). If any particular motion whose perturbations are to be investigated is selected, one has to make sure first that this undisturbed motion satisfies equations (34)–(40).

In order to derive the perturbation equations one has now to substitute in the equations previously mentioned, namely, (18)–(20) and (25)–(29), for the quantities with asterisks the sums (2)–(4). Furthermore f^* has to be replaced by $F + f$, where F refers to the boundary in the undisturbed position while f denotes the variation of the boundary due to disturbance. The resulting equations can be simplified, as explained on page 402, because

1. The undisturbed quantities must satisfy (34)–(40), and
2. Terms of second or higher order in the perturbation quantities can be neglected.

Consider, for instance, equation (18) which may now be written

$$\frac{\partial \mathbf{V}}{\partial t} + \frac{\partial \mathbf{v}}{\partial t} + \underline{\mathbf{V}\cdot\nabla\mathbf{V}} + \mathbf{V}\cdot\nabla\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{V} + \underline{\underline{\mathbf{v}\cdot\nabla\mathbf{v}}} + \underline{2\boldsymbol{\Omega} \times \mathbf{V}}$$

$$+ 2\boldsymbol{\Omega} \times \mathbf{v} = -\left(\frac{1}{Q + q}\right)\nabla(P + p) - \underline{\nabla\Phi}.$$

The first term on the right-hand side may be expanded as follows,

$$\left(\frac{1}{Q+q}\right)\nabla(P+p)$$
$$= \frac{1}{Q}\left(1 - \frac{q}{Q} + \frac{q^2}{Q^2} - \cdots\right)\nabla(P+p)$$
$$= \underline{\frac{1}{Q}\nabla P} - \frac{q}{Q^2}\nabla P + \underline{\underline{\frac{q^2}{Q^3}\nabla P}} - \cdots + \frac{1}{Q}\nabla p - \underline{\underline{\frac{q}{Q^2}\nabla p}} + \cdots.$$

Here the singly underlined terms vanish because they form together the undisturbed equation. The doubly underlined terms can be omitted because they are small of higher order. In a similar manner the other perturbation equations can be obtained so that the complete system becomes

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{V}\cdot\nabla\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{V} + 2\boldsymbol{\Omega}\times\mathbf{v} = -\frac{\nabla p}{Q} + \frac{q\nabla P}{Q^2}, \tag{41}$$

$$\frac{\partial q}{\partial t} + \operatorname{div}(q\mathbf{V}) + \operatorname{div}(Q\mathbf{v}) = 0, \tag{42}$$

$$\frac{\partial q}{\partial t} + \mathbf{V}\cdot\nabla q + \mathbf{v}\cdot\nabla Q - \gamma\left(\frac{\partial p}{\partial t} + \mathbf{V}\cdot\nabla p + \mathbf{v}\cdot\nabla P\right) = 0, \tag{43}$$

$$F(\mathrm{r}, t) + f(\mathrm{r}, t) = 0, \tag{44}$$

$$\frac{\partial f}{\partial t} + \mathbf{V}\cdot\nabla f + \mathbf{v}\cdot\nabla F = 0, \qquad \frac{\partial f}{\partial t} + \mathbf{V}'\cdot\nabla f + \mathbf{v}'\cdot\nabla F = 0, \tag{45}$$

$$P(\mathrm{r}, t) - P'(\mathbf{r}, t) + p(\mathrm{r}, t) - p'(\mathrm{r}, t) = 0, \tag{46}$$

$$\frac{\partial}{\partial t}(p - p') + \mathbf{V}\cdot\nabla(p - p') + \mathbf{v}\cdot\nabla(P - P') = 0, \qquad \frac{\partial}{\partial t}(p - p') + \mathbf{V}'\cdot\nabla(p - p') + \mathbf{v}'\cdot\nabla(P - P') = 0. \tag{47}$$

In equations (44) and (46) the undisturbed terms alone are no longer zero in spite of (37) and (39), because these equations hold only if the values at the undisturbed boundary are substituted for r, while in (44) and (46) the values at the disturbed boundary have to be inserted into F, P, and P'. On the other hand, it is permissible to replace in the perturbation quantities contained in the equations (44)–(47) the values at the disturbed position of the boundary by the values at the undisturbed position because this substitution produces only an error of higher order. Consider, for instance, the perturbation pressure p. For the moment let r be the undisturbed position of the boundary, $\mathrm{r} + \Delta\mathrm{r}$ the disturbed position where $\Delta\mathrm{r}$ must be small of the same order as the other perturbation quantities. Then

$$p(\mathrm{r} + \Delta\mathrm{r}, t) = p(\mathrm{r}, t) + (\Delta\mathrm{r})\cdot\nabla p(\mathrm{r}, t),$$

where the second term on the right-hand side is smaller than the preceding by one order of magnitude.

V. Bjerknes has pointed out that the perturbation equations (41)–(47) can be obtained from the equations of the undisturbed motion (34)–(40) by forming the variation of these equations and substituting the perturbation quantities wherever variations of the undisturbed quantities appear.

The Lagrangian System of Perturbation Equations. In the formulation of the perturbation equations in the Lagrangian form it will be assumed that at the time $t = t_0$ the position and the parameters of state for the undisturbed particle are the same as in the undisturbed case. This assumption implies that the perturbation is induced suddenly at the time t_0. It has the advantage that the three parameters a, b, and c, which identify each particle, are not only the components of the radius vector $\mathbf{R}_0$ for the undisturbed motion at the time t_0, but also for the disturbed motion since r_0, the deviation of the disturbed from the undisturbed position at t_0, vanishes. The equations for the undisturbed motion in the Lagrangian system are now obtained by substituting for the appropriate quantities with asterisks in (22), (23), (14), (30), (32), and (33) the corresponding quantities characterizing the undisturbed motion which are denoted by capital letters. Thus,

$$\nabla\mathrm{R}\cdot\left[\frac{\partial^2\mathrm{R}}{\partial t^2} + 2\boldsymbol{\Omega}\times\left(\frac{\partial\mathrm{R}}{\partial t}\right)\right] = -\frac{\nabla P}{Q} - \nabla\Phi, \tag{48}$$

$$Q\left(\frac{\partial\mathrm{R}}{\partial a}\right)\times\left(\frac{\partial\mathrm{R}}{\partial b}\right)\cdot\left(\frac{\partial\mathrm{R}}{\partial c}\right) = Q_0, \tag{49}$$

$$Q = Q(a, b, c, P), \tag{50}$$

$$F_0(a, b, c) = 0, \qquad F_0(a', b', c') = 0, \tag{51}$$

$$[F(\mathrm{R}, t) = 0] \equiv [F'(\mathrm{R}', t) = 0], \tag{52}$$

$$P(\mathrm{R}, t) = P'(\mathrm{R}, t) \quad \text{for} \quad \mathrm{R} = \mathrm{R}'. \tag{53}$$

The physical equation (50) has here been written somewhat differently from (14) by making use of the fact that the parameters A, B, and C, $\cdots$ may vary from particle to particle and are therefore functions of a, b, and c.

The perturbation equations are now obtained by making in these same six previously mentioned equations the substitutions indicated by (1), (3), and (4) and taking into account also that the surface of discontinuity will in general change its position because of the perturbation motion. The resulting equations can be simplified by the same two considerations as stated on page 407 for the Eulerian equations. It is to be noted that the force potential Φ depends on the position in space. Hence, while in the undisturbed motion it is given by $\Phi(\mathrm{R})$, it will be given by $\Phi(\mathrm{R} + \mathrm{r})$ for the same particle after the beginning of the perturba-

tion. According to Taylor's theorem,

$$\Phi(\mathbf{R} + \mathbf{r}) = \Phi(\mathbf{R}) + \mathbf{r}\cdot\nabla_R\Phi + \text{higher-order terms.}$$

Here, the operator ∇_R denotes space differentiation with respect to the undisturbed position vector R. Similarly, for the vector of the earth's rotation,

$$\Omega(\mathbf{R} + \mathbf{r}) = \Omega(\mathbf{R}) + (\mathbf{r}\cdot\nabla_R)\Omega.$$

The boundary condition (51) given above for the undisturbed state remains the same for the disturbed state since the condition refers to the time t_0 when the particles were in their undisturbed position. The functions $F(\mathbf{R}, t)$ and $F'(\mathbf{R}', t)$ in (52) are modified in the perturbed motion because the particles at the boundary whose position vectors in the undisturbed motion were R and R′ are now R + r and R′ + r′, respectively. According to Taylor's theorem,

$$F(\mathbf{R}+\mathbf{r}, t) = F(\mathbf{R}, t) + \mathbf{r}\cdot\nabla_R F + \text{higher-order terms,}$$

and an analogous relation holds for $F'(\mathbf{R}' + \mathbf{r}', t)$. Thus, because of (52) the kinematic boundary condition for the disturbed motion becomes as written below in (58). In deriving the dynamic boundary condition (59), use is made of the facts that the same particles are considered in the undisturbed and the disturbed case and that in the perturbation quantities the disturbed position may be replaced by the undisturbed position. Thus the system of perturbation equations in Lagrangian form becomes

$$\begin{aligned}\nabla\mathbf{R}\cdot\left\{\frac{\partial^2\mathbf{r}}{\partial t^2} + 2\Omega\times\left(\frac{\partial\mathbf{r}}{\partial t}\right) + \left[(\mathbf{r}\cdot\nabla_R)\Omega\right]\times\left(\frac{\partial\mathbf{R}}{\partial t}\right)\right\} \\ + \nabla\mathbf{r}\cdot\left[\frac{\partial^2\mathbf{R}}{\partial t^2} + 2\Omega\times\left(\frac{\partial\mathbf{R}}{\partial t}\right)\right] = -\frac{\nabla p}{Q} \\ + \frac{q(\nabla P)}{Q^2} - \nabla(\mathbf{r}\cdot\nabla_R\Phi),\end{aligned} \tag{54}$$

$$\begin{aligned}&\frac{q}{Q}\left(\frac{\partial\mathbf{R}}{\partial a}\right)\times\left(\frac{\partial\mathbf{R}}{\partial b}\right)\cdot\left(\frac{\partial\mathbf{R}}{\partial c}\right) \\ &\quad + \left(\frac{\partial\mathbf{r}}{\partial a}\right)\times\left(\frac{\partial\mathbf{R}}{\partial b}\right)\cdot\left(\frac{\partial\mathbf{R}}{\partial c}\right) \\ &\quad + \left(\frac{\partial\mathbf{R}}{\partial a}\right)\times\left(\frac{\partial\mathbf{r}}{\partial b}\right)\cdot\left(\frac{\partial\mathbf{R}}{\partial c}\right) \\ &\quad + \left(\frac{\partial\mathbf{R}}{\partial a}\right)\times\left(\frac{\partial\mathbf{R}}{\partial b}\right)\cdot\left(\frac{\partial\mathbf{r}}{\partial c}\right) = 0,\end{aligned} \tag{55}$$

$$q = \gamma p, \tag{56}$$

$$F_0(a, b, c) = 0, \qquad F_0'(a', b', c') = 0, \tag{57}$$

$$\mathbf{r}\cdot\nabla_R F = \mathbf{r}'\cdot\nabla_R F', \qquad \mathbf{R} = \mathbf{R}', \tag{58}$$

$$p(\mathbf{R}, t) = p'(\mathbf{R}', t), \qquad \mathbf{R} = \mathbf{R}'. \tag{59}$$

As in the case of the Eulerian system the perturbation equations (54) to (59) can be obtained from the equations (48) to (53) by forming the variation of these equations.

A comparison of the three basic equations, namely, of motion, of continuity, and physical changes in the Eulerian system and in the Lagrangian system, shows that the equation of motion is of the first order in the Eulerian form (41) while it is of the second order in the Lagrangian form (54). The equation of continuity in the Eulerian form (42) as well as in the Lagrangian form (55) is of the first order. But the physical equation is a first-order differential equation in the Eulerian system (43) while in the Lagrangian system it has the simple form (56) so that the perturbation density q can, with the aid of these equations, be eliminated directly from the equations of motion and continuity. In the Eulerian system, on the other hand, additional differentiations will in general be necessary in order to eliminate the perturbation density, so that for a compressible fluid the Eulerian equations are not simpler than the Lagrangian equations. The Eulerian system is simpler than the Lagrangian system only for an incompressible and homogeneous fluid, the case mostly considered in classical hydrodynamics.

From a physical viewpoint it may be remarked that meteorological observations are made at a given locality and not following individual air particles so that the Eulerian rather than the Lagrangian method is used in this case. However, when trajectories or the motion of air masses are studied the Lagrangian method offers a more direct approach than the Eulerian method. It is, of course, always possible to make the transition from the results in the one system to those in the other. When the Lagrangian system has been used the positions of the particles and their pressures and densities are given. By a differentiation the particle velocities can be obtained, and the distribution in space of the velocity, pressure, and density can then be found because the position of the particles in space is known. When the Eulerian system is used the particle trajectories can be found from the known velocity distribution.

In some instances a certain simplification may be achieved in the work leading to the mathematical solution depending on whether the problem is formulated in the Eulerian or Lagrangian system. But no general statements about the relative difficulty of the one or the other approach can be made, and it has to be seen by direct comparison whether the Lagrangian or the Eulerian method is more advantageous.

Perturbation Equations for Special Coordinate Systems. From the general equations given in the two preceding sections the special forms suitable for a specific problem can be derived. It will depend on the particular fluid model to be considered, in particular on the boundary conditions, what simplifications can be made in the system of perturbation equations, and which type of coordinates is most suitable. As long as the earth can be considered as flat, a rectangular Cartesian coordinate system is the best choice, unless the motion has a circular symmetry. In the latter case, cylindrical polar coordinates are most appropriate. For large-scale disturbances on the earth its curvature has to be taken into account, and in this case spherical polar coordinates represent the most suitable frame of reference.

Some simple typical examples of problems to be treated by the perturbation method will be considered later. Here we shall discuss the transition from the vector form of the equations to the coordinate form. This transition presents little or no difficulty in the case of straight-line coordinates, such as the Cartesian rectangular coordinate system, but it is somewhat more awkward if curvilinear coordinates, such as spherical or cylindrical polar coordinates, are to be used. The reason for this difficulty is that in the case of curvilinear coordinates the change in direction along the coordinate curves has to be taken into consideration.

The problem of writing the perturbation equations in arbitrary curvilinear coordinates has been solved by V. Bjerknes [3, 4], with the use of Lagrange's equations of the second kind, but the resulting equations will not be reproduced here since the vector equations given previously permit also the derivation of the equations in any particular coordinate system. In order to show by an example how the equations for the undisturbed and the disturbed motion can be derived for a curvilinear coordinate system from the vector equations given above, spherical polar coordinates will be chosen, and the equations will be written in the Eulerian system.

Depending on the amount of vector and tensor analysis assumed to be known this derivation naturally becomes shorter or longer. Here only the more elementary theorems of vector analysis will be used. Consider a right-hand rectangular Cartesian coordinate system whose z-axis, for convenience, is parallel to the earth's axis. Let ϑ, λ, and r be the colatitude, longitude, and distance, respectively, from the earth's center.[1] Further, let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ be the unit vectors in the x, y, and z directions, respectively. Then the radius vector

$$\begin{aligned}\mathbf{r} &= \mathbf{i}x + \mathbf{j}y + \mathbf{k}z \\ &= \mathbf{i}r\sin\vartheta\cos\lambda + \mathbf{j}r\sin\vartheta\sin\lambda + \mathbf{k}r\cos\vartheta.\end{aligned} \tag{60}$$

Further, since the radius vector $\mathbf{r}$ is also a function of λ, ϑ, and r

$$d\mathbf{r} = \left(\frac{\partial\mathbf{r}}{\partial\vartheta}\right)d\vartheta + \left(\frac{\partial\mathbf{r}}{\partial\lambda}\right)d\lambda + \left(\frac{\partial\mathbf{r}}{\partial r}\right)dr. \tag{61}$$

Here $\partial\mathbf{r}/\partial\vartheta$ is evidently a vector tangential to the curve of intersection between the coordinate surfaces

$$\lambda = \text{const} \quad \text{and} \quad r = \text{const},$$

or as it may be called briefly a "direction vector" in the direction ϑ. Similarly $\partial\mathbf{r}/\partial\lambda$ and $\partial\mathbf{r}/\partial r$ are direction vectors in the directions λ and r, respectively. These three direction vectors play in the polar coordinate system a role similar to that of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in the Cartesian system except that they are not unit vectors, a shortcoming which is immaterial here. By appropriate differentiations of (60) the following relations are obtained:

$$\begin{aligned}\frac{\partial\mathbf{r}}{\partial\vartheta} &= \mathbf{i}r\cos\vartheta\cos\lambda + \mathbf{j}r\cos\vartheta\sin\lambda - \mathbf{k}r\sin\vartheta, \\ \frac{\partial\mathbf{r}}{\partial\lambda} &= -\mathbf{i}r\sin\vartheta\sin\lambda + \mathbf{j}r\sin\vartheta\cos\lambda, \\ \frac{\partial\mathbf{r}}{\partial r} &= \mathbf{i}\sin\vartheta\cos\lambda + \mathbf{j}\sin\vartheta\sin\lambda + \mathbf{k}\cos\vartheta.\end{aligned} \tag{62}$$

From these three equations it is found that

$$\begin{aligned}\mathbf{i}r\sin\vartheta &= \left(\frac{\partial\mathbf{r}}{\partial\vartheta}\cos\vartheta + \frac{\partial\mathbf{r}}{\partial r}r\sin\vartheta\right)\sin\vartheta\cos\lambda \\ &\qquad - \frac{\partial\mathbf{r}}{\partial\lambda}\sin\lambda, \\ \mathbf{j}r\sin\vartheta &= \left(\frac{\partial\mathbf{r}}{\partial\vartheta}\cos\vartheta + \frac{\partial\mathbf{r}}{\partial r}r\sin\vartheta\right)\sin\vartheta\sin\lambda \\ &\qquad + \frac{\partial\mathbf{r}}{\partial\lambda}\cos\lambda, \\ \mathbf{k}r &= -\frac{\partial\mathbf{r}}{\partial\vartheta}\sin\vartheta + \frac{\partial\mathbf{r}}{\partial r}r\cos\vartheta.\end{aligned} \tag{63}$$

In the following we shall for the moment consider only the undisturbed motion. The extension to the disturbed motion will be recognized easily. From (61) it follows by individual differentiation with respect to time that

$$\mathbf{V} = \frac{d\mathbf{R}}{dt} = \left(\frac{\partial\mathbf{R}}{\partial\vartheta}\right)\dot{\Theta} + \left(\frac{\partial\mathbf{R}}{\partial\lambda}\right)\dot{\Lambda} + \left(\frac{\partial\mathbf{R}}{\partial r}\right)\dot{R}. \tag{64}$$

Here the dots indicate individual differentiation with respect to time. The following relations exist between the linear and angular velocity components,

$$r\dot{\Theta} = V_\vartheta, \qquad r\sin\vartheta\dot{\Lambda} = V_\lambda, \qquad \dot{R} = V_r. \tag{65}$$

For the present the notation in (64) will be retained. With the foregoing relations the Coriolis term $\boldsymbol{\Omega}\times\mathbf{V}$ can be evaluated. Since $\boldsymbol{\Omega}$ is in the direction of z, $\boldsymbol{\Omega} = \Omega\mathbf{k}$. Further, $\mathbf{V}$ may be expressed by its components in the directions of x, y, and z with the aid of (64) and (63) and the vector product of $\boldsymbol{\Omega}$ and $\mathbf{V}$ may be evaluated. In the resulting expression the unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ can be replaced again by the direction vectors $\partial\mathbf{R}/\partial\vartheta$, $\partial\mathbf{R}/\partial\lambda$, and $\partial\mathbf{R}/\partial r$ so that

$$\begin{aligned}2\boldsymbol{\Omega}\times\mathbf{V} = &-2\Omega\left\{\sin\vartheta\left[\left(\frac{\partial\mathbf{R}}{\partial\vartheta}\right)\cos\vartheta + \left(\frac{\partial\mathbf{R}}{\partial r}\right)r\sin\vartheta\right]\dot{\Lambda}\right. \\ &\left. + \left[\dot{\Theta}\cot\vartheta + \frac{\dot{R}}{r}\right]\left[\frac{\partial\mathbf{R}}{\partial\lambda}\right]\right\}.\end{aligned} \tag{66}$$

1. It is also possible to use elliptic coordinates if the flattening of the earth is to be taken into account [20]. For most meteorological considerations, however, this factor can be neglected, and the sum of gravity and centrifugal force due to the earth's rotation can be assumed in the direction to the center of the earth.

The operator

$$\nabla = \frac{1}{r^2}\left(\frac{\partial \mathrm{R}}{\partial \vartheta}\right)\left(\frac{\partial}{\partial \vartheta}\right) + \frac{1}{r^2 \sin^2 \vartheta}\left(\frac{\partial \mathrm{R}}{\partial \lambda}\right)\left(\frac{\partial}{\partial \lambda}\right) + \left(\frac{\partial \mathrm{R}}{\partial r}\right)\left(\frac{\partial}{\partial r}\right). \tag{67}$$

This relation can for instance be verified directly from the form of ∇ in Cartesian rectangular coordinates with the aid of (63). Further,

$$\mathbf{V}\cdot\nabla = \dot{\Theta}\left(\frac{\partial}{\partial \vartheta}\right) + \dot{\Lambda}\left(\frac{\partial}{\partial \lambda}\right) + \dot{R}\left(\frac{\partial}{\partial r}\right). \tag{68}$$

In order to obtain the expression $\mathbf{V}\cdot\nabla\mathbf{V}$ which appears in the equations of motion the direction vectors have also to be differentiated with respect to ϑ, λ, and r. This differentiation can be carried out with the aid of (62), and the resulting expressions can again be written as direction vectors according to (63). It will be found that

$$\begin{aligned}\mathbf{V}\cdot\nabla\mathbf{V} = {} & \left(\frac{\partial \mathrm{R}}{\partial \vartheta}\right)\left[(\mathbf{V}\cdot\nabla)\dot{\Theta} + 2\dot{\Theta}\frac{\dot{R}}{r} - \dot{\Lambda}^2 \sin\vartheta\cos\vartheta\right] \\ & + \left(\frac{\partial \mathrm{R}}{\partial \lambda}\right)\left[(\mathbf{V}\cdot\nabla)\dot{\Lambda} + 2\dot{\Theta}\dot{\Lambda}\cot\vartheta + 2\dot{\Lambda}\frac{\dot{R}}{r}\right] \\ & + \left(\frac{\partial \mathrm{R}}{\partial r}\right)\left[(\mathbf{V}\cdot\nabla)\dot{R} - \dot{\Theta}^2 r - r\dot{\Lambda}^2 \sin^2\vartheta\right].\end{aligned} \tag{69}$$

The expression for the divergence in polar coordinates is sufficiently well known so that it need only be quoted here:

$$\operatorname{div}\mathbf{V} = \frac{\partial \dot{\Theta}}{\partial \vartheta} + \frac{\partial \dot{\Lambda}}{\partial \lambda} + \frac{\partial \dot{R}}{\partial r} + (\cot\vartheta)\dot{\Theta} + \frac{2\dot{R}}{r}. \tag{70}$$

With these foregoing transformations the equations (34) to (40) can now be written in polar coordinates. The three component equations of (34) can be obtained if, after the necessary substitutions are made from (64) and (66)–(69), the factors of each of the direction vectors are equated separately. In writing down these equations the linear velocities have been substituted according to (65). Then

$$\mathbf{V}\cdot\nabla = V_\vartheta\left(\frac{\partial}{r\partial\vartheta}\right) + V_\lambda\left(\frac{\partial}{r\sin\vartheta\,\partial\lambda}\right) + V_r\left(\frac{\partial}{\partial r}\right), \tag{71}$$

and the equations of motion assume the following form:

$$\frac{\partial V_\vartheta}{\partial t} + \mathbf{V}\cdot\nabla V_\vartheta + \frac{V_r V_\vartheta}{r} - V_\lambda^2\left(\frac{\cot\vartheta}{r}\right) - 2\Omega\cos\vartheta\, V_\lambda = -\frac{1}{Q}\left(\frac{\partial P}{r\partial\vartheta}\right), \tag{72}$$

$$\frac{\partial V_\lambda}{\partial t} + \mathbf{V}\cdot\nabla V_\lambda + \frac{V_r V_\lambda}{r} + V_\vartheta V_\lambda\left(\frac{\cot\vartheta}{r}\right) + 2\Omega(V_\vartheta\cos\vartheta + V_r\sin\vartheta) = -\frac{1}{Q}\left(\frac{\partial P}{r\sin\vartheta\,\partial\lambda}\right), \tag{73}$$

$$\frac{\partial V_r}{\partial t} + \mathbf{V}\cdot\nabla V_r - \frac{V_\vartheta^2}{r} - \frac{V_\lambda^2}{r} - 2\Omega\sin\vartheta\, V_\lambda = -\frac{1}{Q}\left(\frac{\partial P}{\partial r}\right) - \frac{\partial \Phi}{\partial r}. \tag{74}$$

The equation of continuity is

$$\frac{\partial Q}{\partial t} + \mathbf{V}\cdot\nabla Q + Q\operatorname{div}\mathbf{V} = 0, \tag{75}$$

where

$$\operatorname{div}\mathbf{V} = \frac{\partial(\sin\vartheta\, V_\vartheta)}{r\sin\vartheta\,\partial\vartheta} + \frac{\partial V_\lambda}{r\sin\vartheta\,\partial\lambda} + \frac{1}{r^2}\frac{\partial(r^2 V_r)}{\partial r} \tag{76}$$

and the physical equation as well as the boundary conditions have the same form as that given for the Eulerian system of perturbation equations on page 407 except that the operator $\mathbf{V}\cdot\nabla$ is now of the form given by (71).

As explained before, the perturbation equations can be written down directly by forming the variation δ of the equations of the undisturbed motion and replacing the quantities δV_ϑ, δV_λ, δV_r, δP, and δQ by the corresponding perturbation quantities. This procedure which is more rapid than the one used in the two preceding sections leads to the following equations:

$$\frac{\partial v_\vartheta}{\partial t} + \mathbf{V}\cdot\nabla v_\vartheta + \mathbf{v}\cdot\nabla V_\vartheta + \frac{v_r V_\vartheta}{r} + \frac{V_r v_\vartheta}{r} - \frac{2V_\lambda v_\lambda\cot\vartheta}{r} - 2\Omega\cos\vartheta\, v_\lambda = -\frac{1}{Q}\left(\frac{\partial p}{r\partial\vartheta}\right) + \frac{q}{Q^2}\left(\frac{\partial P}{r\partial\vartheta}\right), \tag{77}$$

$$\frac{\partial v_\lambda}{\partial t} + \mathbf{V}\cdot\nabla v_\lambda + \mathbf{v}\cdot\nabla V_\lambda + \frac{v_r V_\lambda}{r} + \frac{V_r v_\lambda}{r} + \frac{v_\vartheta V_\lambda\cot\vartheta}{r} + \frac{V_\vartheta v_\lambda\cot\vartheta}{r} + 2\Omega(v_\vartheta\cos\vartheta + v_r\sin\vartheta) = -\frac{1}{Q}\left(\frac{\partial p}{r\sin\vartheta\,\partial\lambda}\right) + \frac{q}{Q^2}\left(\frac{\partial P}{r\sin\vartheta\,\partial\lambda}\right), \tag{78}$$

$$\frac{\partial v_r}{\partial t} + \mathbf{V}\cdot\nabla v_r + \mathbf{v}\cdot\nabla V_r - \frac{2V_\vartheta v_\vartheta}{r} - \frac{2V_\lambda v_\lambda}{r} - 2\Omega\sin\vartheta\, v_\lambda = -\frac{1}{Q}\left(\frac{\partial p}{\partial r}\right) + \frac{q}{Q^2}\left(\frac{\partial P}{\partial r}\right), \tag{79}$$

$$\frac{\partial q}{\partial t} + \mathbf{V}\cdot\nabla q + \mathbf{v}\cdot\nabla Q + Q\operatorname{div}\mathbf{v} + q\operatorname{div}\mathbf{V} = 0. \tag{80}$$

The expressions $\mathbf{v}\cdot\nabla$ and $\operatorname{div}\mathbf{v}$ are analogous to (71) and (76), respectively. The physical equation and the boundary conditions have again the same form as the analogous equations in the Eulerian system of perturbation equations.

EXAMPLES OF THE USE OF THE PERTURBATION METHOD

The present article deals with a method of handling theoretical problems rather than with a theory and interpretation of atmospheric phenomena. Therefore, the discussion of the following examples stresses this method rather than the application of the results to

specific meteorological problems. Even the method is discussed only to the extent to which it illustrates the application of the perturbation equations. The additional procedures of mathematical physics which lead to a solution of the problem are mentioned only very briefly.

Only a few fairly simple examples of the use of the perturbation method are given here which illustrate the various types of problems that can be handled in this manner.

Perturbations in Two Incompressible Fluid Currents, in the Lagrangian System. As a first example of the application of perturbation equations to fluid motions, a very simple case will be considered. This case has been discussed in the textbooks on hydrodynamics, although as a rule only with the aid of the Eulerian system. Here the Lagrangian equations will be used. Assume that we have two incompressible homogeneous fluid layers, both in constant horizontal motion. Let the density of the lower fluid be Q, its velocity U, while the density and velocity for the upper fluid are Q' and U', respectively. The earth's rotation will be neglected. The effects of capillary forces which are important for short water waves will also be disregarded. Further, let the lower boundary be horizontal and rigid at $Z = c = 0$, let the internal discontinuity be at the level h, and the free surface be at H, while the thickness of the upper layer may be denoted by h'. The last two statements imply that both the internal and the free surface are horizontal in the undisturbed case. This is physically obvious and will be corroborated by the equations for the undisturbed motion.

It is sufficient to consider the motion in a vertical XZ plane only. It is

$$X = a + Ut, \qquad Z = c. \tag{81}$$

The equations of motion and of continuity reduce, according to (47) and (49) to

$$0 = \frac{\partial P}{\partial a}, \tag{82}$$

$$0 = -\frac{1}{Q}\frac{\partial P}{\partial c} - g, \tag{83}$$

$$Q = Q_0 = \text{const.} \tag{84}$$

The subscript zero for the density can, of course, be omitted. Equations analogous to (81)–(84) hold for the upper layer. Since, according to (82) and (83), the pressure is a function of the vertical coordinate c only, it follows from the boundary condition (53) that the internal and free surfaces must both be horizontal. If (83) is integrated and if the outside pressure at the free surface vanishes,

$$P' = gQ'(H - c), \tag{85}$$

$$P = gQ'h' + gQ(h - c). \tag{86}$$

The integration constants have been chosen so that the two equations satisfy also the condition that the pressure be continuous at the internal surface. Since according to (53) at the free surface $P' = 0$, and at the internal surface $P - P' = 0$, it follows that both these surfaces must be horizontal in the undisturbed case.

The perturbation equations of motion and of continuity follow from (54) and (55), namely,

$$\frac{\partial^2 x}{\partial t^2} = -\frac{1}{Q}\frac{\partial p}{\partial a} - \frac{\partial}{\partial a}(gz), \tag{87}$$

$$\frac{\partial^2 z}{\partial t^2} = -\frac{1}{Q}\frac{\partial p}{\partial c} - \frac{\partial}{\partial c}(gz), \tag{88}$$

$$\frac{\partial x}{\partial a} + \frac{\partial z}{\partial c} = 0, \tag{89}$$

and analogous equations follow for the upper layer. From this system of three equations the three unknown variables x, z, and p can be determined as functions of the coordinates at the time $t = t_0$, namely a and c, and of the time t. Since the system is not only linear, but also has constant coefficients, exponential or trigonometric functions will represent solutions. It may be assumed that the functional dependence on a, c, and t can be expressed in the form

$$\begin{aligned} x, z, p &= A, C, D \exp\{i\alpha(X - \sigma t) + \beta Z\} \\ &= A, C, D \exp\{i\alpha[a - (\sigma - U)t] + \beta c\}, \end{aligned} \tag{90}$$

where A, C, and D are constants. The complex form for the periodicity term is more convenient than the trigonometric form and is therefore used here. For the subsequent physical interpretation either the real or the imaginary part of the solution or a linear combination of both can be used. The form of the solution given above represents a wave in the a direction of the length $2\pi/\alpha$ and moving with the speed σ. Special attention should be called to the appearance of U (and U' for the upper layer) in the exponent since the perturbation quantities remain small only with regard to the *undisturbed* position of the particle, but not with respect to its *initial* position.

The expressions (90) do not satisfy the condition that the perturbation quantities vanish at the time $t = t_0$ but it can be shown by transition to the corresponding Eulerian equations that the expressions developed here represent solutions of the problem.

Substitution of (90) into equations (87)–(89) leads to a system of linear homogeneous equations for A, C, and D. The system has nontrivial solutions only if its determinant vanishes. It follows that $\beta = \pm\alpha$. Two of the three constants A, C, and D may now be expressed by the third, and in view of the condition to be satisfied at the horizontal lower boundary, A and D may be expressed by C. Thus

$$A = i\frac{\beta}{a}C, \qquad D = -Q\frac{g\beta - \alpha^2(\sigma - U)^2}{\beta}C, \tag{91}$$

and analogous expressions hold for the upper layer. Because of the two values for β there are two different solutions, and a linear combination of the two solutions represents also a solution of the system of differential equations. When the two arbitrary constants

are denoted by C and C^* it follows that

$$z = [Ce^{\alpha c} + C^*e^{-\alpha c}] \exp \{i\alpha[a - (\sigma - U)t]\}.$$

Since, according to (58), z must vanish at the rigid lower boundary ($c = 0$), it follows that

$$C = -C^* = \frac{K}{2}, \text{ say.}$$

Consequently,

$$\begin{aligned} x &= [iK \cosh \alpha c] \exp \{i\alpha[a - (\sigma - U)t]\}, \\ z &= [K \sinh \alpha c] \exp \{i\alpha[a - (\sigma - U)t]\}, \\ p &= -QK[g \sinh \alpha z - \alpha(\sigma - U)^2 \cosh \alpha z] \\ &\quad \times \exp \{i\alpha[a - (\sigma - U)t]\}. \end{aligned} \tag{92}$$

In the upper layer the pressure is given by

$$\begin{aligned} p' = -Q'\{&g(C'e^{\alpha c} + C^{*\prime}e^{-\alpha c}) \\ &- \alpha(\sigma - U')^2(C'e^{\alpha c} - C^{*\prime}e^{-\alpha c})\} \\ &\times \exp \{i\alpha[a - (\sigma - U')t]\}. \end{aligned}$$

This expression must vanish at the free upper surface ($c = H$) according to (59), so that

$$\begin{aligned} &[g - \alpha(\sigma - U')^2]C'e^{\alpha H} \\ &= -[g + \alpha(\sigma - U')^2]C^{*\prime}e^{-\alpha H} = \frac{K'}{2}[g^2 - \alpha^2(\sigma - U')^4], \end{aligned}$$

where K' is a new constant introduced for convenience. Consequently,

$$\begin{aligned} x' &= iK'[g \cosh \alpha(c - H) \\ &\quad + \alpha(\sigma - U')^2 \sinh \alpha(c - H)] \\ &\quad \times \exp \{i\alpha[a - (\sigma - U')t]\}, \\ z' &= K'[g \sinh \alpha(c - H) \\ &\quad - \alpha(\sigma - U')^2 \cosh \alpha(c - H)] \\ &\quad \times \exp \{i\alpha[a - (\sigma - U')t]\}, \\ p' &= -Q'K'[g^2 - \alpha^2(\sigma - U')^4] \sinh \alpha(c - H) \\ &\quad \times \exp \{i\alpha[a - (\sigma - U')t]\}. \end{aligned} \tag{93}$$

At the internal boundary the two conditions (58) and (59) become

$$\left.\begin{aligned} z &= z' \\ p &= p' \end{aligned}\right\} \quad \text{for} \quad X = X' \quad \text{and} \quad Z = Z' = h.$$

Strictly speaking, these two conditions as well as the condition at the free upper surface are to be satisfied at the disturbed position of the boundary, but only an error of higher order is incurred as explained previously (p. 408) if the condition is made to hold at the undisturbed position. The two conditions at the internal boundary give two homogeneous linear equations for K and K', and again the determinant must vanish if nontrivial solutions are to exist. Thus the following equation is obtained:

$$\begin{aligned} &Q[g - \alpha(\sigma - U)^2 \coth \alpha h][g - \alpha(\sigma - U')^2 \coth \alpha h'] \\ &\quad = Q'[g^2 - \alpha^2(\sigma - U')^4]. \end{aligned} \tag{94}$$

An equation of this type is referred to as a secular or frequency equation. It permits the determination of the wave velocity, and thus the frequency, as a function of the wave length $2\pi/\alpha$ and of the physical parameters of the system. It is hardly necessary to discuss this equation in any detail because the wave motion in a fluid system such as is being considered here is too well known. In order to illustrate stability investigations one special example may be considered, however, namely that both layers are so deep compared to the wave length that

$$\coth \alpha h' = \coth \alpha h = 1.$$

In this case we may write

$$\begin{aligned} &Q[g - \alpha(\sigma - U)^2][g - \alpha(\sigma - U')^2] \\ &\quad = Q'[g + \alpha(\sigma - U')^2][g - \alpha(\sigma - U')^2]. \end{aligned} \tag{95}$$

A first solution of this equation is given by the following expression,

$$\sigma = U' + \left(\frac{g}{\alpha}\right)^{1/2}. \tag{96}$$

This value of σ gives the wave velocity of surface waves in deep water. It is physically plausible that such waves can exist in an infinitely deep layer with a free upper surface, even if the lower boundary is an internal discontinuity rather than a rigid plane, because the amplitude of the surface waves decreases to zero at the boundary. If this type of wave is disregarded, we obtain from (95) a quadratic equation for σ whose roots are

$$\begin{aligned} \sigma = &\frac{UQ + U'Q'}{Q + Q'} \\ &\pm \left(\frac{g}{\alpha}\frac{Q - Q'}{Q + Q'} - \frac{QQ'(U - U')^2}{(Q + Q')^2}\right)^{1/2}. \end{aligned} \tag{97}$$

This familiar expression shows that the wave velocity consists of two terms, the "convective" term which represents a weighted mean of the velocities in both layers and the "dynamic" term which depends on the density and the wind discontinuity. The latter term can evidently become imaginary for sufficiently small density differences and wave lengths or for sufficiently large wind discontinuities. Then we may write

$$\sigma = \sigma_1 \pm i\sigma_2,$$

and the periodicity terms of the perturbation equations become

$$\exp \{i\alpha[a - (\sigma_1 - U)t] \mp \alpha\sigma_2 t\}.$$

Thus the perturbation quantities depend exponentially on the time and, since one of the two exponentials has a positive exponent, a solution exists which increases exponentially with time, indicating that the motion is unstable. The particular type of instability arising here is referred to as shearing instability since it is due to the wind shear.

Once the solutions (90) with the various relations

between the different constants have been found it is, of course, possible to satisfy any required initial conditions by suitable linear combinations of expressions of the form (90).

Wave Motions in a Compressible Atmosphere. As an example of motion in a compressible medium, the case of an atmosphere may be considered where the temperature decreases linearly with the elevation at the rate ϵ. In the undisturbed state the atmosphere may be at rest. The lower boundary of this atmosphere may be at $c = -h^I$; there may be an internal discontinuity at $c = 0$ where the temperature, and consequently the density, changes abruptly. The upper surface is free. Its height is determined by the fact that temperature, pressure, and density all vanish here if the temperature decreases linearly with altitude. The earth's rotation will again be neglected; this is permissible for small-scale phenomena such as microbarometric oscillations or billow clouds.

It follows from these assumptions that in the undisturbed state the hydrostatic equation must be satisfied:

$$g = -\frac{1}{Q}\frac{\partial P}{\partial c},$$

and that pressure, temperature, and density are functions of c only, specifically,

$$\frac{P}{P_0} = \left(\frac{T}{T_0}\right)^{g/R\epsilon},$$

$$\frac{Q}{Q_0} = \left(\frac{T}{T_0}\right)^{(g/R\epsilon)-1}.$$

Further, we shall assume that changes of state are adiabatic. Then, according to (15a), the coefficient of piezotropy

$$\gamma = \frac{1}{\lambda RT}$$

and, according to (10), the coefficient of barotropy

$$\Gamma = \frac{1 - R\epsilon/g}{RT}.$$

Both coefficients become equal if

$$\epsilon = \bar{\epsilon} = \left(\frac{g}{R}\right)\frac{\lambda - 1}{\lambda}, \tag{98}$$

the adiabatic lapse rate.

Since the motion may again be assumed as two-dimensional the perturbation equations are given according to (54) and (55), if the perturbation density is replaced by the perturbation pressure according to (56),

$$\frac{\partial^2 x}{\partial t^2} + \frac{1}{Q}\left(\frac{\partial p}{\partial a}\right) + \frac{\partial}{\partial a}(gz) = 0,$$

$$\frac{\partial^2 z}{\partial t^2} + \frac{1}{Q}\left(\frac{\partial p}{\partial c}\right) + \frac{\partial}{\partial c}(gz) + g\gamma\left(\frac{p}{Q}\right) = 0, \tag{99}$$

$$\frac{p}{Q} = -\frac{1}{\gamma}\left(\frac{\partial x}{\partial a} + \frac{\partial z}{\partial c}\right).$$

The second of these equations may also be written in the form

$$\frac{\partial^2 z}{\partial t^2} + \frac{\partial}{\partial c}\left(\frac{p}{Q}\right) + \frac{\partial}{\partial c}(gz) - g(\Gamma - \gamma)\left(\frac{p}{Q}\right) = 0.$$

In general the system (99) has coefficients which depend on the elevation c. Only if the atmosphere is isothermal, $\epsilon = 0$, γ and Γ are constants, and in this case the coefficients are constant if p/Q rather than p is considered as an unknown variable. The case of an isothermal atmosphere is therefore particularly easy to deal with.

In the more general case of a nonisothermal atmosphere the functional dependence of the unknown variables x, z, and p on the height must be left open, and it may be assumed that

$$x,\ z,\ p = A(c),\ C(z),\ D(c)\exp[i\alpha(a - \sigma t)]. \tag{100}$$

If these expressions are substituted in (99), the system may be transformed into a second-order ordinary differential equation for one of the three functions A, C, and D of c. For instance, the differential equation for the pressure amplitude D becomes

$$\frac{d^2D}{dz^2} + \left(\frac{g}{R} - \epsilon\right)\left(\frac{1}{T}\right)\frac{dD}{dz} + \left\{-\alpha^2 + \frac{1}{RT}\left[\frac{\alpha^2\sigma^2}{\lambda} + (\bar{\epsilon} - \epsilon)\frac{Rg}{\sigma^2}\right]\right\} = 0. \tag{101}$$

When D is known, the vertical amplitude C can be obtained from the following expression,

$$QC = \frac{\sigma^2\frac{dD}{dz} + gD}{\alpha^2\sigma^4 - g^2}. \tag{102}$$

Similarly, the horizontal amplitude A can be expressed by D but this function need not be known to satisfy the boundary conditions.

The differential equation (101) can with the aid of the substitutions $D = e^{\alpha c}X(c)$ and $y = -2\alpha T/\epsilon$ be transformed into the confluent hypergeometric equation,

$$y\frac{d^2X}{dy^2} + \left[y - \frac{g}{R\epsilon} + 1\right]\frac{dX}{dy} - \left[\frac{\alpha\sigma^2}{2\epsilon R} + \frac{(\bar{\epsilon} - \epsilon)g}{2\alpha\epsilon\sigma^2}\right]X = 0. \tag{103}$$

The case of one atmospheric layer has been discussed by Lamb [13] and V. Bjerknes and collaborators [4].

In the case of an internal discontinuity surface where such phenonema as billow clouds arise it may be assumed that the wave motion does not extend far upward or downward from the interface. An approximate solution may then be obtained in the following manner. Since

$$T = T_0\left(1 - \frac{\epsilon c}{T_0}\right),$$

multiplication of (101) by $1 - \epsilon c/T_0$ permits one to

write (101) in the form

$$L(D) = \frac{\epsilon c}{T_0}\left(\frac{d^2D}{dc^2} - \alpha^2 D\right),$$

where L stands for the operator acting on D on the left-hand side of (101) when T is replaced by T_0. If we put

$$D = D_0 + \left(\frac{\epsilon}{T_0}\right) D_1 + \left(\frac{\epsilon}{T_0}\right)^2 D_2 + \cdots,$$

comparison of the coefficients of ϵ/T_0 shows that

$$L(D_0) = 0, \tag{104}$$

$$L(D_1) = c\left(\frac{d^2D_0}{dc^2} - \alpha^2 D_0\right), \text{ etc.} \tag{105}$$

The differential equation (104) for the zero approximation can readily be solved since it is homogeneous and has constant coefficients. It is, of course, the equation obtained from (101) if for the variable coefficient T in this equation its constant value T_0 at the interface is blithely substituted. The correction terms D_1, etc., can also be computed by elementary means since they involve the solution of a second-order linear and inhomogeneous equation with constant coefficients. It can be shown that the series for D converges, and for practical purposes when relatively short waves such as billow clouds are studied the zero approximation is sufficiently accurate. According to (104),

$$D_0 = e^{-\mu c/2}(K_1 e^{Nc} + K_2 e^{-Nc}), \tag{106}$$

where K_1 and K_2 are integration constants and

$$\mu = \frac{g/R - \epsilon}{T_0}, \tag{107}$$

$$N^2 = \left(\frac{\mu}{2}\right)^2 + \alpha^2 - \frac{\alpha^2\sigma^2/\lambda + (\tilde{\epsilon} - \epsilon)Rg/\sigma^2}{RT_0}. \tag{108}$$

In order to distinguish the upper and lower layer, primes will be added to the appropriate letters to indicate that these quantities refer to the upper layer. After D_0 has been determined for the lower layer, C_0 is found from (102). Since the vertical displacement must vanish at the lower rigid boundary where $c = -h$, it follows that a relation must exist between the two constants K_1 and K_2. With the introduction of a suitable new constant

$$C_0 = Ke^{-\mu c/2}\frac{\sinh N(c + h)}{Q\alpha^2(\sigma^4\alpha^2 - g^2)} \tag{109}$$

and

$$D_0 = Ke^{-\mu c/2}\left[\frac{\sigma^2 N \cosh N(c + h)}{\alpha^2(\varkappa_1 + g)(\varkappa_2 - g)} + \frac{(\sigma^2\mu/2 - g)\sinh N(c + h)}{\sigma^2(\varkappa_1 + g)(\varkappa_2 - g)}\right], \tag{110}$$

where $\varkappa_1 = \sigma^2(N - \mu/2)$ and $\varkappa_2 = \sigma^2(N + \mu/2)$. At the top of the upper layer ($h' = T_0'/\epsilon'$) the perturbation pressure must vanish ($D_0'(h') = 0$), so that with the introduction of a new integration constant K',

$$D_0' = K'e^{-\mu' c/2}\sinh N'(c - h'), \tag{111}$$

and consequently,

$$C_0' = K'e^{-\mu' c/2}\left[\frac{-(\sigma^2\mu'/2 - g)\sinh N'(c - h')}{Q'(\sigma^4\alpha^2 - g^2)} + \frac{\sigma^2 N' \cosh N'(c - h')}{Q'(\sigma^4\alpha^2 - g^2)}\right]. \tag{112}$$

At the interface ($c = 0$) the perturbation pressures and the vertical displacements must be continuous. Thus

$$D_0(0) = D_0'(0) \text{ and } C_0(0) = C_0'(0).$$

These conditions represent two linear and homogeneous equations for K and K', and a nontrivial solution exists only if the determinant of the foregoing system vanishes. This condition leads to the equation of frequency

$$Q_0\frac{\sigma^2 N \coth Nh + \sigma^2\mu/2 - g}{[\sigma^2(N - \mu/2) + g][\sigma^2(N + \mu/2) - g]} = -\frac{Q_0'}{\sigma^2\mu'/2 - g + \sigma^2 N' \coth N'h'}, \tag{113}$$

which gives the relation between wave velocity, wave length, and the parameters of the fluid system. Since σ is contained in the abbreviations N and N', the equation is transcendental. An approximation which is satisfactory for small-scale oscillations can be obtained by putting the hyperbolic cotangents equal to one which is strictly correct only for infinitely deep fluid layers. Then (113) changes into an algebraic equation which can be evaluated [9]. After an approximation has been obtained it can be used to substitute more accurate values for the hyperbolic cotangents and a new value for σ can be computed if necessary. Furthermore, the correction terms D_1, etc., can be computed, although it is found that the zero approximation gives a sufficiently good quantitative result, for instance in the theory of billow clouds.

Long Waves on a Rotating Plane. As an example of the application of perturbation equations when the earth's rotation is taken into account, the so-called trough formula given by Rossby [18] may be derived. The earth will be regarded as a flat, horizontal plane and the atmosphere as incompressible and homogeneous. The motion, disturbed as well as undisturbed, will be assumed as horizontal. If the atmosphere has a horizontal velocity U in the x-direction,

$$X = a + Ut, \qquad Y = b, \qquad Z = c.$$

It follows from (48) that for the undisturbed motion,

$$0 = -\frac{1}{Q}\frac{\partial P}{\partial a},$$

$$2\Omega_z U = -\frac{1}{Q}\frac{\partial P}{\partial b},$$

$$-g = -\frac{1}{Q}\frac{\partial P}{\partial c}.$$

The equation of continuity is satisfied for this type of motion since the density Q is constant. In the third equation above, the term with Ω_y has been omitted since it is small compared to g. In the following perturbation equations Ω_x and Ω_y will be set equal to zero as is customary when motion on a rotating plane is considered. Further, only the variation of Ω_z in the y-direction will be taken into account, a restriction which is justified if the undisturbed motion is predominantly zonal. Then, according to (54) and (55), the perturbation equations become

$$\frac{\partial^2 x}{\partial t^2} - 2\Omega_z \frac{\partial y}{\partial t} + 2\Omega_z U \frac{\partial y}{\partial a} + \frac{1}{Q}\frac{\partial p}{\partial a} = 0, \tag{114}$$

$$\frac{\partial^2 x}{\partial t^2} + 2\Omega_z \frac{\partial x}{\partial t} + 2\Omega_z U \frac{\partial y}{\partial b} + y \frac{\partial (2\Omega_z)}{\partial b} + \frac{1}{Q}\frac{\partial p}{\partial b} = 0, \tag{115}$$

$$\frac{\partial x}{\partial a} + \frac{\partial y}{\partial b} = 0. \tag{116}$$

In the last equation the variation of the height of the undisturbed fluid in the y-direction has not been taken into account. This height variation provides the meridional pressure gradient required by the undisturbed zonal current. It is equal to the small inclination of the isobaric surfaces and therefore presumably of little effect on the perturbation motion.

It may be assumed that

$$\eta = \frac{\partial (2\Omega_z)}{\partial b} = \text{const.}$$

The foregoing system of equations has coefficients which depend on b. It can therefore be satisfied if the unknown variables x, y, and p are assumed to be unknown functions of b and have the periodicity factor

$$\exp\{i\alpha[a - (\sigma - U)t]\}.$$

Instead of adopting this procedure we may directly eliminate all unknown variables but one by differentiations, a method which could also have been used in the preceding examples. The continuity equation (116) can be satisfied by a function $\psi(a, b, t)$ analogous to the stream function such that

$$x = \frac{\partial \psi}{\partial b}, \qquad y = -\frac{\partial \psi}{\partial a}.$$

By cross differentiation of (114) and (115) the vorticity equation is obtained,

$$\frac{\partial^2}{\partial t^2}(\nabla^2 \psi) + \eta \frac{\partial}{\partial t}\left(\frac{\partial \psi}{\partial a}\right) = 0, \tag{117}$$

where

$$\nabla^2 \equiv \frac{\partial^2}{\partial a^2} + \frac{\partial^2}{\partial b^2}.$$

Equation (117) is satisfied by the following expression

$$\psi = A \exp\{i\alpha[a - (\sigma - U)t] + i\delta b\}, \tag{118}$$

where A is an arbitrary constant, provided that

$$\sigma - U = -\frac{\eta}{\alpha^2 + \delta^2}. \tag{119}$$

The last relation is the trough formula which relates the speed of long waves in a zonal current to their wave length and width and to the parameters of the system. The formula has been discussed and applied widely in meteorological work.

Long Waves on a Rotating Globe. As an example of the application of the perturbation equations to motion on a sphere the same problem as in the previous section will be considered for a spherical fluid layer [11]. It will be assumed that the angular velocity of the undisturbed current, κ, is constant and that it is in the direction of the geographic longitude λ, so that the linear zonal velocity

$$V_\lambda = \kappa E \sin\vartheta, \tag{120}$$

where E is the earth's radius, and ϑ is the colatitude. Since vertical motion is neglected, we may substitute in the equations for polar coordinates on page 411 the earth's radius E instead of r because the vertical dimensions of the fluid layer are small compared to E. The following relations are thus obtained for the undisturbed motion. From (72)

$$-(\kappa + 2\Omega)\cos\vartheta\, V_\lambda = -\frac{1}{Q}\left(\frac{\partial P}{E\partial\vartheta}\right). \tag{121}$$

From (73) it follows that the undisturbed pressure field is independent of λ. The two terms on the left-hand side of (74) are so much smaller than the acceleration of gravity that hydrostatic equilibrium may be assumed, as in the preceding example. The equation of continuity is evidently satisfied by the assumed undisturbed motion.

As in the plane case the perturbation motion may be purely horizontal. Then the two equations for the horizontal motion are, according to (77) and (78), if (120) is noted,

$$\frac{\partial v_\vartheta}{\partial t} + \kappa \frac{\partial v_\vartheta}{\partial \lambda} - 2(\kappa + \Omega)\cos\vartheta\, v_\lambda = -\frac{1}{Q}\left(\frac{\partial P}{E\partial\vartheta}\right), \tag{122}$$

$$\frac{\partial v_\lambda}{\partial t} + \kappa \frac{\partial v_\lambda}{\partial \lambda} + 2(\kappa + \Omega)\cos\vartheta\, v_\vartheta = -\frac{1}{Q}\left(\frac{\partial p}{E\sin\vartheta\,\partial\lambda}\right). \tag{123}$$

Since static equilibrium is assumed, the third equation of motion need not be considered.

In analogy to (116) the small effects of the meridional variation of the height of the free surface may be disregarded so that the equation of continuity becomes, according to (80) and (76),

$$\left[\frac{1}{E\sin\vartheta}\right]\left[\frac{\partial(\sin\vartheta\, v_\vartheta)}{\partial\vartheta} + \frac{\partial v_\lambda}{\partial\lambda}\right] = 0. \tag{124}$$

Equation (124) may be satisfied by a stream function χ so that

$$v_\vartheta = -\frac{\partial \chi}{E\sin\vartheta\,\partial\lambda}, \qquad v_\lambda = \frac{\partial \chi}{E\partial\vartheta}. \tag{125}$$

By cross differentiation of equations (122) and (124) it follows that

$$\left(\frac{\partial}{\partial t}+\kappa\frac{\partial}{\partial\lambda}\right)\left[\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\,\frac{\partial\chi}{\partial\vartheta}\right)+\frac{1}{\sin\vartheta}\,\frac{\partial^2\chi}{\partial\lambda^2}\right] + 2(\Omega+\kappa)\,\frac{\partial\chi}{\partial\lambda}=0\,. \tag{126}$$

If it is assumed that the perturbations are waves of the frequency β and the wave number m,

$$\chi = Cf(\vartheta)\exp\left[i(\beta t+m\lambda)\right], \tag{127}$$

where $f(\vartheta)$ must satisfy the equation,

$$\frac{1}{\sin\vartheta}\,\frac{d}{d\vartheta}\left(\sin\vartheta\,\frac{df}{d\vartheta}\right) + \left[2(\Omega+\kappa)(\beta+\kappa m)-\frac{m^2}{\sin^2\vartheta}\right]f=0\,. \tag{128}$$

It is known that this equation has solutions which are regular at the poles only if

$$2(\Omega+\kappa)(\beta+\kappa m)=n(n+1), \tag{129}$$

where n is an integer. The last relation represents the frequency equation, and $f(\vartheta)$ is then an associated Legendre function, $P_n^m(\cos\vartheta)$. From (129) the period of the oscillation can be obtained as a function of the earth's rotation, zonal wind velocity (expressed by the angular velocity κ), wave number m, and the integer n. This last number determines the meridional extent of the oscillation since

$$f(\vartheta)=C\sin^m\vartheta\,\frac{d^m(P_n\cos\vartheta)}{d(\cos\vartheta)^m}=P_n^m(\cos\vartheta).$$

Thus $f(\vartheta)$ vanishes at the poles and at $n-m$ circles of latitude which are symmetrical to the equator. The equator itself is a nodal parallel if $n-m$ is odd.

Quasi-static and Quasi-geostrophic Approximations. In a discussion of the simplifications made to obtain tractable models of atmospheric motions it is appropriate to consider the hypothesis of "quasi-static" motion. This hypothesis is based on the assumption that, in each vertical, equilibrium exists not only before the start of the motion, but also during motion, although the equilibrium may change with time and from one vertical to another. This assumption is, of course, the basis of all the evaluations of upper-air soundings and oceanographic soundings which are performed with the aid of integrals of the hydrostatic equation. The quasi-static hypothesis implies that the vertical accelerations of the motion can be neglected compared to the acceleration of gravity. Since the latter is in general much larger than the former, the quasi-static hypothesis appears quite plausible, but it is impossible to give an a priori justification for it as Solberg [20], especially, has emphasized. Its most satisfactory justification is to be found in the fact that it gives results which in many instances are in agreement with the observations.

Historically, the quasi-static assumption was first introduced in the theory of tides. It is also successfully used more generally in the theory of "long" waves, that is of waves whose length is large compared to the depth of the fluid layer. The notation "quasi-static," which characterizes the special dynamic nature of the fluid motion more clearly than the expressions "tidal" or "long" waves, was introduced by V. Bjerknes [1] when he generalized the quasi-static treatment from incompressible and homogeneous fluid layers to autobarotropic layers.

In order to illustrate the method, consider a fluid layer which is at rest in the undisturbed case. It will also be supposed that the coordinate system is nonrotating and that all motions take place in a vertical a, c plane. The undisturbed pressure and density depend then only on the c coordinate and satisfy the hydrostatic equation

$$gQ=-\frac{\partial P}{\partial c}\,. \tag{130}$$

The horizontal equation of motion may be written, according to (54),

$$\frac{\partial^2 x}{\partial t^2}+\frac{\partial}{\partial a}\left(gz+\frac{p}{Q}\right)=0. \tag{131}$$

Because the vertical acceleration of motion may be neglected and because of (56) and of (130), the equation for the vertical component becomes

$$\frac{\partial}{\partial c}\left(gz+\frac{p}{Q}\right)=\frac{g(\Gamma-\gamma)p}{Q}\,, \tag{132}$$

where Γ and γ are the coefficients of barotropy and piezotropy, respectively. In the last equation the undisturbed pressure P may be introduced for the height coordinate c with the aid of (130). Then

$$\frac{\partial}{\partial P}\left(gz+\frac{p}{Q}\right)=-\frac{(\Gamma-\gamma)p}{Q^2}\,. \tag{133}$$

The equation of continuity

$$\frac{q}{Q}+\frac{\partial x}{\partial a}+\frac{\partial z}{\partial c}=0$$

may be written with the aid of the physical equation and of the hydrostatic equation (130),

$$\frac{1}{Q}\,\frac{\partial x}{\partial a}-g\,\frac{\partial z}{\partial P}+\frac{\gamma p}{Q^2}=0\,. \tag{134}$$

By combination of (133) and (134) it follows that

$$\frac{\partial x}{\partial a}+\frac{\partial p}{\partial P}=0. \tag{135}$$

This equation relates the perturbation pressure to the horizontal divergence. It expresses the pressure as an effect of mass transport. If (131) is differentiated with respect to P and equation (134) is used, one obtains the relation:

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial x}{\partial P}\right)-\frac{(\Gamma-\gamma)}{Q^2}\,\frac{\partial p}{\partial a}=0. \tag{136}$$

The last equation becomes particularly simple if the fluid is autobarotropic. In this case the horizontal displacement is independent of the vertical coordinate. It is this property of the system of equations for quasi-static motion which permits the most important simplifications of which use is made, for instance, in the theory of ocean tides where as a rule only incompressible, homogeneous fluids are considered. For fluids which are not autobarotropic this simplification no longer holds, but it is in some instances possible to consider a fluid model which consists of a number of autobarotropic layers where the approach to more realistic baroclinic models is effected by assuming autobarotropic relations which differ from layer to layer. A rather comprehensive discussion of quasi-static motions has been given by Eliassen [8].

The limitations of the quasi-static hypothesis especially in its applications on the tidal theory have been discussed by Solberg [20]. The differential equation remains of the same type under the quasi-static assumption regardless of whether the oscillations are longer or shorter than half a sidereal day; but when the vertical acceleration is taken into account it changes from the elliptical through the parabolic to the hyperbolic type as the period increases to more than twelve sidereal hours. Under these conditions it appears possible that oscillations with periods longer than half a sidereal day are not reproduced very accurately by the quasi-static hypothesis, or that additional types of oscillations are possible which are not given by the quasi-static method.

Another way in which the hydrodynamic equations might successfully be simplified for investigations in theoretical meteorology and oceanography is indicated by the empirically known fact that the large-scale motions of the free atmosphere are nearly geostrophic. It thus appears possible when attention is to be focused on these large-scale motions to achieve substantial simplifications by assuming that the wind field is very nearly geostrophic. Such a "geostrophic" approximation has been developed by Charney [5] in systematic form from considerations of the orders of magnitude of the various meteorological variables. The most important step in this "quasi-geostrophic method" is the elimination of the horizontal divergence of motion from the system of hydrodynamic equations before the introduction of the geostrophic approximation. This elimination can be effected conveniently by means of the equation of continuity together with an equation expressing the conservativeness of a quantity such as the potential temperature. While for the vertical vorticity component, for instance, it is permissible to use the geostrophic approximation directly, the same is not true for the horizontal divergence. From a consideration of the orders of magnitude of $\partial u/\partial x$ and $\partial v/\partial y$ it follows that for the large-scale motions, these terms are one order of magnitude larger than their sum, the horizontal divergence. Since the geostrophic deviation, that is the difference between actual and geostrophic wind, is also one order of magnitude smaller than the wind itself it follows that a computation of the horizontal divergence by the geostrophic wind would not be a satisfactory approximation. By the elimination of the horizontal divergence in the manner indicated above, a consistent system of approximate equations for large-scale motions is obtained.

A linearization of this quasi-geostrophic system of equations has been used by Charney and Eliassen [6] as the basis for a numerical method to predict the future pressure distribution. This method and its extension to nonlinear mathematical models is discussed elsewhere in this Compendium.[2] But it is relevant to the topic of the present article to state that even the linear models based on the assumption of small perturbations have given satisfactory forecasts in a number of cases.

CONCLUSION

The atmospheric perturbation equations are a tool of theoretical meteorology which is to be used in the investigation of problems of atmospheric dynamics. In considering future work in this field we shall therefore concern ourselves with a general survey of types of problems and lines of attack in which the perturbation method may appropriately be used.

As was pointed out in the discussion of the basic assumptions of the perturbation theory, one of the important problems which is appropriately treated by the perturbation method is that of the stability or instability of a given atmospheric flow pattern. It was explained there that the method consists in superimposing a small perturbation on the basic flow pattern and investigating whether such a perturbation would increase with time or not. In the first case the basic flow pattern would be unstable, in the second case stable.

It is immediately apparent that such an instability investigation is not necessarily complete. When it has been found that an originally small perturbation increases with time it can be concluded that the perturbation equations do not describe satisfactorily the development of the perturbation beyond a certain state because sooner or later the perturbation will have grown to such a magnitude that the terms of higher than the first order in the equations can no longer be neglected. Thus, after the perturbation has grown to certain dimensions its future behavior can no longer be predicted by the original perturbation equations. It is possible that the growth of the perturbation ceases when this state is reached and, in fact, it is generally observed that atmospheric perturbations do not exceed certain limits. The foregoing remarks are not meant to imply that the study of stability and instability based on the perturbation method is unsatisfactory. The point is that, by means of the perturbation method, we can make statements about the development of the perturbation only for a limited time interval. It would evidently be highly desirable to extend these investigations in such a manner that they permit us

2. Consult "Dynamic Forecasting by Numerical Process" by J. G. Charney, pp. 470–482.

to follow the future life history of an unstable perturbation. In order to do so one has to find solutions of the complete hydrodynamic equations in their nonlinear forms, a mathematical problem which is much more complicated than the solution of linear problems because no general mathematical procedures exist for its solution. It is conceivably possible that solutions of the nonlinear problem of the later development of unstable perturbations can be obtained by successive approximations, starting out with the perturbation as described by the linear perturbation equations and obtaining the next approximation as a perturbation of the first approximation. However, it is doubtful if such a method would be simpler than a more direct approach which starts immediately with a consideration of the nonlinearized hydrodynamic equations. No matter which method of approach will be found feasible, a quantitative study of developing perturbations beyond their nascent linear stages appears indispensable for the future development of the perturbation theory.

Such a study may also be expected to shed additional light on another question connected with the perturbation theory, namely, the problem of how an unstable flow pattern can develop. If it is found that a given state of fluid motion is unstable, the implication is that any small disturbance will increase indefinitely in amplitude and lead to a breakdown of the original, undisturbed state. Since in a fluid system such as the atmosphere or the ocean such small disturbances are always present, it is difficult to see offhand how an unstable state could develop and exist for any appreciable length of time unless during a certain state of its development the factors which built up this state are more effective than those which contribute to its eventual breakdown due to its inherent instability. Perhaps the effect of friction is responsible for delaying the breakdown. In the stability investigations it is always assumed that the undisturbed state whose instability one wishes to investigate is given a priori. Supplementary investigations of the mechanism leading to its development are evidently in order. Again, as in the case of the nonlinear perturbations, it is doubtful whether the linear perturbation equations are the appropriate mathematical tool for such a study.

Even though the foregoing remarks indicate that for a future development of instability studies the investigation of nonlinear problems, either by approximation methods or by direct means, will be an important step forward, a continuation of studies along the lines conducted so far will deserve an important place in theoretical meteorology. For a first investigation whether a given state of motion is unstable, the linearized perturbation equations are evidently the appropriate system of equations. Since the model atmospheres whose stability conditions have been investigated so far are all only more or less close approximations to reality, it is evident that much work remains to be done in the study of progressively more realistic models. In view of the rapidly mounting difficulties of such analyses, as baroclinic wind fields, compressibility, and horizontally and vertically variable temperature distributions are taken into account, it is not likely that all fruitful problems which can be treated by the theory of linear perturbations will find an early solution.

It is necessary to make simplifying assumptions such as those discussed in the last section, namely, that the motion is quasi-static or quasi-geostrophic. These assumptions are physically plausible and permit the omission of certain terms in the equations. In view of the complexity of the equations such procedures are indispensable. In these simplifications, terms are neglected which are of smaller orders of magnitude than others. This may give rise to errors in the subsequent analysis because in the reduction of the system of differential equations to one equation it is necessary to carry out differentiations in order to eliminate all unknown variables but one. It is conceivable that while of two quantities one may be considerably larger than the other their derivatives may be of the same order of magnitude. Consequently, the omission of terms in the original system of equations may lead to erroneous equations when the necessary eliminations are carried out. This possibility has been discussed by Queney [16] who pointed to this as a possible explanation of some contradictory results which have been obtained by different authors. It is not proposed to discuss this question here but at any rate this controversy shows as stated before that a great number of problems remain to be settled with the aid of the perturbation equations.

The problems to be studied by means of the perturbation method should include questions not only of the stability and instability of given fluid states, but also of the fluid motions caused by these perturbations. The solution of the linearized equations of fluid motion has led to many satisfactory descriptions of fluid motions as, for instance, the theory of tidal and surface waves, of sound waves, and of atmospheric waves both where a smaller scale is involved, for instance, in the theory of mountain waves and of billow clouds, and where motions of a much larger scale are considered.

It may finally be repeated that very little work has been done so far to extend the perturbation theory to viscous fluids. Even though friction plays presumably only a minor role in the free atmosphere its effect is quite important near the ground and at very high levels, in the ionosphere, where the kinematic viscosity must reach large values. Along the same lines, presumably more attention will have to be given to the problem of heat conduction in perturbation motions. Such problems as the land and sea breeze, or the monsoon circulations, require that the conduction from the earth's surface into the atmosphere and the eddy conduction within the atmosphere be taken into account. Among other nonadiabatic processes which have received little attention in connection with the perturbation theory is radiation. Studies of perturbation motions when radiative transfer of heat occurs can presumably shed additional light on atmospheric circu-

lations. It would be easy to extend this list of new problems considerably. But it may suffice to say that the perturbation method almost invariably will have to be used to get a first insight into these problems because in all branches of mathematical physics linear differential equations are the only ones which can in general be dealt with successfully. Even in those deplorably rare instances when nonlinear problems can be handled, the linearized problem and its solution gives a valuable and helpful guidance in the solution of the nonlinear problem.

REFERENCES

1. Bjerknes, V., "On Quasi Static Wavemotion in Barotropic Fluid Strata." *Geofys. Publ.*, Vol. 3, No. 3 (1923).
2. —— "Die atmosphärischen Störungsgleichungen." *Beitr. Phys. frei. Atmos.*, 13:1–14 (1927).
3. —— "Über die hydrodynamischen Gleichungen in Lagrangescher und Eulerscher Form und ihre Linearisierung für das Studium kleiner Störungen." *Geofys. Publ.*, Vol. 5, No. 11 (1929).
4. —— and others, *Physikalische Hydrodynamik*. Berlin, J. Springer, 1933.
5. Charney, J. G., "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2 (1948).
6. —— and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
7. Craig, R. A., "A Solution of the Nonlinear Vorticity Equation for Atmospheric Motion." *J. Meteor.*, 2:175–178 (1945).
8. Eliassen, A., "The Quasi-static Equations of Motion with Pressure as Independent Variable." *Geofys. Publ.*, Vol. 17, No. 3 (1949).
9. Haurwitz, B., "Über Wellenbewegungen an der Grenzfläche zweier Luftschichten mit linearem Temperaturgefälle," *Beitr. Phys. frei. Atmos.*, 19:47–54 (1932).
10. —— "Über die Wellenlänge von Luftwogen," 2. Mitt. *Beitr. Geophys.*, 37:16–24 (1932).
11. —— "The Motion of Atmospheric Disturbances on the Spherical Earth." *J. mar. Res.*, 3:254–267 (1940).
12. Jeffreys, H., "On the Dynamics of Wind." *Quart. J. R. meteor. Soc.*, 48:29–47 (1922).
13. Lamb, H., "On Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 84:551–572 (1910).
14. Langwell, P. A., "Forced Convection Cell Circulation in Clear Air." *Trans. Amer. geophys. Un.*, 32:7–14 (1951).
15. Neamtan, S. M., "The Motion of Harmonic Waves in the Atmosphere." *J. Meteor.*, 3:53–56 (1946).
16. Queney, P., "Adiabatic Perturbation Equations for a Zonal Atmospheric Current." *Tellus*, 2:35–51 (1950).
17. Rombakis, S., "Über ein Integral der nichtlinearen hydrodynamischen Gleichungen und seine Anwendung in der Meteorologie." *Z. Meteor.*, 2:241–244 (1948).
18. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semipermanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
19. Solberg, H., "Integrationen der atmosphärischen Störungsgleichungen." *Geofys. Publ.*, Vol. 5, No. 9 (1928).
20. —— "Über die freien Schwingungen einer homogenen Flüssigkeitsschicht auf der rotierenden Erde. I." *Astrophys. norveg.*, 1:237–340 (1936).
21. Starr, V. P., and Neiburger, M., "Potential Vorticity as a Conservative Property." *J. mar. Res.*, 3:202–210 (1940).

THE SOLUTION OF NONLINEAR METEOROLOGICAL PROBLEMS BY THE METHOD OF CHARACTERISTICS

By JOHN C. FREEMAN

The Institute for Advanced Study

INTRODUCTION

The Method of Characteristics. The method of characteristics is only one means of solving hyperbolic differential equations or systems of differential equations with real characteristics. Problems of this kind are also solved by related methods, such as simple wave theory, the Riemann method of integration, and the marching method. In addition, in a nonlinear system, shocks or jumps occur which cannot be studied by means of systems of equations with real characteristics, but must be studied by other methods. Loosely speaking, all methods mentioned above can be grouped under the single heading—the method of characteristics. If this definition is used, the method of characteristics is not new to meteorology. Richardson used the marching method correctly in the initial example in his monumental work on numerical weather prediction [15, pp. 6–10]. Rossby [16] used the Riemann method of integration to discuss the effect of a line source of planetary waves in a barotropic atmosphere. In all of its phases, however, the method of characteristics has been applied most often and most thoroughly to problems in gas dynamics and the flow of a shallow layer of fluid with a free surface. Riemann devised his method of integration to solve the problem of one-dimensional unsteady flow of a gas in a pipe, and most other methods have been developed with such problems in mind. Courant and Friedrichs [5] give a bibliography of this work.

Early Work on Hydraulic Jumps in the Atmosphere. Mathematically speaking, the theory of flow of a shallow fluid with a free surface is, to a certain degree of approximation, the same as the theory of flow of air under a shallow (1000–10,000 ft) inversion in the atmosphere [7]. A very striking phenomenon in the flow of shallow water with a free surface is the *bore* or *jump*. A model of the jump is the breaker on a gently sloping beach. A breaker moving over a surface, whose horizontal dimensions are about ten times the height of the breaker, may be viewed as a line along which the height and speed of water change abruptly. In fact, in some cases there is a very shallow outgoing flow near the shore, and the jump or breaker is a transition to a very deep incoming flow. When such a jump is stationary in a channel, it is called a hydraulic jump. If the force of gravity is modified by a buoyancy factor (which is unity in the case of flow in water, and depends on the ratio of the densities above and below an inversion in the atmosphere), the laws governing flow under an inversion are the same as those governing flow of shallow water. Thus, jumps can occur in the atmosphere.

A jump in the atmosphere was first recognized as such (to the writer's knowledge) and discussed by M. McGurrin [14] of the San Bruno Weather Bureau Forecast Center. He recognized that the jump might be the cause of several weather phenomena and gave a rather complete discussion of the equations for a steady-state jump in an inversion. His work has not been published. Mr. McGurrin has addressed local meetings of the American Meteorological Society in California, and probably because of his work the concept of a jump in an inversion is not new to many meteorologists. It should be emphasized here that McGurrin's paper [14] shows that he was completely aware that the steady-state jump could occur in the atmosphere on many scales and in many synoptic situations. He describes a stationary jump in the height of a fog bank. The winds blow through this jump as they blow into the San Bruno valley in California.

The method of characteristics is the application of methods related to those employed by Richardson [15] and Rossby [16] to problems related to those considered by McGurrin. Rossby [16] predicted that a study of the internal waves in the atmosphere would show that they develop sharp forward boundaries because of their dispersive character and the existence of a maximum value of energy propagation. The waves on an inversion are a limiting case of internal waves and have the properties which he predicted.

MATHEMATICAL FOUNDATIONS

Quasi-linear Differential Equations. We can use the method of characteristics, under certain conditions, on equations of the following type:

$$A_1 \frac{\partial u}{\partial t} + B_1 \frac{\partial u}{\partial x} + C_1 \frac{\partial w}{\partial t} + D_1 \frac{\partial w}{\partial x} = E_1, \qquad (1)$$

$$A_2 \frac{\partial u}{\partial t} + B_2 \frac{\partial u}{\partial x} + C_2 \frac{\partial w}{\partial t} + D_2 \frac{\partial w}{\partial x} = E_2. \qquad (2)$$

In these equations (for the moment) the independent variables x and t are Cartesian coordinates in the x, t plane. The dependent variables u and w are unknown functions of x and t. The coefficients A_1, B_1, etc., are known functions of x, t, u, and w. Any of them can be constant or zero. These equations are nonlinear if any of the coefficients, A_1, B_1, etc., are functions of u or w. They are called quasi-linear because they are linear in the derivatives of u and w and the method of solution does not depend strongly on their nonlinearity. This type of equation has been studied by many authors, and bibliographies concerning such studies appear in Courant and Friedrichs [5] and Isenberg [12]. The discussion here is very much like that of Isenberg.

Definition and Determination of Characteristics. The

characteristics of the system of equations (1) and (2) are defined as lines along which the partial derivatives $\partial u/\partial t$, $\partial u/\partial x$, $\partial w/\partial t$, and $\partial w/\partial x$ cannot be determined directly from the equations. These lines are very useful. It can be shown that these are lines along which discontinuities in the partial derivatives occur. For instance, in the rapid steady-state flow of water along a curb, a small notch in the curb causes small discontinuities in the derivative of the height of the water and a line of disturbed water slants out and downstream from the notch. Such lines and similar ones in steady-state gas flows are called *Mach lines*. Thus, in a steady-state flow, disturbances along the characteristics can actually be seen.

In order to find the characteristics of equations (1) and (2) a well-known equation from elementary calculus must be used. If u and w are functions of x and t, then

$$du = \frac{\partial u}{\partial t}\,dt + \frac{\partial u}{\partial x}\,dx, \tag{3}$$

and

$$dw = \frac{\partial w}{\partial t}\,dt + \frac{\partial w}{\partial x}\,dx. \tag{4}$$

Equations (1)–(4) are now looked upon as a system of linear algebraic equations for the unknown functions $\partial u/\partial t$, $\partial u/\partial x$, $\partial w/\partial t$, and $\partial w/\partial x$. They have the matrix

$$\begin{Vmatrix} A_1 & B_1 & C_1 & D_1 & E_1 \\ A_2 & B_2 & C_2 & D_2 & E_2 \\ dt & dx & 0 & 0 & du \\ 0 & 0 & dt & dx & dw \end{Vmatrix}. \tag{5}$$

From this it is seen that the value of $\partial u/\partial t$ at any point (x, t) is expressed in terms of u, w, x, and t by the equation

$$\frac{\partial u}{\partial t} = \begin{vmatrix} E_1 & B_1 & C_1 & D_1 \\ E_2 & B_2 & C_2 & D_2 \\ du & dx & 0 & 0 \\ dw & 0 & dt & dx \end{vmatrix} \div \begin{vmatrix} A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \\ dt & dx & 0 & 0 \\ 0 & 0 & dt & dx \end{vmatrix}. \tag{6}$$

The condition that $\partial u/\partial t$ be indeterminate is that the denominator and numerator of the right-hand side of equation (6) be zero. This can be expressed for the denominator as follows:

$$\begin{aligned}(A_1C_2 - A_2C_1)\,dx^2 \\ - (A_1D_2 - A_2D_1 + B_1C_2 - B_2C_1)\,dx\,dt \\ + (B_1D_2 - B_2D_1)\,dt^2 = 0.\end{aligned} \tag{7}$$

This equation is solved (by using the quadratic formula) for dx/dt or dt/dx, whichever is desired. This is best done in specific cases where usually many of the coefficients, A_1, B_1, etc., are zero, but the symbolic solutions will be useful:

$$\begin{aligned} C_+\text{:}\quad & \frac{dx}{dt} = \alpha_+, \\ C_-\text{:}\quad & \frac{dx}{dt} = \alpha_-. \end{aligned} \tag{8}$$

These equations define the two characteristics (C_+, C_-) of equations (1) and (2). Note that the characteristics are functions of u, w, x, and t and not of their derivatives.

The Equations of Compatibility. The numerator of equation (6) must be zero where the denominator is zero. This condition is expressed by writing out the determinant in the numerator to obtain

$$\begin{aligned}\left(B_1\frac{du}{dt} - E_1\frac{dx}{dt}\right)\left(C_2\frac{dx}{dt} - D_2\right) \\ - \left(B_2\frac{du}{dt} - E_2\frac{dx}{dt}\right)\left(C_1\frac{dx}{dt} - D_1\right) \\ - \frac{dw}{dt}\frac{dx}{dt}(C_1D_2 - D_1C_2) = 0.\end{aligned}$$

Since this is to be true along the characteristics, the appropriate values of dx/dt are substituted in the equation. Then two equations of compatibility, one for each of the two roots dx/dt of equation (7), are obtained,

$$\begin{aligned}\left(B_1\frac{du}{dt} - E_1\alpha_+\right)(C_2\alpha_+ - D_2) \\ - \left(B_2\frac{du}{dt} - E_2\alpha_+\right)(C_1\alpha_+ - D_1) \\ - \frac{dw}{dt}\alpha_+(C_1D_1 - D_1C_2) = 0\end{aligned} \tag{9}$$

along $\dfrac{dx}{dt} = \alpha_+$, and

$$\begin{aligned}\left(B_1\frac{du}{dt} - E_1\alpha_-\right)(C_2\alpha_- - D_2) \\ - \left(B_2\frac{du}{dt} - E_2\alpha_-\right)(C_1\alpha_- - D_1) \\ - \frac{dw}{dt}\alpha_-(C_1D_1 - C_2D_1) = 0\end{aligned} \tag{10}$$

along $\dfrac{dx}{dt} = \alpha_-$.

Numerical Integration by the Method of Characteristics. The conditions (9) and (10) form the basis for the method of numerical integration which is called the method of characteristics. The approximation is made that if two points (t_i, x_i) and (t_j, x_j) are very close together on a curve, then the slope of the curve is

$$\frac{dx}{dt} = \frac{x_j - x_i}{t_j - t_i}.$$

Similar approximations are made concerning du/dt and dw/dt. The values of u and w are given at two points (t_1, x_1) and (t_2, x_2) not connected by a characteristic and it is required to find u and w at some other point (t_3, x_3) (not given) where it is unknown. We define t_3 and x_3 as the intersection of the characteristic line with slope α_+ through (t_1, x_1) and the characteristic line with slope α_- through (t_2, x_2). Therefore, t_3 and

x_3 can be found by solving the two simultaneous equations

$$\frac{x_3 - x_1}{t_3 - t_1} = \alpha_+; \qquad \frac{x_3 - x_2}{t_3 - t_2} = \alpha_-.$$

These values of t_3 can be substituted in the equations (9) and (10):

$$\left(B_1\frac{u_3 - u_1}{t_3 - t_1} - E_1\alpha_+\right)(C_2\alpha_+ - D_2) - \left(B_2\frac{u_3 - u_1}{t_3 - t_1} - E_2\alpha_+\right)(C_1\alpha_+ - D_1) - \frac{w_3 - w_1}{t_3 - t_1}\alpha_+(C_1D_2 - D_1C_2) = 0,$$

$$\left(B_1\frac{u_3 - u_2}{t_3 - t_2} - E_1\alpha_-\right)(C_2\alpha_- - D_2) - \left(B_2\frac{u_3 - u_2}{t_3 - t_2} - E_2\alpha_-\right)(C_1\alpha_- - D_1) - \frac{w_3 - w_2}{t_3 - t_2}\alpha_-(C_1D_2 - D_1C_2) = 0,$$

and these two equations can be solved for u_3 and w_3 because t_3 and x_3 are known. By proceeding in this manner from a line (such as the boundary of a flow, or an initial state) in the x, t plane, along which u and w are known, values of u and w over a portion of the (x, t) plane can be found. For instance, in Fig. 1 the

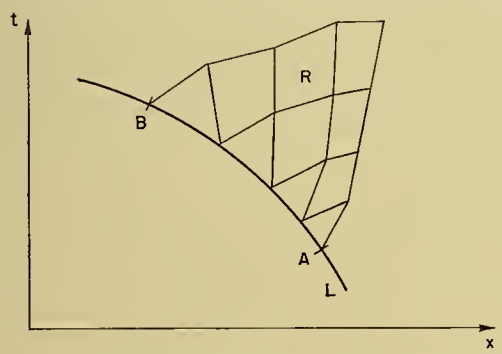


FIG. 1.—The region in which u and w can be computed from values given on L.

data between A and B on L allow a computation of the values of u and w in the hatched region R. For a full discussion of such "domains of dependence" and the related "regions of influence" the reader is referred to Courant and Friedrichs [5]. This method of integrating even the most complicated of such equations is easy to derive but is difficult to carry out in practice. Some simpler problems are easier to discuss.

Simple Waves. If, in equations (1) and (2), $E_1 = E_2 = 0$ and A_1, A_2, B_1, B_2, etc., are not functions of x or t but only functions of u and w, then the factor $1/dt$ can be removed from equations (9) and (10). These equations then define two functions of u and w (in implicit form, to be sure):

$$H(u,w) = K_+ \qquad \text{along } C_+,$$
$$L(u,w) = K_- \qquad \text{along } C_-,$$

where in general the constants K_+ are different for each C_+ characteristic and similarly for K_- and C_-. Courant and Friedrichs have shown that a region of constant u and w, say u_0 and w_0, is separated from a region of varying u and w by a characteristic. Assume that this is a C_+ characteristic. Since C_- characteristics intersect C_+ characteristics, some of the C_- characteristics are common to regions in which u and w are constant (u_0 and w_0) and to regions in which they vary. Along such C_- characteristics we can say

$$L(u, w) = L(u_0, w_0) \qquad \text{along } C_-.$$

This is true along every C_- characteristic extending into the region of constant u and w. Thus along such C_- characteristics u is a function of w only. Since these C_- characteristics cover all of a certain region, we need only have the assurance that we are considering motion inside that region to know that u is a function of w only. Inside such a region, if w remains constant along a line, u remains constant and vice versa.

We now focus our attention on the C_+ characteristics in this region where w is a function of u only. We can say $w = k(u)$; we then have

$$H(u, k(u)) = M(u) = K_+ \qquad \text{along } C_+.$$

This shows that u is constant along C_+. From equation (6) the slope of C_+ is given by

$$\frac{dx}{dt} = \alpha_+.$$

Now we have assumed (for simple waves) that α_+ is a function of u and w. Since u has been shown to be constant along C_+, and therefore w is constant along C_+, this means that α_+ is constant along C_+ or that C_+ has a constant slope and is then a straight line. When a flow has a family of straight characteristics it is called a *simple wave*. Much of Courant and Friedrichs's book *Supersonic Flow and Shock Waves* is concerned with simple waves and the reader is referred to it for a comprehensive discussion of them.

The two basic types of simple waves are the *expansion wave* and the *compression wave*. The names are derived from gas dynamics, but apply equally well to the expansion or compression of the distance between two straight characteristic lines in the general case presented above. If the given values of one of the dependent variables u and w, say u, along a line L are such that the straight characteristics in a simple wave diverge as they move away from L, then there is an expansion wave in that region (see Fig. 4). Note that, since u is constant along these lines, as we move away from L along a characteristic, the distance to the nearest point where u differs by a fixed amount is increasing. Thus gradients of u are decreasing. A more concrete example of such an expansion wave is given in the next section. If the given values of u are such that the straight characteristics converge, there is a compression wave in the region (see Fig. 2) and the gradients of u are increasing.

Envelope of the Characteristics. Figure 2 shows that a family of converging straight lines defines a double

envelope which encloses all points in the x, t plane covered by two or more characteristics. The meaning of such an envelope depends almost entirely on the physical nature of the problem and of the quantity u. The envelope has meaning if it is possible to have two values of u at a point. Usually the assumptions involved in deriving equations (1) and (2) are violated before the envelope is formed by the characteristics. If the region in which equations (1) and (2) cannot be used is very small and motion inside it is not important, it is assumed to be so thin that it can be drawn as a line in the x, t plane and it is called a *jump*. It is so called because there are jumps in values of u and w on going across such a line. Such an assumption can be made in gas dynamics where the jump is called a *shock*, in hydraulics where it is called a *hydraulic jump*, in flow under an inversion where it is called a *pressure jump*, and to a certain extent in flow in a planetary jet stream where the jump is called a *block*.

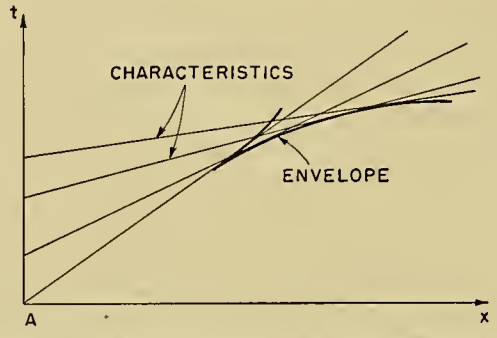


FIG. 2.—The characteristics of a compression wave.

FLOW UNDER AN INVERSION

Statement of the Problem. The changes in height of an inversion and the flow beneath an inversion are of great interest to synoptic meteorologists. The frontal contour charts of the Canadian Weather Service [6] bring home the fact that motions of the frontal surface are important even far behind its intersection with the ground. The behavior of low-level winds during and just after the passage of a surface cold front in the Western Plains is usually a topic of discussion among forecasters. Finally, the forecasters in tropical areas are interested in the variation of the winds below the trade wind inversion. It is the writer's opinion that a mathematical theory of flow under an inversion will help to describe these phenomena and perhaps will eventually lead to quantitative forecasting techniques. Several phenomena have been described by means of the simplest such theory (see Freeman [7], Abdullah [1], and Tepper [19]). Flow under a widespread inversion in the atmosphere can be approximated, at least somewhat better than qualitatively, by the flow of a shallow layer of liquid under a deep layer of almost the same density. The details of the validity of this approximation can be found elsewhere [7, 8, 18]. The pressure in a shallow layer of fluid flowing under a deep layer of fluid at rest is given by the formula

$$P = (h' - h)\rho' g + (h - z)\rho g,$$

where h is the height of the fluid with density ρ, and h' is the much greater height of the surface of the fluid with density ρ' ($\rho' < \rho$); z is the height of the point at which the pressure is measured, and g is the acceleration of gravity. If this is substituted in the two-dimensional equations of motion (neglecting the earth's rotation), we obtain the following results:

$$\frac{du}{dt} = -\frac{1}{\rho}\frac{\partial P}{\partial x} = -g\left(1 - \frac{\rho'}{\rho}\right)\frac{\partial h}{\partial x}, \tag{11}$$

$$\frac{dv}{dt} = -\frac{1}{\rho}\frac{\partial P}{\partial y} = -g\left(1 - \frac{\rho'}{\rho}\right)\frac{\partial h}{\partial y}. \tag{12}$$

If the flow under the interface at h does not depend on z, the equation of continuity is

$$\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = 0. \tag{13}$$

Equations (11), (12), and (13) are the equations of motion and continuity for two-dimensional unsteady flow of a shallow layer of liquid under a deep liquid of smaller density. These equations are difficult to integrate as they stand, and very likely high-speed computing machines would be required to solve even the simplest problems involving all three variables x, y, and t. If the dimensions are restricted, however, results can be found numerically or graphically with comparative ease.

Time-Dependent Flow under an Inversion. If there is no variation in the y direction during the time a flow is studied, it can be investigated as a one-dimensional unsteady flow. If v is zero initially and remains zero, equation (12) need not be considered and equations (11) and (13) become

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + g\left(1 - \frac{\rho'}{\rho}\right)\frac{\partial h}{\partial x} = 0 \tag{14}$$

and

$$\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + h\frac{\partial u}{\partial x} = 0, \tag{15}$$

respectively. This is a system of quasi-linear partial differential equations of the type discussed in the previous section. If, in equations (1) and (2), the conditions

$$\begin{aligned} A_1 &= 1, & B_1 &= u, & C_1 &= 0, \\ A_2 &= 0, & B_2 &= h, & C_2 &= 1, \\ D_1 &= g\left(1 - \frac{\rho'}{\rho}\right), & D_2 &= u, \\ E_1 &= 0, & u &= u, \\ E_2 &= 0, & w &= h, \end{aligned} \tag{16}$$

are substituted, equations (14) and (15) result. If the values (16) are substituted in equation (7), the characteristics of (14) and (15) become

$$\begin{aligned} \frac{dx}{dt} &= u \pm \sqrt{gh\left(1 - \frac{\rho'}{\rho}\right)} \\ &= u \pm c. \end{aligned} \tag{17}$$

The values from equation (16), substituted in equations (9) and (10), give the results

$$u + 2c = \text{const} \qquad \text{along } \frac{dx}{dt} = u + c \qquad (18)$$

and

$$u - 2c = \text{const} \qquad \text{along } \frac{dx}{dt} = u - c. \qquad (19)$$

Thus the numerical or graphical computing scheme described for equations (1) and (2) can be performed using equations (18) and (19).

Simple Waves.[1] Since A_1, etc., are not explicit functions of x and t and $E_1 = E_2 = 0$, with proper initial conditions there can be simple waves in the model described by equations (14) and (15). This is a powerful method of investigating the results of certain initial conditions under an inversion. In addition, these equations are a suitable set to make the idea of a simple wave clear. A simple wave results when u and h are constant along the straight characteristic lines

$$\frac{dx}{dt} = u + c. \qquad (20)$$

Since these straight lines can be determined from the initial conditions, or conditions along any line in the x, t plane, u and h can be found throughout a region of the x, t plane by drawing the characteristic lines covering that region.

Expansion Waves. An expansion wave under an inversion is best demonstrated by the following example. The air under an inversion is at rest in a channel formed by two mountain ranges which restrict motion normal to them. It is held at rest by meteorological conditions at the termination of the mountain ranges. If these conditions change so that the air under the inversion begins to flow out from between the ranges, an expansion wave results. Figure 3 shows the height of the inversion and the wind speed along a cross section through the center of the channel under these conditions. The flow described qualitatively in Fig. 3 can be described quantitatively in the x, t plane (see Fig. 4). The observation of u at A and the initial height of the inversion at rest form the boundary conditions that determine the flow to the right of A. These values of u determine c and therefore $u + c$. Since u is constant along the lines $dx/dt = u + c$, we can compute it throughout the region between the mountains. This example emphasizes the ease with which the flow can be discussed if it is made up of simple waves. Note from formula (19), which applies in this case, that under these conditions the inversion height decreases as the speed of flow outward from between the mountains increases, that is, an increasing speed of east wind is a decreasing wind in this coordinate system. An example of an expansion wave that has synoptic importance will be given in the discussion of squall lines. It should be emphasized that the example of this paragraph is used for its definiteness and simplicity. Expansion waves in the atmosphere need not occur between mountain ranges and need not

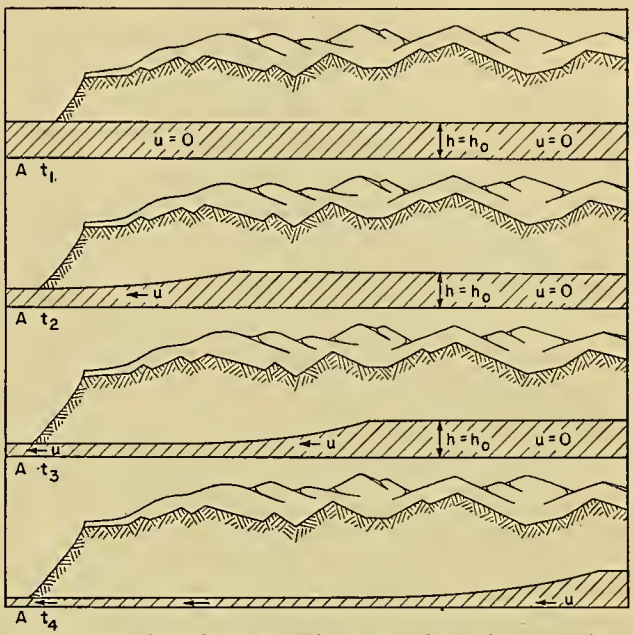


Fig. 3.—Motion of an expansion wave shown in successive cross sections.

be associated with air initially at rest. The whole system of flow (in particular, the zone of still air) can have a constant velocity of any direction superimposed on

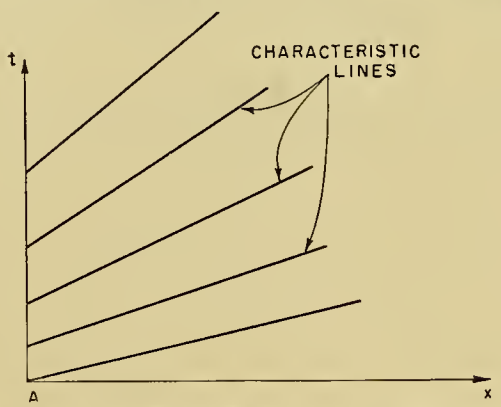


Fig. 4.—The characteristics of an expansion wave.

it. The rigid boundary is not necessary. Any flow which is essentially one-dimensional can be studied by the means presented here.

Compression Waves. If the flow at the mouth of the valley changes from zero velocity to a flow into the valley, a compression wave results. A cross section of such a compression wave for successive times is given in Fig. 5. The x, t diagram for such a flow is given in Fig. 6. Again the values of u and the inversion height at A are the boundary conditions that determine the flow. We have seen that a compression wave leads to an envelope of the characteristic lines. In this case the envelope cannot persist for any great length of time so that a pressure jump forms.

1. The word "wave" is used here in its general sense, familiar to meteorologists in "cold wave," "frontal wave," etc. No periodicity or sinusoidal properties are implied.

Pressure Jumps. The envelope of the characteristics illustrated in Fig. 2 has a definite physical meaning in the case of flow under an inversion. It means that the

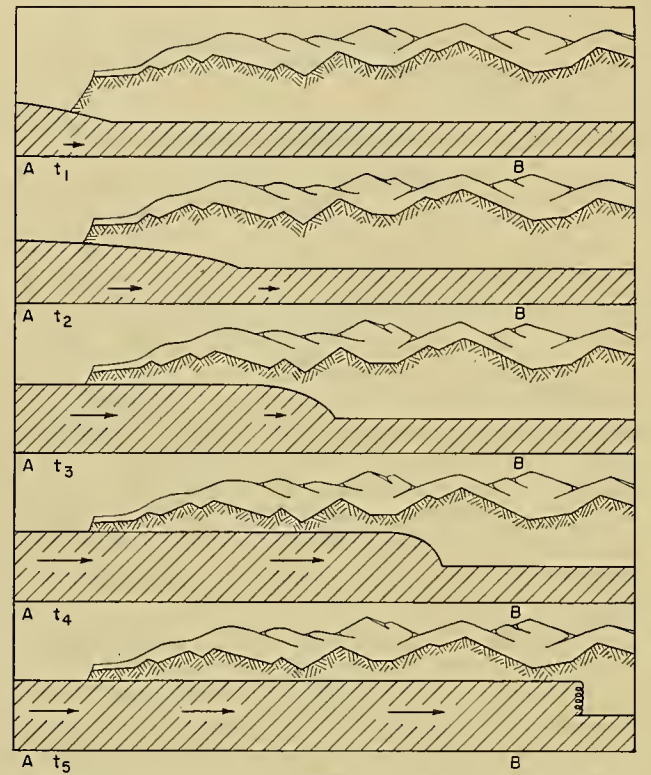


Fig. 5.—Motion of a compression wave shown in successive cross sections. The slope of the wave increases until it "breaks" and a pressure jump is formed.

higher values of inversion height have overtaken the lower and are covering the same values of x. This hypothetical situation is illustrated in Fig. 7. This is a

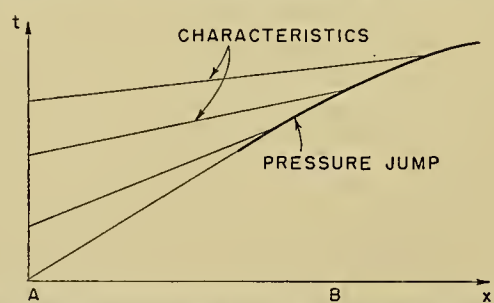


Fig. 6.—The characteristics of a compression wave in an inversion. The pressure jump is represented as a line in the x, t plane.

very unstable situation from a mechanical standpoint. The density ρ is greater than the density ρ' so that the overhanging air with density ρ will fall down into the rest of the fluid. This will usually occur before such an advanced unstable state as the one illustrated in Fig. 7 will occur. The zone in which this air is falling down is usually a rather narrow zone of the flow and is marked by chaotic motion of the air. Such a zone is illustrated in the cross section for t_5 in Fig. 5, and the x, t diagram is shown in Fig. 6. The heavy line in Fig. 6 is called a pressure jump. It is assumed to be a very narrow region in the x, t plane in which the transition from a low inversion height with still air (in this case) to a high inversion height and moving air is accomplished. The

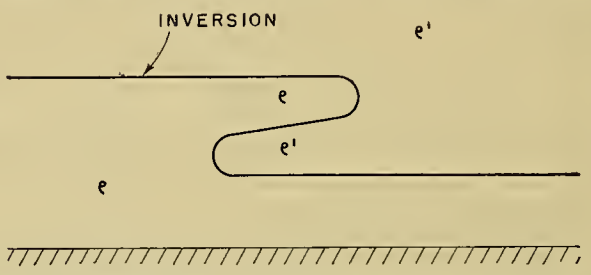


Fig. 7.—The shape an inversion would have after an envelope of the characteristics had formed if the pressure jump did not occur first.

best indication that this is a narrow region is given by synoptic data which will be discussed later. The fact that the hydraulic jump is confined to a narrow zone in the flow in a channel is an indication that this intimately related phenomenon will also be narrow. A barograph placed at station A (Fig. 5) will have a trace similar to trace A in Fig. 8, but an observer at station B will see a trace with a marked jump in the pressure (see trace B in Fig. 8). This effect of the jump in the inversion

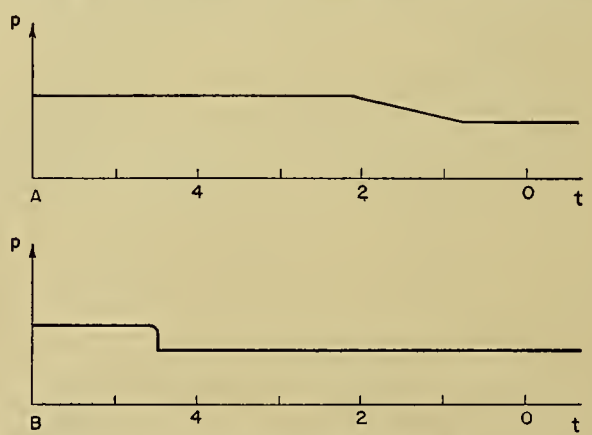


Fig. 8.—Barograph traces recorded during the history of a compression wave.

height on the pressure record at the station led Tepper [19] to adopt the name "pressure jump" for this phenomenon. The pressure surge (after it breaks) of Abdullah [1], the squall line (not the pressure pulse) of Brunk [4], and the jump of Freeman [7] are all pressure jumps. If a flow under an inversion is essentially one-dimensional and if for any reason there is acceleration of the air into a region covered by the inversion at a constant height, a pressure jump will form (if the Coriolis force is neglected). This is just a general description in words of the phenomenon illustrated in Fig. 5. The jump forms because if there is a region of constant inversion height, the flow is a simple wave. If the air is accelerating into this region, it is a compression wave and the straight characteristics converge to form the pressure jump.

Weather Associated with a Pressure Jump. Tepper [19] has described the surface weather phenomena that accompany the passage of a pressure jump. He refers to a larger-scale pressure jump than the one described between the mountains. The origin of the jump will be discussed in the next section, but the scale is one of a change from an inversion height of 2000–3000 ft to one of 5000–10,000 ft with temperature differences across the inversion of 1–5C. The jump in height is moving faster than the conditionally unstable air above. If this air were dry, the phenomena associated with precipitation would not occur. The data over a network of 55 stations in a rectangle 8 mi by 20 mi were averaged. Several outstanding phenomena accompanied the pressure jump of May 16–17, 1948: the wind shift, the temperature break, the rain gush, the wind-speed maximum, and the maximum pressure difference.[2] These terms are almost self-explanatory and a complete description may be obtained from Tepper's original paper. The magnitudes of some of these quantities are given below.

Average total rise in pressure	2.3 mb
Average maximum precipitation intensity	0.02 in. min^{-1}
Average maximum surface wind speed	27 mph
Average maximum rate of temperature fall	1C min^{-1}
Average total wind shift	75°

The foregoing tabulation gives the conditions averaged over all 55 stations, but the pressure jumps are better known and most respected for the more extreme weather phenomena. These are described in some detail by D. T. Williams [23] and are summarized here.

Pressure: The pressure would rise 2–6 mb during time intervals of the order of 5 min. Gradients were as steep as 1–2 mb mi^{-1}. Following the abrupt rise in pressure there was a leveling off or a slow fall.

Wind: Winds ahead of the line were usually southerly and an almost instantaneous shift to westerly or northwesterly accompanied its passage. The peak gust over the whole network was of the order of 50 mph. In many cases the winds blew at right angles to the isobars.

Precipitation: Rain in the first mile behind the pressure jump was usually light, becoming heavy farther back.

Temperature: Temperatures fell sharply upon the passage of a pressure jump. Maximum drops occurred while heavy rain was falling and usually the decrease was proportional to the rate of rainfall. With the ending of the period of heavy rain, temperatures rose slowly, frequently returning to near their original levels.

Such phenomena are naturally of interest to meteorologists who have expressed their interest throughout the years by marking such pressure jumps and their accompanying precipitation as prefrontal squall lines on weather maps. Harrison and Orendorff [11] describe the synoptic aspects of the squall line in great detail. They show that it is certainly not the result of an upper front. However, the importance of the barograph trace and the pressure pattern in locating squall lines has not been fully recognized. The careful and accurate analysis of the Daily Weather Map of the U. S. Weather Bureau since 1946 has made it increasingly evident that the squall line is a real phenomenon. In fact the squall lines and fronts on these maps move in a manner so similar to that of a shock wave and piston in theoretical discussions of gas flow in a tube that a connection only waited for someone who had sufficient knowledge of both motions. This connection was established by Tepper [19] in his detailed synoptic study of the squall line of May 16–17, 1949. He established that this particular squall line was a pressure jump and proposed that most persistent, well-developed, squall lines are pressure jumps.

The Prefrontal Squall Line. Tepper [19] has proposed that the prefrontal squall line is a line of showers and thunderstorms that is started by a pressure-jump line. A study of Fig. 5 shows that if the pressure jump represents a large jump in the height of the inversion, it will have the same effect as a rapidly moving cold front on the air above the inversion. In fact, since the slope of the advancing surface is usually much greater in a pressure jump, the release of whatever moisture is in the upper air should be more rapid and therefore more spectacular. It is believed that the lifting of the air above the inversion by the advancing surface of the inversion causes most of the precipitation in squall lines. (It is also possible that the sudden lifting of the air under the inversion will cause saturation and that the resultant change in the lapse rate will allow the air to break through the inversion and thus the origin of the storms will be in the lower air mass. This possibility is disregarded in this discussion because such a break-through would violate the theoretical model for which equations (14) and (15) are true. Of course, this does not mean that it cannot or does not occur.) Tepper [19] emphasizes that the storms that make up the squall line are caused by the mechanical lifting along a pressure jump and that the jump can occur with no resulting storms or precipitation of any type if the air above the inversion is dry enough. Tepper's study of prefrontal pressure jumps developed from an attempt to discover some theoretical explanation for the intense pressure gradients that occurred during the seven squall lines that were studied by Williams [23]. Williams' work was a synoptic study of the dense network of stations that the Cloud Physics Project maintained near Wilmington, Ohio. This study made it clear that there was a very narrow region in which the pressure changed very rapidly. This region also moved rapidly and was associated with the thunderstorms that made up the squall lines which he was studying. An idea of the width of the zone of rising pressure can be obtained from the accompanying map (Fig. 9). Tepper established that the surface phenomena discussed by Williams occurred in a sequence and moved in such a way that a jump in the height of an inversion (or a stable layer) was the most likely cause of the phenomena. He proposed a model of a squall line based on the work of Freeman

2. H. R. Byers maintains that these properties (the pressure trace in particular) are common in varying degrees to any line of thunderstorms. (Statement from the floor of the 108th meeting of the American Meteorological Society, at Tallahassee, Florida, December 1950.)

[7]. Naturally the line in the x, t plane along which data are known need not be a vertical line corresponding to a fixed value of x. As in the more general case of the previous section, it can be any line in the x, t plane

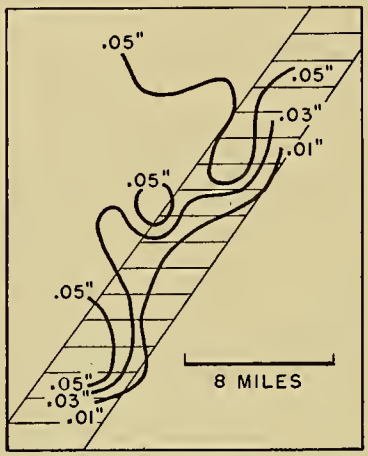


FIG. 9.—A "squall line" magnified in the same ratio as the scale of miles on a micromap showing the position of a pressure jump.

which is not a characteristic line. Accordingly, Tepper assumes that he knows the motion of a steep cold front on the weather map. This cold front is adjacent to a region in which there is an inversion and the air under the inversion is in some equilibrium state with equal inversion height and equal west-wind components throughout. (The north-south wind components and the Coriolis parameter are neglected entirely in this particular study.) The cold front then begins to accelerate and this pushes the inversion up in its vicinity. This compression wave ultimately develops into a pressure jump. In the meantime the front decelerates to its previous speed, or a lower speed, to the west and an expansion wave moves out behind the pressure jump. Some successive cross sections describing this phenomenon appear in Fig. 10.

Interaction of an Expansion Wave and a Pressure Jump. The whole area in which the characteristic lines in Fig. 4 are diverging is called an expansion wave. Such a wave has two speeds: it approaches from the left with a speed given by the slope of the first diverging characteristic and leaves to the right with a speed given by the slope of the last diverging characteristic. The approximate speed of a pressure jump can be computed by a method demonstrated to the writer by C.-G. Rossby. If the jump is assumed to have an unchanging shape and to be moving with constant speed V, the law that the difference in pressure force is equal to the loss in momentum can be expressed as

$$(u_1 - V)\rho u_1 h_1 - (u_2 - V)\rho u_2 h_2 = \int_0^{h_2} P_2\,dz - \int_0^{h_1} P_1\,dz. \tag{21}$$

Equation (21) reduces to

$$\rho(u_1 - V)u_1 h_1 - \rho(u_2 - V)u_2 h_2 = \frac{g}{2}(\rho - \rho')(h_2^2 - h_1^2). \tag{22}$$

The continuity of mass is expressed by the similar equation:

$$(u_1 - V)h_1 = (u_2 - V)h_2\,. \tag{23}$$

If the value of $(u_2 - V)$ from equation (23) is substituted in equation (22), the resulting equation can be solved for the following four possible values of V:

$$V_{1\pm} = u_1 \pm \left[g\left(1 - \frac{\rho'}{\rho}\right)\frac{h_2}{h_1}\frac{h_2 + h_1}{2}\right]^{1/2}, \tag{24}$$

$$V_{2\pm} = u_2 \pm \left[g\left(1 - \frac{\rho'}{\rho}\right)\frac{h_1}{h_2}\frac{h_2 + h_1}{2}\right]^{1/2}, \tag{25}$$

where $V_{1+} = V_{2+}$ and $V_{1-} = V_{2-}$.

If the jump is from a low value of h_1 to a higher value of h_2, it can be seen immediately that if we set

$$V = u + a \tag{26}$$

to correspond to the slope of the characteristics

$$\frac{dx}{dt} = u + c, \tag{27}$$

where

$$c = \left[gh\left(1 - \frac{\rho'}{\rho}\right)\right]^{1/2},$$

that

$$a_1 > c_1 \tag{28}$$

and

$$a_2 < c_2\,. \tag{29}$$

This is a quantitative statement of the following rule expressed for gas dynamics by Courant and Friedrichs [5]: Each member of a sequence made up of alternate expansion waves and jumps in the same family of characteristics will move to overtake the member preceding it. This is true because the expansion waves move at the speed $u + c$ and the jumps move at the

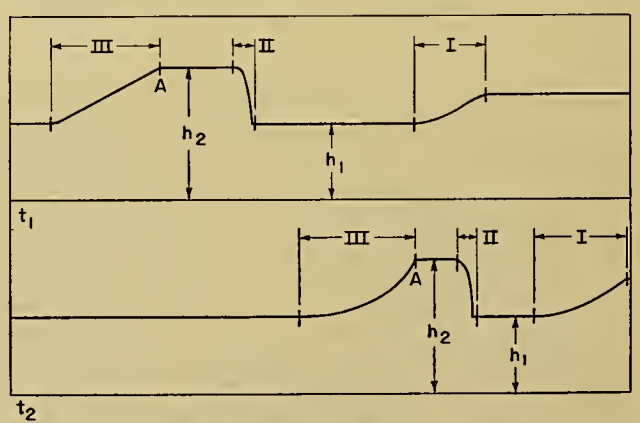


FIG. 10.—A series of simple waves and pressure jumps showing how they overtake and modify each other.

speed $u + a$. This rule is illustrated in Fig. 10 where it can be seen that the left-hand side of expansion wave I moving at $u_1 + c_1$ is being overtaken by jump II

moving at $u_1 + a_1$ because $a_1 > c_1$ and that jump II, in its turn moving at $u_2 + a_2$, is being overtaken by the right-hand side of expansion wave III moving at $u_2 + c_2$ because $a_2 < c_2$. This phenomenon was first demonstrated synoptically by Tepper [19], who showed that on May 16, 1948 the pressure maximum which coincides with point A in Fig. 10 had an average speed of 54 mph over the network stations at Wilmington, Ohio, while the jump had a speed of 45 mph. This phenomenon of overtaking leads to a modification of the expansion waves and jumps involved. The exact nature of this modification in all cases and the results of all "interactions" as they are called are not known. In some simple cases an approximation to the actual conditions can be made that has some validity. Fortunately, one of these simple cases is important in the study of squall lines. One manner of dissipation of a pressure jump is for it to be modified by a following expansion wave (Fig. 10).

Map Analysis with a Pressure Jump. The small amount of space and time occupied by a pressure jump has been discussed previously. The small dimensions in space of the pressure jump are emphasized in Fig. 9, where the line marking a squall line from the Daily Weather Map of the U. S. Weather Bureau is magnified in the same proportion as the scale of miles. These lines

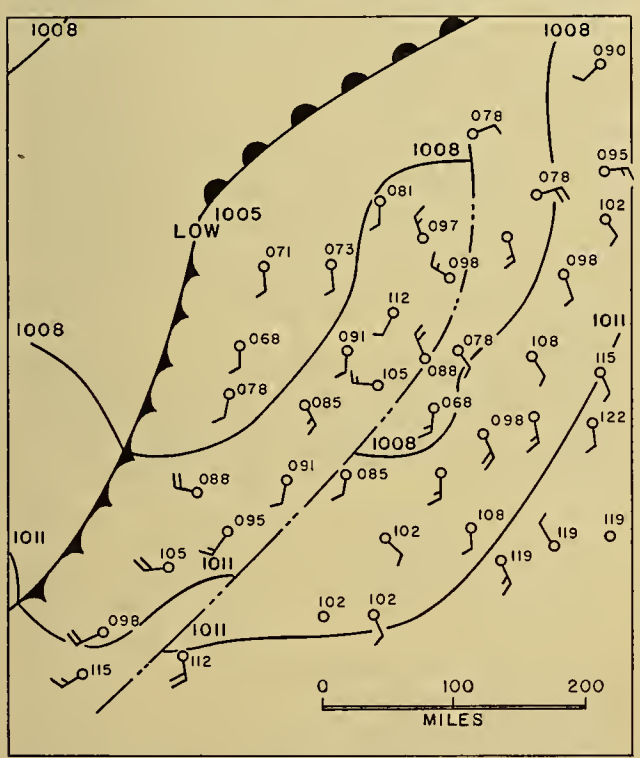


FIG. 11.—Proposed method of drawing isobars in the vicinity of a pressure jump. The map is for 2230 EST May 16, 1948 and the squall line is through northwestern Pennsylvania and southeastern Ohio.

are four miles wide and it can be seen that this one covers most of the zone of rising pressure. With this chart in mind it has been proposed [9] that the same approximation used in the mathematical analysis be made, and *that the pressure jump be analyzed as a discontinuity in the field of pressure.* This method of analysis would have the following advantages: (1) more of the winds would blow nearly parallel to the isobars, (2) the analysis would fit more of the pressures than is usually possible with conventional methods (as can be seen from Fig. 11), and (3) the analysis would correspond more closely to the proposed mathematical theory of pressure jumps. This type of analysis has the disadvantage that the isobaric field is not correct in the immediate vicinity of the pressure jump, but Fig. 9 shows that the area that is not accurately represented is almost covered by the line marking the pressure jump. Pressure jumps can be located rather accurately on a well-plotted synoptic map by this method. An example of such an analysis is given in Fig. 11. Pressure jumps could be located very accurately indeed if the suggestion were followed of sending a special report on the hourly network when a pressure jump passed the station.

Transmission of Energy by Means of Finite Disturbances in Inversion Height. Abdullah [1], working independently, established the existence of a pressure jump on the inversion behind a cold front. He advanced the hypothesis that the pressure jump in cool air was the medium that carried the energy of distant cold air to the vicinity of a cyclone and that this energy was then available to aid the formation or deepening of a cyclone on the front separating the cool air from the warm air. Abdullah also suggested that pressure jumps of finite lateral extent could be one of the "disturbances" needed to start the formation of a wave on a front in the frontal theory of cyclones. By means of an energy computation that can be found in his paper, he showed that for the usual dimensions of waves on a front a pressure jump of ordinary dimensions will supply more than ten times the amount of energy that is found in the frontal wave. Abdullah used the Riemann method of integration. The method is particularly well adapted for the study of a situation in which the boundary conditions appear in closed form. Assuming that the coldest air begins moving at $t = 0, x = 0$ with a constant acceleration α, he showed that a compression wave moving into air with a constant height h_0 of the inversion will break into a pressure jump (*i.e.*, the envelope of characteristics will form) at a point

$$T_0 = \frac{2}{3\alpha}\sqrt{gh_0\left(1 - \frac{\rho'}{\rho}\right)}, \tag{30}$$

$$x_0 = \frac{2}{3\alpha}gh_0\left(1 - \frac{\rho'}{\rho}\right). \tag{31}$$

Using these formulas, he constructed a table showing computed and observed distances (x) from the front at which breaking into a pressure jump occurred. The essential information from his table is presented in Table I in which T is the mean temperature of the cool air, and ΔT is the difference in temperature between the cool and the warm air. The best synoptic example of a pressure jump on a large scale to appear in the

literature is provided by Abdullah who showed how the energies of motion and the rise in inversion height are transmitted to the leading edge of the cool air.

TABLE I. DISTANCE OF PRESSURE JUMP FROM FRONT

Initial time of 12-hr period	h (m)	T (°K)	ΔT (°C)	x computed (mi)	x observed (mi)
0400Z Jan. 19, 1947	1700	253	5.0	490	410
1600Z Mar. 23, 1947	2000	259	3.5	670	550
1600Z Mar. 4, 1947	1800	274	4.0	436	390

Time-Dependent Flow as a Forecast Tool. The usefulness of the method of characteristics in making forecasts has been emphasized before [7, 21]. The success of Abdullah [1] in predicting the time of breaking and of Freeman [7] in predicting the height of the tropical inversion emphasizes this feature. Of course, the solution of any equation with time as an independent variable gives a method of prediction. In most sciences the problem is to predict the behavior of a system with a given set of initial conditions which will be applied repeatedly. In weather forecasting the prediction is usually made from a set of initial conditions that is never repeated. The method of characteristics is particularly suited for such a system because any set of initial conditions can be fitted to any accuracy desired, and regularity of the functions involved is not required.

In any region in which the flow under an inversion can be considered one-dimensional and the effects of the Coriolis force are not important there are two procedures for making forecasts that might be found useful. The most straightforward procedure is that of forecasting the weather along part of the x-axis from a given set of initial conditions on a larger part of that axis. The x-axis is a line in the x, t plane and the characteristics can be built up from it as in Fig. 12. Such a

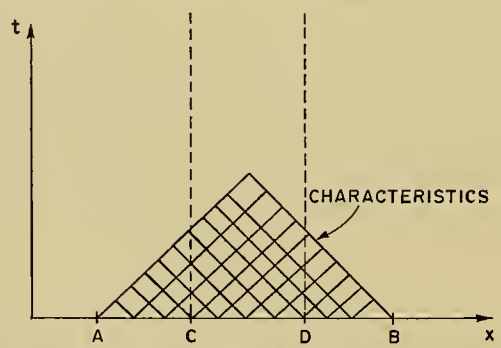


FIG. 12.—Area for which a prediction can be made from initial data on the x-axis.

method should be useful over large land areas in the tropics, such as Africa and South America, and probably in island-studded oceans like the northern part of the South Pacific Ocean. This method could also be developed into a method of making forecasts from data gathered from aircraft flights, particularly those which make periodic observations of the inversion height. The most useful method between stations (in the tropics particularly) is outlined in Fig. 13. The procedure above, or a good estimate of the conditions between A and B, is used initially and then new characteristics are found from information at A and B only. This pro-

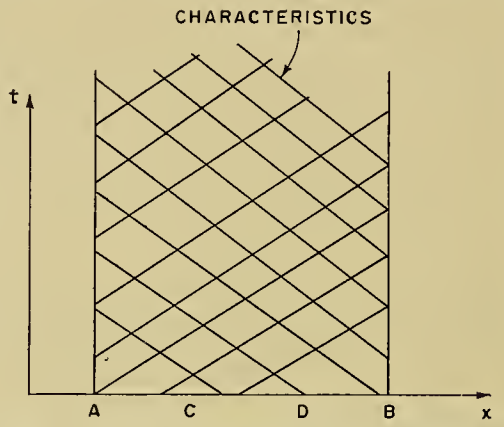


FIG. 13.—Area for which continuing predictions can be made from a series of weather observations (data along the t-axis) at two discrete points.

cedure could be used to great advantage to predict conditions between two islands in the easterlies. Of course, to be used to best advantage, A and B should be a reasonable distance to the east and west of the region for which forecasts are desired. Forecasts for C and D can be made for appreciable periods. The usefulness of this method as a forecast tool for anything other than the qualitative aspects of flow under an inversion in the middle latitudes awaits the incorporation of the effects of the Coriolis force into the method. High-speed computing machines will undoubtedly be needed to forecast two-dimensional time-dependent flow under an inversion. The extension to two dimensions will eliminate many of the restrictive assumptions involved in making a one-dimensional problem out of one that has such inherent two-dimensional aspects as the Coriolis force.

Equations for Steady-State Flow under an Inversion. If we assume that there is steady-state flow under an inversion, that is, that there are no changes in the flow with time, then equations (11)–(13) become

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\left(1 - \frac{\rho'}{\rho}\right)\frac{\partial h}{\partial x}, \tag{32}$$

$$u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\left(1 - \frac{\rho'}{\rho}\right)\frac{\partial h}{\partial y}, \tag{33}$$

$$u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} + h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \tag{34}$$

This is a system of equations similar to equations (1) and (2). It is complicated slightly by the presence of a third unknown h, but the same method of analysis is used on this set of equations. Three families of characteristics exist and the determinants used in equation (6) are of order six rather than four. In this case this third characteristic is a streamline and the vorticity $\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ is constant along it. If we assume the flow is irrotational, this characteristic is not used. The flow

can then be studied in the x, y plane like equations (1) and (2). This equation is complicated by the fact that real characteristics exist only under certain conditions, that is, if

$$u^2 + v^2 > gh\left(1 - \frac{\rho'}{\rho}\right). \tag{35}$$

The study of steady-state flows has not been applied in synoptic meteorology to date. The only published discussions of a steady two-dimensional flow are the climatological example in the next paragraph and the conjecture of Tepper discussed below [19].

An Example of a Two-Dimensional Steady Flow. An example of the type of flow described in the previous section has been given by Freeman [9]. The chart from the *Climatic Charts of the Oceans* [22] showing mean resultant winds off the west coast of South America in October is reproduced in Fig. 14. This is to be compared

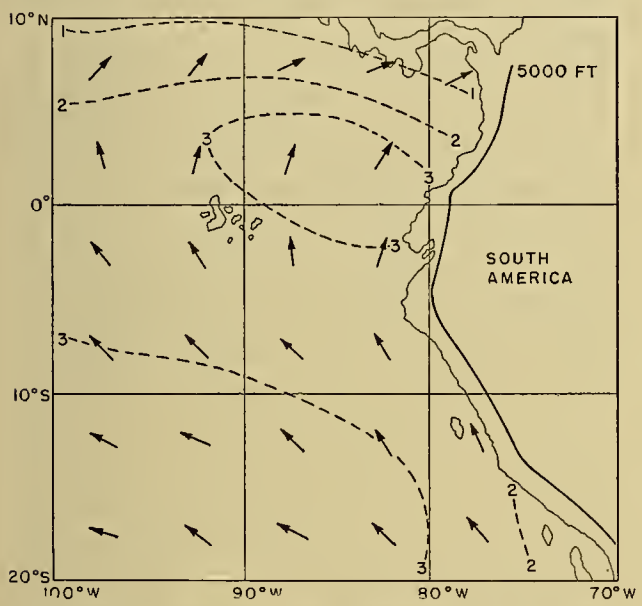


Fig. 14.—The resultant wind field east of South America in October from the *Climatic Charts of the Oceans*. The dashed lines are isopleths of Beaufort force.

to the Prandtl-Meyer type of expansion wave in Fig. 15. This expansion wave is computed from the following initial data: Winds of 5 m sec^{-1} (force 3) under an inversion of 1C at 0.65 km are blowing parallel to the coast of South America and pass the bend in the coast line. The wind speed is assumed to be somewhat higher than the observed surface wind to allow for slowing by surface friction. The height and strength of the inversion is that found by Alpert [2]. Points of similarity of these flows are that the change in wind direction occurs farther north as the longitude increases and that the winds increase their speed as they turn around the corner. The slowing down of the resultant winds north of 5°N is to be expected because the expansion wave does not always extend to that latitude and when it does it is more likely to be modified by the Northern Hemisphere wind systems. This may not be the only factor causing these winds, but it must certainly be an important one, since the data fit the theoretical model so well.

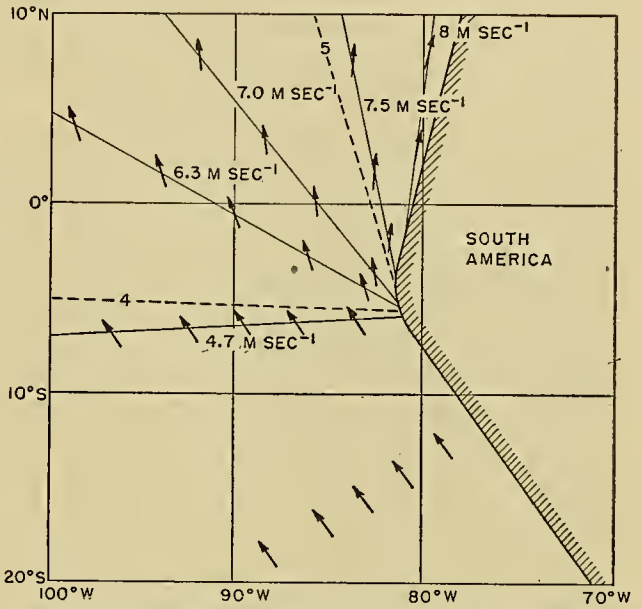


Fig. 15.—The computed wind field if the air under an inversion undergoes a Prandtl-Meyer expansion around the bend in the west coast of South America.

Interaction of Two Pressure-Jump Lines. It has been proposed by Tepper [20] that the intersection of two pressure-jump lines should be a region of preferred occurrence of tornadoes. He cited the discussion of interactions by Courant and Friedrichs [5] to lend support to this conjecture. When two shock waves intersect in the two-dimensional flow of a compressible fluid a vortex sheet is formed. This vortex sheet has been found experimentally in gas flows and in flow of water in a channel. Such a vertical vortex sheet should form behind the intersection of two pressure jumps. If the circular properties of a tornado are similar to those of the dust devils studied by N. R. Williams [24], then such a region of strong shear is a very likely zone for tornado formation.

Tepper [20] found that, in five out of the seven tornado situations he studied, the tornadoes formed within fifty miles of the intersection of two pressure jumps.

CURRENT PROBLEMS AND SUGGESTIONS FOR RESEARCH

The application of the method of characteristics to meteorological problems is an active field of research with many topics of interest. Most of the problems under consideration at the present time are concerned with flow under an inversion. The path to follow in the solution of many of these problems is clearly marked. Other problems are being attacked with success by Tepper. The success of Abdullah [1] in explaining the "secondary cold front" or "postfrontal squall line" and of Tepper in explaining the occurrence of the "squall line" indicates that many phenomena on the same and smaller scales can be studied with this theory in mind.

The usefulness of the method of characteristics is not confined to the study of flow under an inversion. Any problem in which certain elements are described by equations such as (1) and (2), with real characteristics, can be studied by these methods. It is hoped that this article will be useful to every meteorologist with such problems and that some may even be encouraged to study some of the problems mentioned below.

The Effect of the Coriolis Force on Flow under an Inversion. In all of the discussions of nonlinear flow under an inversion that have appeared in the literature the Coriolis parameter is neglected. The Coriolis force works to turn the winds parallel to the isobars, but it takes time for this force to act. If a parcel of air is under the influence of a pressure gradient for only a short time, it will follow the path dictated by the equations without Coriolis force so there is some justification in neglecting this parameter in the pioneering stages of such a study. If the value of the theory of flow under an inversion to middle-latitude forecasting is to increase, however, the Coriolis parameter must be included in the equations. Some of the physical effects of this force can be seen immediately. Suppose that a cold front is holding an inversion of constant height at rest at point x_0 and time t_0. At t_0 the front begins to recede, moving westward from the inversion. An expansion wave will be formed and the slope in the inversion height will be such that pressure will increase to the east. If the front continues to recede, the air parcels near it and beneath the inversion will be under the influence of such a pressure gradient for a long period of time and will therefore move northward, giving south winds in the now lower-pressure zone to the east of the front. If, on the other hand, the front pushes slowly into the inversion, the new pressure gradients will cause north winds. The equations that tell how much south or north wind is to be expected and how long it will blow are being studied at the present time.

The Characteristics of a Circular Vortex and the "Rings" of a Hurricane. The rings or spirals of violent convective activity in hurricanes have been observed by many authors [13, 25]. We have seen in earlier sections that the Mach lines and the envelopes of Mach lines are at least approximately the lines along which physical disturbances move in a flow. The Mach lines of a circular vortex in a fluid with a free surface are spirals almost tangent to the streamlines near the center of the vortex, making a larger angle with them as the distance from the center increases. If the presence of a stability retarding the vertical motion of air involved in a hurricane can be established, the theory of flow under an inversion can be applied and might lead to an explanation of these spirals. To the writer's knowledge no one is working on this problem at the present time.

Interactions and Frontogenesis. The discussion at the end of the previous section indicates that a vortex sheet (or wind shear line) is formed when two strong pressure jumps are involved in an interaction. If these vortex sheets can bring about frontogenesis, and this occurrence is frequent enough to be important to synoptic meteorologists, this should be a fruitful field of investigation. From synoptic experience we can say that if the vortex sheet persists it is very likely to bring about frontogenesis. A wind-shear line usually develops into a front because there usually are horizontal temperature gradients in an air mass. This problem is not being studied at the time of this writing.

The "Expansion-Wave Storm." Brunk [4] has made a complete and very interesting study of a wind storm in 1944. The winds over the northern part of the United States were from the east under a quasi-stationary front. A small wave moved from west to east along this front (rather rapidly) and, as it moved, the winds to the north of it became very strong from the east. This storm is now being studied as an expansion wave moving from the west followed by a compression wave and a subsequent pressure jump on the inversion north of the quasi-stationary front. The effect of the Coriolis force is being included in this investigation. The problems of flow under an inversion are very well suited for complementary work by theoreticians and synoptic meteorologists.

Blocking Waves in a Planetary Jet Stream. Rossby [17] has shown that an analogue to the stationary hydraulic jump can occur in a planetary jet stream. The jet stream he uses is defined as a symmetric stream of fast-flowing air moving through stationary surroundings. He showed that if the variation of the Coriolis parameter is considered and the momentum transport for a given volume transport is expressed as a function of velocity, the momentum transport has a minimum value. This means that for any other value of the momentum transport there are two possible flows with this volume transport. Thus, a necessary condition for the sudden change from one state of flow to another is satisfied. This is exactly what happens in a hydraulic jump. Rossby advances the theory that this analogue to the hydraulic jump appears on the weather map as a blocking wave of the type discussed by Berggren, Bolin, and Rossby [3]. Rossby's paper naturally suggested that a system of equations of a nonstationary flow of this type might be developed and that possibly they would be similar to equations (1) and (2). If the velocity profile of the jet stream is such that the vorticity in the jet is f_0 (*i.e.*, $u = u_0 + \beta y^2/2$), the jet stream has width a, the acceleration in the y direction can be neglected, and the geostrophic approximation can be made outside the jet stream, then the equations describing the flow of the stream are

$$\frac{\partial u_0}{\partial t} + \left(u_0 - \frac{\beta a^2}{2}\right)\frac{\partial u_0}{\partial x} = 0,$$

$$\frac{\partial a}{\partial t} + \left(u_0 + \frac{\beta a^2}{2}\right)\frac{\partial a}{\partial x} + a\frac{\partial u_0}{\partial x} = 0.$$

These are equations of the type (1) and (2). Simple waves are possible, and the compression waves lead to envelopes of the characteristics. Needless to say, this problem is being studied actively and in detail. The success of the method of characteristics in the study of this problem emphasizes the fact that the usefulness of this method to meteorologists is not confined to the

study of flow under an inversion. It is sincerely hoped that all meteorologists will keep this method in mind as a possible means of obtaining quantitative results in a problem.

REFERENCES

1. Abdullah, A. J., "Cyclogenesis by a Purely Mechanical Process." *J. Meteor.*, 6:86–97 (1949).
2. Alpert, L., "Notes on the Weather and Climate of Seymour Island, Galapagos Archipelago." *Bull. Amer. meteor. Soc.*, 27:200–209 (1946).
3. Berggren, R., Bolin, B., and Rossby, C.-G., "An Aerological Study of Zonal Motion, Its Perturbations and Break-Down." *Tellus*, Vol. 1, No. 2, pp. 14–37 (1949).
4. Brunk, I. W., "The Pressure Pulsation of 11 April 1944." *J. Meteor.*, 6:181–187 (1949).
5. Courant, R., and Friedrichs, K. O., *Supersonic Flow and Shock Waves*. New York, Interscience, 1948.
6. Crocker, A. M., Godson, W. L., and Penner, C. M., "Frontal Contour Charts." *J. Meteor.*, 4:95–99 (1947).
7. Freeman, J. C., "An Analogy between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5:138–146 (1948).
8. —— "Reply: The Usefulness of Incompressible Models." *J. Meteor.*, 6:287 (1949).
9. —— "Map Analysis in the Vicinity of a Pressure Jump." *Bull. Amer. meteor. Soc.*, 31:324–325 (1950).
10. —— "The Wind Field of the Equatorial East Pacific as a Prandtl-Meyer Expansion." *Bull. Amer. meteor. Soc.*, 31:303–304 (1950).
11. Harrison, H. T., and Orendorff, W. K., "Pre-cold-frontal Squall Lines." *United Air Lines Meteor. Dept. Circ.* No. 16 (1941).
12. Isenberg, J. S., *The Method of Characteristics in Compressible Flow, Part I*. Tech. Rep. No. F-TR-1173A-ND, Hdqtrs. Air Materiel Command, Wright Field, Dayton, Ohio, 1947.
13. Maynard, R. H., "Radar and Weather." *J. Meteor.*, 2:214–226 (1945).
14. McGurrin, M., *The Principle of Open Channel Flow Applied to the Formation of Tornadoes*. Unpublished, available at the Library of the U. S. Weather Bureau, Washington, D. C., 1942.
15. Richardson, L. F., *Weather Prediction by Numerical Process*. Cambridge, University Press, 1922.
16. Rossby, C.-G., "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2:187–204 (1945).
17. —— "On the Dynamics of Certain Types of Blocking Waves." *J. Chinese geophys. Soc.*, 2:1–13 (1950).
18. Sherman, L., "The Neglect of Compressibility in the Flow of Gas at Low Speeds." *J. Meteor.*, 6:286–287 (1949).
19. Tepper, M., "A Proposed Mechanism of Squall Lines—The Pressure Jump Line." *J. Meteor.*, 7:21–29 (1950).
20. —— "On the Origin of Tornadoes." *Bull. Amer. meteor. Soc.*, 31:311–314 (1950).
21. —— and Freeman, J. C., "Analogy between Equatorial [Easterlies] and Supersonic Flows." *J. Meteor.*, 6:226 (1949).
22. U. S. Weather Bureau, *Atlas of Climatic Charts of the Oceans*. W. B. Publ. No. 1247, Washington, D. C., 1938.
23. Williams, D. T., "A Surface Micro-Study of Squall-Line Thunderstorms." *Mon. Wea. Rev. Wash.*, 76:239–246 (1948).
24. Williams, N. R., "Development of Dust Whirls and Similar Small-Scale Vortices." *Bull. Amer. meteor. Soc.*, 29:106–117 (1948).
25. Wexler, H., "Structure of Hurricanes as Determined by Radar." *Ann. N. Y. Acad. Sci.*, 48:821–844 (1947).

HYDRODYNAMIC INSTABILITY*

By JACQUES M. VAN MIEGHEM

University of Brussels

Introduction

The concept of dynamic instability appeared along with the initial developments in the theory of atmospheric disturbances [2]. In the method of perturbations, which V. Bjerknes introduced into meteorology, a "small motion" is superimposed on a state of "simple motion" of the air. This "simple motion" is always a permanent motion characterized, at every point of the medium, by equilibrium of the forces perpendicular to the direction of the flow. These forces are gravity, the pressure gradient (hydrostatic equilibrium), the Coriolis force (state of geostrophic motion), and the centrifugal force (circular vortex). The "simple motion," therefore, is evidently a state of hydrodynamic equilibrium.

The "small motion" is simply a nascent perturbation, defined by quantities small enough that the values of the products and squares of the quantities as well as of their derivatives are negligible. Furthermore, it is assumed to begin with that the perturbation is restricted to a train of plane sinusoidal waves, propagated with constant velocity c in the same sense as the simple motion. This perturbation must satisfy linearized equations derived from Euler's dynamical equations and from the equations of continuity and of the adiabatic process; from these the amplitude relations are obtained. In addition it is necessary to satisfy restrictions imposed by external boundaries (surface of the globe, free surface) and, eventually, by internal boundaries (frontal surface, tropopause). There generally results a relation between the quantities describing the physical and dynamical state of the "simple motion," the wave length L, and the velocity of propagation c. When this "dispersion equation" admits only real roots for c, the perturbation is said to be stable regardless of the value of the wave length L; on the other hand, when the equation admits only imaginary roots for c, the motion is referred to as unstable. In general this equation admits real roots, for certain values of L and imaginary roots for others, that is, the stability or instability of a perturbation depends on its wave length. The condition for which the dispersion equation has only imaginary roots for c expresses the dynamic instability criterion in the sense of V. Bjerknes and H. Solberg [2]. This criterion, which depends not only on the state of simple motion, but also, and more significantly, on the wave length L, is thus a *selective* criterion that can be applied only to a flow pattern consisting of a perturbation of the type "plane sinusoidal waves" propagated uniformly in the direction of the simple motion.

This dynamic instability of selective character has been studied extensively since the first work of V. Bjerknes and H. Solberg, notably by B. Haurwitz, C. L. Godske, Z. Sekera, J. G. Charney, P. Queney, E. T. Eady, and others. On the other hand, the stability of the simple motion, upon which the nascent perturbation is superimposed, was not considered until very much later, although logically it should have been studied in the very beginning. It is evident, without further amplification, that the stability or instability constitutes an essential characteristic of the simple motion and must appear as such in the dispersion equation.

Kleinschmidt [17, 18] was the first to show that, along with the hydrostatic instability of masses at rest in the gravity field, there exists a hydrodynamic instability of air masses whose motion obeys the law of geostrophic flow. Although the hydrostatic equation amounts to an excellent approximation to the equation of atmospheric dynamics in a vertical direction, the classical criteria of stability and instability for the vertical distribution of the air are strictly applicable only in an atmosphere without wind, in which there are only vertical displacements and where only vertical forces, such as gravity and the vertical pressure gradient, come into play. In reality, on a synoptic scale, we must take account of the horizontal pressure gradient, the Coriolis force, and the centrifugal force, as well as of the vertical forces. In this case the criteria of stability and instability of hydrostatic equilibrium lose their validity, and it is expedient to substitute for them criteria of stability and instability of hydrodynamic equilibrium.

In general, the atmosphere is said to be stable (or unstable) at any point O at any instant t for any direction r when a particle displaced from this point, at this instant and in the direction r, acquires a relative motion with respect to the environment at O returning it toward (or carrying it away from) the equilibrium position O.

The Stability of Geostrophic Motion

Let us consider a mass of air in a state of geostrophic motion (hydrodynamic equilibrium in the sense of Kleinschmidt [17, 18]). In this case the isobaric surfaces P = const and the isentropic surfaces Θ = const of the air are cylindrical surfaces with horizontal generatrices parallel to the geostrophic current; the equipotential surfaces, Φ = const, of the gravity field are assumed to be planes parallel to the horizontal plane at any given point O in the interior of the air mass. We refer the motion of the air to a right-handed, rectangular, Cartesian coordinate system $Oxyz$, at rest with respect

* Translated from the original French.

to the earth and having the axis Ox directed parallel to the geostrophic current $\mathbf{u}$ and the axis Oz directed toward the zenith above point O. In this system we thus have $u_x \equiv u(y, x) > 0$, $u_y \equiv u_z \equiv 0$, $P \equiv P(y, z)$, $\Theta \equiv \Theta(y, z)$, $\Phi \equiv \Phi(z)$, and the components of the earth's rotation ω are given by $\omega_x \equiv \omega \cos\phi \sin\alpha$, $\omega_y \equiv \omega \cos\phi \cos\alpha$, and $\omega_z \equiv \omega \sin\phi$, where ϕ is the latitude of point O and α is the angle between the vector $\mathbf{u}$ and the west-to-east direction. Let us now introduce Exner's function $\Pi \equiv c_p(P/100)^{R/c_p}$ (in which P is expressed in centibars), where c_p stands for the specific heat at constant pressure and R for the specific ideal gas constant for air. Finally, to facilitate the presentation, any orientation parallel to the current $\mathbf{u}$ will be called *longitudinal* and any orientation perpendicular to this current will be called *transverse*.

The law of geostrophic motion expresses the equilibrium of the transverse forces $-\nabla\Phi, -\Theta\nabla\Pi$, and $-2\omega \times \mathbf{u}$, that is,

$$2\omega \times \mathbf{u} + \Theta\nabla\Pi + \nabla\Phi = 0. \tag{1}$$

By taking the curl of *equation* (1) *of hydrodynamic equilibrium*, we find the *equilibrium condition of Margules*,

$$2\omega\cdot\nabla u = [\nabla\Theta \times \nabla\Pi]_x \equiv N, \tag{2}$$

which expresses the integrability of equation (1) with respect to Π. In the equilibrium condition, therefore, the gradient of the air speed is proportional to the number N of isobaric-isosteric solenoids per unit of transverse area.

We shall now consider a certain particle, in the interior of the air mass which is in hydrodynamic equilibrium, carried along by the current $\mathbf{u}$ and occupying at any instant t_0 the position of an arbitrarily chosen point O; at the instant t_0 we apply to it a transverse impulse; then at the same instant we take away the impulse. Let $A(x, y, z)$ be the position occupied by the perturbed particle at instant $t > t_0$; let $\mathbf{v}(v_x, v_y, v_z)$ be the velocity of this particle relative to current $\mathbf{u}$ at A at the instant t, and let $\mathbf{V}(V_x, V_y, V_z)$ be the velocity of this same particle relative to the earth at the same instant; we then have

$$V_x \equiv \frac{dx}{dt} = u + v_x, \qquad V_y \equiv \frac{dy}{dt} = v_y, \qquad V_z \equiv \frac{dz}{dt} = v_z, \tag{3}$$

where the operator d/dt stands for the substantial derivative with respect to time t in the reference system $Oxyz$, at rest with respect to the earth.

We now propose to study the motion of the perturbed particle A around its equilibrium position O. This relative motion will be assumed to be a small motion; furthermore, we shall neglect the effects of friction, conduction, and radiation on the displaced particle, and we shall assume that its displacement involves no perturbation of the pressure field. Consequently the perturbed particle retains the potential temperature Θ_0 which it had at point O, and at any instant $t > t_0$ its pressure is equal to that at the point A which it is occupying at that instant. The motion of the perturbed particle $A(x, y, z)$ relative to the earth is then determined by the vector equation

$$\frac{d\mathbf{V}}{dt} + 2\omega \times \mathbf{V} + \Theta_0\nabla\Pi + \nabla\Phi = 0 \tag{4}$$

and by the initial conditions

$$x = y = z = 0, \quad t = t_0, \quad V_x = u_0 \equiv u(0, 0), \quad V_y = v_y^0, \quad V_z = v_z^0, \tag{5}$$

where v_y^0 and v_z^0 are the transverse components of the initial velocity communicated at O. By eliminating $\nabla\Phi$ between (4) and (1), we obtain

$$\frac{d\mathbf{V}}{dt} + 2\omega \times \mathbf{v} = (\Theta - \Theta_0)\nabla\Pi. \tag{6}$$

We observe that if, at the initial instant t_0, we apply the same impulse to all the particles which occupy the points of the Ox axis at this instant, all these particles will perform the same adiabatic, inertial motion (4) or (6). Consequently, the quantities which describe this motion will be independent of the longitudinal coordinate x, like all those which define geostrophic motion. It is thus natural to separate the motion (6) of point A into transverse and longitudinal parts.

In order to obtain the equation of longitudinal motion at A, it suffices to project equation (6) on the axis Ox; if we use (3), we obtain the equation of motion of the projection $A''(x, 0, 0)$ of A on the longitudinal axis through O:

$$dV_x + 2\omega_y\, dz - 2\omega_z\, dy = 0. \tag{7}$$

This equation gives the change dV_x, relative to the earth, which takes place in the longitudinal velocity of the perturbed particle when it undergoes transverse displacement (dy, dz). It is integrable at once; by virtue of (5) and of the relation $u = u_0 + (\partial u/\partial y)_0 y + (\partial u/\partial z)_0 z$, we obtain

$$v_x = \left(2\omega_z - \frac{\partial u}{\partial y}\right)_0 y - \left(2\omega_y + \frac{\partial u}{\partial z}\right)_0 z, \tag{8}$$

where the subscript zero serves as a reminder that the quantities in parentheses must be replaced by their values at point O. We see, therefore, that the longitudinal motion of point A is fixed, as soon as we know its transverse motion. The latter is governed by the equations obtained by projecting relation (6) on the transverse axes Oy and Oz; following (3), the equations of motion of the projection $A'(0, y, z)$ of A on the transverse plane through O are

$$\frac{dv_y}{dt} - 2\omega_x v_z = \psi_y, \quad \frac{dv_z}{dt} + 2\omega_x v_y = \psi_z, \tag{9}$$

where ψ_y and ψ_z are the components of the transverse force

$$\psi \equiv (\Theta - \Theta_0)\nabla\Pi - 2\omega \times \mathbf{v}_x \tag{10}$$

which the surrounding medium exerts on the unit mass of the displaced particle. This transverse force con-

sists of two parts: the first, perpendicular to the isobaric surfaces, reduces to the hydrostatic buoyant force of Archimedes, in the absence of the geostrophic current $\mathbf{u}$; the second, perpendicular to both the geostrophic current and the earth's axis, is an inertia force due to the earth's rotation. We observe that in an adiabatic atmosphere ($\Theta \equiv \Theta_0 \equiv$ const), where differences between the density of the displaced particle and that of its environment disappear, only the second of the two forces remains. If in (10) we replace $\Theta - \Theta_0$ by $(\partial\Theta/\partial y)_0 y + (\partial\Theta/\partial z)_0 z$ and v_x by its value in (8), we find, whatever the orientation of transverse axes y and z, provided always that the coordinate system xyz is rectangular and right-handed,

$$\psi_y \equiv -a^0_{yy}y - a^0_{yz}z, \quad \psi_z \equiv -a^0_{zy}y - a^0_{zz}z, \tag{11}$$

with

$$\left.\begin{aligned} a_{yy} &\equiv 2\omega_z\left(2\omega_z - \frac{\partial u}{\partial y}\right) - \frac{\partial \Pi}{\partial y}\frac{\partial \Theta}{\partial y}, \\ a_{yz} &\equiv -2\omega_z\left(2\omega_y + \frac{\partial u}{\partial z}\right) - \frac{\partial \Pi}{\partial y}\frac{\partial \Theta}{\partial z}, \\ a_{zy} &\equiv -2\omega_y\left(2\omega_z - \frac{\partial u}{\partial y}\right) - \frac{\partial \Pi}{\partial z}\frac{\partial \Theta}{\partial y}, \\ a_{zz} &\equiv 2\omega_y\left(2\omega_y + \frac{\partial u}{\partial z}\right) - \frac{\partial \Pi}{\partial z}\frac{\partial \Theta}{\partial z}, \end{aligned}\right\} \tag{12}$$

and the symmetry relation $a_{yz} \equiv a_{zy}$ resulting from the equilibrium condition (2).

Making use of equations (11) and (12), which give the components of the force Ψ that acts on perturbed particle A as functions of the physical and dynamical state of the atmosphere at the equilibrium position O of A, and of the transverse components of displacement $\overrightarrow{OA}$, it is easy to determine the nature of the hydrodynamic equilibrium at point O. Depending on whether the applied force (ψ_y, ψ_z) and the displacement (y, z) are of opposite or of the same sense, this force tends to return the displaced particle A toward its equilibrium position O or to carry it away from this position, that is, depending on whether $-\psi_y y - \psi_z z > 0$ or <0, the hydrodynamic equilibrium at O is stable or unstable.

We thus rederive E. Kleinschmidt's criterion: *the state of geostrophic motion at any point, for any transverse direction r_y, r_z at this point, is stable or unstable depending on whether* [17, 18, 33, 37]

$$Q(r_y, r_z) \equiv a_{yy}r_y^2 + 2a_{yz}r_y r_z + a_{zz}r_z^2 > \text{or} < 0, \tag{13}$$

or again,

$$Q(r_y, r_z) \equiv 2(\mathbf{r} \times \boldsymbol{\omega})_x(\mathbf{r} \times \operatorname{curl} \mathbf{U})_x - (\mathbf{r}\cdot\nabla\Theta)(\mathbf{r}\cdot\nabla\Pi) > \text{or} < 0, \tag{14}$$

where

$$\left.\begin{aligned} \operatorname{curl}_x \mathbf{U} &\equiv 2\omega_x, \\ \operatorname{curl}_y \mathbf{U} &\equiv 2\omega_y + \frac{\partial u}{\partial z}, \\ \operatorname{curl}_z \mathbf{U} &\equiv 2\omega_z - \frac{\partial u}{\partial y}, \end{aligned}\right\} \tag{15}$$

are the components of the curl of the absolute geostrophic velocity $\mathbf{U} = \mathbf{u} + \boldsymbol{\omega} \times \mathbf{R}$, in which $\mathbf{R}$ is the radius of rotation of the point relative to the earth's axis. As a special case, in an adiabatic atmosphere where no Archimedean buoyant force is possible, there is said to be *inertial stability* or *inertial instability* [32], depending on whether

$$2(\mathbf{r} \times \boldsymbol{\omega})_x(\mathbf{r} \times \operatorname{curl} \mathbf{U})_x > \text{or} < 0. \tag{16}$$

Next we multiply equations (9) respectively by dy/dt and dz/dt. By adding them we obtain, after integrating, the relation thus found and taking account of initial conditions (5),

$$\tfrac{1}{2}[(v_y)^2 + (v_z)^2] - \tfrac{1}{2}[(v_y^0)^2 + (v_z^0)^2] = -\tfrac{1}{2}[a^0_{yy}y^2 + 2a^0_{yz}yz + a^0_{zz}z^2]. \tag{17}$$

Consequently, depending upon whether the hydrodynamic equilibrium at O is stable or unstable for displacement (y, z), the kinetic energy per unit mass of the particle displaced transversely from (y, z) decreases or increases. The work E_0 which is performed in the course of a displacement (y, z) by the transverse force per unit mass, applied by the surroundings to the displaced particle, will be referred to as *energy of instability* per unit mass, latent at point O. It follows that

$$E_0 \equiv -\tfrac{1}{2}[a^0_{yy}y^2 + 2a^0_{yz}yz + a^0_{zz}z^2]. \tag{18}$$

In the case of stability this work is negative; the surrounding medium opposes displacement (y, z); in fact this displacement ceases as soon as the kinetic energy of the initial impulse is dissipated. Under conditions of instability, on the other hand, this work is positive: the surroundings further the displacement (y, z); to the kinetic energy of the initial impulse is added the energy of instability E_0 set free in the course of the displacement [35, 37]. This gives an interpretation on an energy basis of the quadratic form (13).

We return now to equation (4). By multiplying scalarly with $\mathbf{V}$, we obtain a differential expression which can be integrated directly. This furnishes the first integral of the equation of motion of the perturbed particle,

$$\tfrac{1}{2}[(V_x)^2 + (v_y)^2 + (v_z)^2] + \Theta_0\Pi + \Phi = \text{const}. \tag{19}$$

We set $S(y, z) \equiv \tfrac{1}{2}(V_x)^2 + \Theta_0\Pi + \Phi$, where V_x is given as a function of y and z by (3) and (8). After substituting initial conditions (5) and equation (17), equation (19) assumes the form [18]

$$S(y, z) - S(0, 0) = \tfrac{1}{2}[a^0_{yy}y^2 + 2a^0_{yz}yz + a^0_{zz}z^2]. \tag{20}$$

However, we have assumed that the perturbation of the particle considered is small, so that the right-hand side of (20) constitutes in reality only the first term of the expansion of the difference $S(y, z) - S(0, 0)$ in a Taylor series; therefore, $(\partial S/\partial y)_0 = (\partial S/\partial z)_0 = 0$ and $(\partial^2 S/\partial y^2)_0 \equiv a^0_{yy}$, $(\partial^2 S/\partial y \partial z)_0 \equiv a^0_{yz} \equiv a^0_{zy}$, $(\partial^2 S/\partial z^2)_0 \equiv a^0_{zz}$. *We conclude that a particle A, in the interior of an air mass in geostrophic motion, occupies an equilibrium position O when the sum of its longitudinal kinetic energy per unit mass $\tfrac{1}{2}(V_x)^2$, its specific enthalpy $\Theta_0\Pi$, and its*

potential energy per unit mass Φ is an extremum; depending upon whether this extremum is a minimum or a maximum, the equilibrium at O between gravity, the pressure gradient, and the Coriolis force is stable or unstable [38].

The Quadratic Form of E. Kleinschmidt

Let us now consider the case of vertical displacement ($r_y \equiv 0$, $r_z \neq 0$), the equilibrium being hydrostatic ($u \equiv 0$). In this case the earth's rotation can be neglected; by referring back to (12), we obtain $a_{yy} \equiv a_{yz} \equiv a_{zy} \equiv 0$ and $a_{zz} \equiv (g/\Theta)(\partial\Theta/\partial z)$, consideration having been taken of the hydrostatic equation in the vertical (z) direction. We thus rederive the classical criterion: the vertical distribution of air in the field of gravity g is stable or unstable depending upon whether $\partial\Theta/\partial z >$ or < 0. We observe that the energy of instability released per unit mass in the course of vertical displacement r_z of a particle in hydrostatic equilibrium is given by $-\frac{1}{2}(g/\Theta)(\partial\Theta/\partial z)r_z^2$.

In the case of a geostrophic current ($u \neq 0$), the vertical equilibrium is stable or unstable depending upon whether

$$a_{zz} \equiv 2\omega_y\left(2\omega_y + \frac{\partial u}{\partial z}\right) - \frac{\partial \Pi}{\partial z}\frac{\partial \Theta}{\partial z} > \text{or} < 0.$$

In the atmosphere, however, the second term is always at least one hundred times larger than the first and therefore determines the sign of a_{zz}.

We consider next the case of transverse isentropic displacement, which is defined at any point by $r_y^* \equiv \frac{r\partial\Theta/\partial z}{|\nabla\Theta|}$, $r_z^* \equiv -\frac{r\partial\Theta/\partial y}{|\nabla\Theta|}$. The value of the quadratic form (14) at this point for this displacement can be written independently of the chosen system of coordinates as follows [37]

$$Q(r_y^*, r_z^*) \equiv 2(\boldsymbol{\omega}\cdot\nabla\Theta)(\nabla\Theta\cdot\text{curl U})r^2/(\nabla\Theta)^2. \quad (21)$$

In order to give (21) a convenient analytical form, it is useful to adopt Kleinschmidt's rectangular, right-handed reference system XYZ [18], where axis X coincides with the longitudinal axis x and the transverse axis Y is perpendicular to the earth's axis and points toward the earth's interior. The direction cosines between the new axes X, Y, Z and the former axes x, y, z are as given in Table I. We can set $\kappa \sin\beta = \sin\phi$, $\kappa\cos\beta = \cos\phi\cos\alpha$, and $\kappa^2 = \sin^2\phi + \cos^2\phi\cos^2\alpha$.

Table I. Direction Cosines for Kleinschmidt's Coordinate System

	X	Y	Z
x	1	0	0
y	0	$\sin\beta$	$\cos\beta$
z	0	$-\cos\beta$	$\sin\beta$

In this coordinate system we have $\omega_X \equiv \omega\cos\phi\sin\alpha$, $\omega_Y \equiv 0$, $\omega_Z \equiv \kappa\omega$ and

$$\nabla\Theta\cdot\text{curl U} \equiv \frac{\partial\Theta}{\partial Z}\left(2\kappa\omega - \frac{\delta u}{\delta Y}\right), \quad (22)$$

where the operator

$$\frac{\delta}{\delta Y} \equiv \frac{\partial}{\partial Y} - \left(\frac{\partial\Theta/\partial Y}{\partial\Theta/\partial Z}\right)\frac{\partial}{\partial Z} \quad (23)$$

represents the isentropic derivative in the direction of the transverse axis Y. Following (21),

$$-\tfrac{1}{2}Q(r_y^*, r_z^*) \equiv \kappa\omega\left(\frac{\partial\Theta}{\partial Z}\right)^2\left(\frac{\delta u}{\delta Y} - 2\kappa\omega\right)\frac{r^2}{(\nabla\Theta)^2}. \quad (24)$$

It turns out that geostrophic motion at any point, for a transverse isentropic displacement from this point, is stable or unstable depending on whether

$$\frac{\delta u}{\delta Y} < \text{or} > 2\kappa\omega. \quad (25)$$

Furthermore we note that the right side of (24) gives the energy of instability released per unit mass during an isentropic displacement.

Finally, we come to the general case. Study of the sign of the quadratic form Q is accomplished with the help of its discriminant [18, 37]

$$a \equiv a_{yz}^2 - a_{yy}a_{zz} \equiv 2(\boldsymbol{\omega}\cdot\nabla\Pi)(\nabla\Theta\cdot\text{curl U}) \quad (26)$$

and of the coefficient of one of the squared terms. First we observe that, (1),

$$2\Theta(\boldsymbol{\omega}\cdot\nabla\Pi) = -2\boldsymbol{\omega}\cdot\nabla\Phi = -fg \quad (27)$$

where g represents gravity and f the Coriolis parameter $2\omega_z \equiv 2\omega\sin\phi$. After substituting (22) and (27) into (26), we obtain

$$a \equiv \frac{fg}{\Theta}\frac{\partial\Theta}{\partial Z}\left(\frac{\delta u}{\delta Y} - 2\kappa\omega\right). \quad (28)$$

We next find by reference to equations (12) that the coefficient of the term in $(r_Z)^2$ can be written

$$a_{ZZ} \equiv -\frac{\partial\Pi}{\partial Z}\frac{\partial\Theta}{\partial Z} = \frac{1}{\Theta}\frac{\partial\Theta}{\partial Z}\frac{\partial\Phi}{\partial Z} = \frac{g\sin\phi}{\kappa\Theta}\frac{\partial\Theta}{\partial Z}. \quad (29)$$

Since the sign of quadratic form Q depends upon the signs of a and a_{ZZ}, we see in the final analysis that this sign depends upon the sign of $\partial\Theta/\partial Z$ and the sign of $\delta u/\delta Y - 2\kappa\omega$; whence we can construct Table II [18, 37].

Table II. General Conditions for the Stability of Geostrophic Motion

Case	$\partial\Theta/\partial Z$	$\frac{\delta u/\delta Y}{-2\kappa\omega}$	a	Nature of the hydrodynamic equilibrium
I	+	−	−	Stable whatever the transverse displacement.
II	+	+	+	Unstable for transverse displacements close to the isentropic surface. Stable for transverse displacements close to the Z direction.
III	−	+	−	Unstable whatever the transverse displacement.
IV	−	−	+	Unstable for transverse displacements close to the Z direction. Stable for transverse displacements close to the isentropic surface.

When $a > 0$, that is, when the quadratic form Q at the point considered is positive for certain displacements and negative for others, it becomes zero for two transverse directions d' and d''. For these directions the hydrodynamic equilibrium at the point is neutral ($-Q \equiv \psi_y r_y + \psi_z r_z = 0$). It is simple to show that the angle between these two directions is given by

$$\tan (d'd'') = \frac{2\sqrt{2(\omega \cdot \nabla\Pi)(\nabla\Theta \cdot \text{curl } \mathbf{U})}}{2(\omega \cdot \text{curl } \mathbf{U}) - (\nabla\Pi \cdot \nabla\Theta)} = 10^{-2}. \quad (30)$$

Since the sign of Q at any point is that of $\partial\Theta/\partial Z$ for a displacement parallel to Z and is that of $2\kappa\omega - \delta u/\delta Y$ for a transverse isentropic displacement, and since these signs are different because of the inequality $a > 0$, the directions d' and d'' separate the transverse tangent to the isentropic surface at this point from the line parallel to the Z axis. The directions d' and d'' separate the transverse plane at this point into four regions, opposite to one another in pairs across the vertex. The two regions containing the displacements for which $Q > 0$ (<0) constitute the stable (unstable) sector of the transverse plane at the point considered. Moreover, the isentropic surface passing through this point lies in the unstable sector in Case II and in the stable sector in Case IV [8, 9, 14, 15, 17, 18].

When $a = 0$, the directions d' and d'' both coincide with direction

$$\frac{r_Z}{r_Y} = -\frac{a_{YZ}}{a_{ZZ}} = -\frac{a_{ZY}}{a_{ZZ}} = -\frac{\partial\Theta/\partial Y}{\partial\Theta/\partial Z},$$

that is, with the transverse tangent d to the isentropic surface at the point considered. When a becomes positive, in passing from Cases I and III to Cases II and IV, directions d' and d'' separate from this tangent and include between them an angle defined by (30). This angle is very small, averaging at the most on the order of a few tenths of a degree, and the isentropic surface at the point considered lies in the acute angle between directions d' and d'' from this point [37].

In the atmosphere we generally have $\partial\Theta/\partial Z > 0$, so that only Cases I and II are ordinarily possible, Case I being the most common. Therefore, in general, at any point in the atmosphere the state of geostrophic motion is stable whatever may be the transverse displacements from this point, with the exception of isentropic and quasi-isentropic displacements, when [37]

$$\frac{\delta u}{\delta Y} > 2\kappa\omega \quad \text{or} \quad \nabla\Theta \cdot \text{curl } \mathbf{U} < 0. \quad (31)$$

Hydrodynamic Instability as a Function of Latitude

Middle and High Latitudes. Let us return to the coordinate system $Oxyz$ introduced earlier. In the Northern Hemisphere, we have in this system $\partial\Pi/\partial y < 0$ and $\partial\Theta/\partial y < 0$. At the pole $\phi \equiv 90°$, so that $\omega_x \equiv \omega_y \equiv 0$. In the polar regions ϕ is close to 90° and the two horizontal components of ω are practically zero. Furthermore, we know that in middle latitudes we can neglect all the terms of the dynamic equations which contain $\omega \cos\phi$; when thus simplified the equations are exact, at least to the nearest hundredth. As a first approximation, we can thus eliminate the terms containing components ω_x and ω_y of the earth's rotation from all equations in the preceding sections. In particular we have, (12), [20, 37]

$$\begin{aligned}
a_{yy} &\equiv f\left[f - \dot{\Theta}\frac{\partial}{\partial y}\left(\frac{u}{\Theta}\right)\right] \cong f\left(f - \frac{\partial u}{\partial y}\right) \approx 10^{-8}\ \text{sec}^{-2},\\
a_{yz} &\equiv -f\Theta\frac{\partial}{\partial z}\left(\frac{u}{\Theta}\right) \cong -f\frac{\partial u}{\partial z} \approx 10^{-7}\ \text{to}\ 10^{-6}\ \text{sec}^{-2},\\
a_{zy} &\equiv \frac{g}{\Theta}\frac{\partial\Theta}{\partial y} \approx 10^{-7}\ \text{to}\ 10^{-6}\ \text{sec}^{-2},\\
a_{zz} &\equiv \frac{g}{\Theta}\frac{\partial\Theta}{\partial z} \approx 10^{-4}\ \text{sec}^{-2},
\end{aligned} \quad (32)$$

account having been taken of the simplified equation (1), and we have from (28),

$$a \equiv a_{yz}^2 - a_{yy}a_{zz} \equiv \frac{fg}{\Theta}\frac{\partial\Theta}{\partial z}\left(\frac{\delta u}{\delta y} - f\right), \quad (33)$$

as well as from (2),

$$\begin{aligned}
f\frac{\partial u}{\partial z} &= \frac{\partial\Theta}{\partial y}\frac{\partial\Pi}{\partial z} - \frac{\partial\Theta}{\partial z}\frac{\partial\Pi}{\partial y}\\
&\equiv N \cong -\frac{g}{\Theta}\frac{\partial\Theta}{\partial y} \approx 10^{-7}\ \text{to}\ 10^{-6}\ \text{sec}^{-2}.
\end{aligned} \quad (34)$$

The operator

$$\frac{\delta}{\delta y} \equiv \frac{\partial}{\partial y} - \left(\frac{\partial\Theta/\partial y}{\partial\Theta/\partial z}\right)\frac{\partial}{\partial z} \quad (35)$$

gives the transverse isentropic derivative in the direction of the horizontal y-axis. The sign of form Q for a transverse isentropic displacement (r_y^*, r_z^*) is that of the difference $f - \frac{\delta u}{\delta y}$, and for a vertical displacement r_z is that of a_{zz}, that is, that of $\partial\Theta/\partial z$. Thus we see that at any point the hydrodynamic equilibrium for a vertical displacement from this point is stable or unstable depending on whether the hydrostatic equilibrium in the field of gravity is stable or unstable. Therefore, to the approximation agreed upon here, the vertical gradient $\partial\Theta/\partial z$ of the potential temperature Θ reassumes its classical significance. We can also say that

$$\nu_s^2 \equiv \frac{g}{\Theta}\frac{\partial\Theta}{\partial z} \equiv g\,\frac{\gamma_a - \gamma}{T} \approx 10^{-4}\ \text{sec}^{-2} \quad (36)$$

defines the *static stability* in middle and high latitudes. In (36), T represents the absolute air temperature; γ, the vertical lapse rate of temperature; and γ_a, the adiabatic lapse rate. However, in the general case considered in the preceding section, $\partial\Theta/\partial z$ must be replaced by $\partial\Theta/\partial Z$ to be rigorous.

This having been established, we can adopt without change the reasoning of the previous section and substitute $\partial\Theta/\partial z$ and $\delta u/\delta y - f$ for $\partial\Theta/\partial Z$ and $\delta u/\delta Y - 2\kappa\omega$ in Table II. Since we generally have $\partial\Theta/\partial z > 0$ above the convective level, geostrophic motion in middle and high latitudes is stable for any transverse displacement, except in the neighborhood of the isentropic surfaces when $\delta u/\delta y > f$ [17, 18, 37]. In the extra-

tropical troposphere, $\partial u/\partial y$ is on the average negative and of the order of 10^{-5} sec^{-1}, $\partial u/\partial z$ is positive and of the order of 10^{-3} to 10^{-2} sec^{-1}, while the slope $-\dfrac{\partial\Theta/\partial y}{\partial\Theta/\partial z}$ of isentropic surfaces can attain the value 10^{-2}; consequently, under these conditions, $\delta u/\delta y$ is positive and of the order of 10^{-5} to 10^{-4} sec^{-1}. Then the vertical component $-\delta u/\delta y$ of the vorticity which the velocity field **u** produces in the isentropic surfaces is anticyclonic, and when it exceeds in absolute value the limit $f \approx 10^{-4}$ sec^{-1}, the geostrophic motion becomes unstable within the isentropic surfaces and in their neighborhood. In this case the transverse tangent d to the isentropic surface at any point is, to a first approximation, the bisector of the angle between directions d' and d'' which limit the sector of unstable displacements on either side of this surface at the point considered [37]. The energy of instability released per unit mass in the d direction is satisfactorily given by the approximation $(f/2)(\delta u/\delta y - f)r^2$ deduced from (24).

In the atmosphere we usually have $0 < \delta u/\delta y < f$ and $\partial\Theta/\partial z > 0$, and consequently the state of geostrophic motion is usually stable for any transverse displacement. In this case $a < 0$ and $a_{zz} > 0$. If we admit that the state of hydrodynamic equilibrium can be altered gradually, it may be asked which of the two quantities a and a_{zz} will be the first to change its sign. In reply it is enough to observe that if $a_{zz} = 0$, we have $a > 0$. Then a vanishes before a_{zz}, since for a and a_{zz} to vanish simultaneously it would be necessary that $a_{zz} \equiv a_{zy} \equiv 0$, that is, $\partial\Theta/\partial y \equiv \partial\Theta/\partial z \equiv 0$, a condition which does not occur in the atmosphere. From this follows the proposition [18]: *When the state of hydrodynamic equilibrium is altered, the threshold of hydrodynamic instability in an isentropic surface is reached before that of hydrostatic instability in the field of gravity.* In other words, in the atmosphere, hydrodynamic instability in an isentropic surface will develop before hydrostatic instability in the vertical direction.

In summary, in order for hydrodynamic instability to appear in a region—it will first appear in the isentropic surfaces—the vertical component of vorticity, produced by the geostrophic velocity field in the isentropic surfaces, must increase in absolute value and exceed the critical value f.

In the absence of Archimedean buoyant forces, form Q reduces to $f(f - \partial u/\partial y)r_y^2$. Then there is inertial stability or instability depending on whether $\partial u/\partial y <$ or $> f$. In the atmosphere we generally have $\partial u/\partial y < 0$ in the lower troposphere, as well as in the middle and high troposphere north of the "jet stream"; but south of the "jet stream" $0 < \partial u/\partial y < f$ [32]. The condition of inertial stability is thus generally attained in the atmosphere. We shall say that

$$\nu_i^2 \equiv f(f - \partial u/\partial y) \approx 10^{-8}\,\text{sec}^{-2} \tag{37}$$

defines the *inertial stability*.

Let us now examine more closely the criterion

$$\frac{\delta u}{\delta y} \equiv \frac{\partial u}{\partial y} - \left(\frac{\partial\Theta/\partial y}{\partial\Theta/\partial z}\right)\frac{\partial u}{\partial z} > f \tag{38}$$

for hydrodynamic instability. After $\partial\Theta/\partial y$ in (38) is replaced by its value taken from (34), this criterion assumes the form

$$\left[\frac{g}{\Theta}\frac{\partial\Theta}{\partial z}\right]\left[f\left(f - \frac{\partial u}{\partial y}\right)\right] \equiv \nu_s^2\nu_i^2 < N^2 = \left[f\frac{\partial u}{\partial z}\right]^2 \cong \left[\frac{g}{\Theta}\frac{\partial\Theta}{\partial y}\right]^2, \tag{39}$$

if we assume that static stability is in fact realized ($\nu_s^2 > 0$), which is usually the case. Inequality (39) shows that, whatever the magnitude of the inertial stability ν_i^2 and the static stability ν_s^2, the hydrodynamic instability can appear as soon as the baroclinity exceeds the critical value $N_* = \nu_s\nu_i$. *Instability of geostrophic motion is thus essentially dependent upon the baroclinic character of the atmosphere and can develop only in regions where this character is sufficiently pronounced*, that is, in frontal zones and in the tropical air immediately above and in the polar air immediately below these zones.

Inequality (39) suggests the introduction of the *hydrodynamic stability*

$$\nu_d^2 \equiv f\left(f - \frac{\delta u}{\delta y}\right) \cong \nu_i^2 - \frac{N^2}{\nu_s^2}, \tag{40}$$

which is always less than the inertial stability. We see at once that inertial instability ($\nu_i^2 < 0$ or $\partial u/\partial y > f$) implies *ipso facto* hydrodynamic instability ($\nu_d^2 < 0$); this is, what takes place in a region, generally of small extent, of the tropical troposphere south of the "jet stream" [22–24]. For a given static stability $(g/\Theta)\,\partial\Theta/\partial z$ and baroclinity $-(g/\Theta)\,\partial\Theta/\partial y$, the hydrodynamic stability is greater when the inertial stability is large, that is, when the vorticity $-\partial u/\partial y$, which current **u** produces in horizontal surfaces, is strongly cyclonic ($\partial u/\partial y << 0$); it is weaker when the inertial stability is small, that is, when the vorticity $-\partial u/\partial y$ is weakly cyclonic or anticyclonic ($\partial u/\partial y > 0$). This latter condition occurs in the tropical troposphere south of the "jet stream." In that case hydrodynamic instability can appear for a sufficiently large positive value of $\partial u/\partial y$.

The critical slope $\tan\alpha_\theta^*$ of the isentropic surfaces, defined by $\delta u/\delta y - f = 0$, can be expressed as

$$\tan\alpha_\theta^* = \nu_i^2/N \approx 10^{-2}. \tag{41}$$

The actual slope of these surfaces is given by

$$\tan\alpha_\theta = -\frac{\partial\Theta/\partial y}{\partial\Theta/\partial z} = N/\nu_s^2, \tag{42}$$

so that criterion (39) can also be written $\alpha_\theta > \alpha_\theta^*$. Thus hydrodynamic instability occurs whenever the slope of the isentropic surfaces exceeds its critical value (41).

Let us consider the isentropic surfaces Θ = const, drawn at constant intervals of, for example, $\Delta\Theta = 5°$. To a first approximation, the baroclinity at any point depends on the horizontal distance between two neighboring isentropic surfaces; the static stability depends

on their vertical separation. These two quantities increase or decrease depending on whether those distances decrease or increase. When the isentropic surfaces approach one another while maintaining constant slope, the baroclinity N and the static stability ν_s^2 both increase in such a way that N/ν_s^2 remains constant, so that by virtue of (40) or (41), packing of the isentropic surfaces can involve instability when a sufficient baroclinity is attained. On the other hand, when the slope of the isentropic surfaces increases while their horizontal separation remains constant, the static stability decreases while the baroclinity remains constant; this baroclinity can become critical for a sufficiently small value of the static stability. It follows that in a frontal zone where the static stability and the baroclinity are relatively large, instability can occur only if the baroclinity is particularly pronounced, that is, when the isentropic surfaces are not only very close to one another but also markedly inclined to the horizontal (case of active frontogenesis). On the other hand, in the tropical air immediately above the frontal surface and the polar air immediately below, where the static stability is much less pronounced, a much smaller baroclinity is enough for the threshold of hydrodynamic instability to be attained.

In the stratosphere, where the static stability is large (5×10^{-4} sec^{-2}), but where the baroclinity is relatively weak (quasi-horizontal isentropic surfaces), the state of geostrophic motion is stable [8, 9].

In summary, strong baroclinity and weak static stability favor hydrodynamic instability (see (40)).

The Lower Latitudes. At the equator, $\phi \equiv 0$, and therefore $\omega_x \equiv \omega \sin \alpha$, $\omega_y \equiv \omega \cos \alpha$, and $\omega_z \equiv 0$. By virtue of (1) $\partial \Pi/\partial y = \partial P/\partial y = 0$; furthermore we have (Table I) $\partial/\partial z = -\partial/\partial Y$ and $\partial/\partial y = \partial/\partial Z$. It follows that in (12), if (1) is taken into account, $a_{yy} \equiv a_{yz} \equiv a_{zy} \equiv a_{ZZ} \equiv a_{YZ} \equiv a_{ZY} \equiv 0$ and

$$a_{zz} \equiv a_{YY} \equiv 4\omega^2 \cos^2 \alpha + \Theta \frac{\partial}{\partial z}\left(\frac{u}{\Theta}\right) 2\omega \cos \alpha + \frac{g}{\Theta}\frac{\partial \Theta}{\partial z}. \tag{43}$$

All these formulas hold in the neighborhood of the equator, providing we assume that, as we approach the equator, $\partial P/\partial y$ approaches zero as does $\sin \phi$.

In equatorial latitudes the quadratic form Q reduces to the degenerate form $a_{zz} r_z^2$, which is thus independent of r_y; consequently, at the equator, out of all the possible transverse displacements only the vertical ones permit an energy exchange between the displaced particle and its environment. At lower latitudes the force acting on a displaced particle is vertical.

Finally, the hydrodynamic equilibrium in equatorial regions is stable or unstable depending on whether $a_{zz} >$ or < 0, that is, whether

$$\nu_s^2 \equiv \frac{g}{\Theta}\frac{\partial \Theta}{\partial z} > \text{or} < g\,\frac{2\omega \cos \alpha(2\omega \cos \alpha + \partial u/\partial z)}{2\omega u \cos \alpha - g} \cong -2\omega \frac{\partial u}{\partial z} \cos \alpha. \tag{44}$$

Since $|\, 2\omega(\partial u/\partial z) \cos \alpha \,| \leqq 10^{-6}$ sec^{-2}, we observe that *the threshold of hydrostatic instability in equatorial regions practically coincides with the threshold of hydrostatic instability in the vertical direction* [18].

Inertial Oscillations of the Geostrophic Current

The orbital motion of perturbed particle $A(x, y, z)$ is the motion of this particle relative to the geostrophic current $\mathbf{u}$. Referring to the coordinate system $Oxyz$ considered earlier, in which the origin O is the equilibrium position of A, the orbital velocity $\mathbf{V} - \mathbf{u}$ of A is defined by its coordinate components (3)

$$\left.\begin{aligned} \dot{x} = v_x = V_x - u = \frac{dx}{dt} - u, \\ \dot{y} = v_y = V_y = \frac{dy}{dt}, \\ \dot{z} = v_z = V_z = \frac{dz}{dt}. \end{aligned}\right\} \tag{45}$$

The dot indicates the substantial derivative with respect to time t in a reference system embedded in the geostrophic motion. Deriving (45) for time t in the reference system $Oxyz$, at rest with respect to the earth, we obtain

$$\left.\begin{aligned} \frac{dV_x}{dt} = \ddot{x} + \frac{\partial u}{\partial y}\dot{y} + \frac{\partial u}{\partial z}\dot{z}, \\ \frac{dV_y}{dt} = \frac{dv_y}{dt} = \ddot{y}, \\ \frac{dV_z}{dt} = \frac{dv_z}{dt} = \ddot{z}, \end{aligned}\right\} \tag{46}$$

account having been taken of the fact that v_x, v_y, and v_z are independent of the longitudinal coordinate x. By referring to equations (8) and (9) and to (11), (12), and (15), we obtain the equation of longitudinal orbital motion for the perturbed particle A,

$$\dot{x} = y \operatorname{curl}_Z^0 \mathbf{U} - z \operatorname{curl}_y^0 \mathbf{U}, \tag{47}$$

and the differential system of transverse orbital motion is

$$\left.\begin{aligned} \ddot{y} - 2\omega_x \dot{z} = -a_{yy}^0 y - a_{yz}^0 z, \\ \ddot{z} + 2\omega_x \dot{y} = -a_{zy}^0 y - a_{zz}^0 z. \end{aligned}\right\} \tag{48}$$

This system defines the motion of A', the projection of A on the transverse plane through O. We see at once that the transverse orbital motion is independent of the longitudinal motion and that the latter follows from the former. It is therefore sufficient to determine the motion defined by (48).

The problem of integrating a system with constant coefficients (48) is classical; we know that its general integral is a linear combination of circular and hyperbolic functions of time t. The form of this general integral depends on the nature of the roots of the characteristic equation of the system (48); this latter is

simply Solberg's equation of pulsations or orbital circular frequencies ν_0 of inertial oscillation [31], that is,

$$\nu_0^4 - b\nu_0^2 - a = 0, \tag{49}$$

where a is the discriminant (26) of Kleinschmidt's quadratic form Q (13) and $b \equiv a^0_{yy} + a^0_{zz} + 4\omega_x^2$.

We recall that to a positive root ν_0^2 of (49) there corresponds a periodic motion with an elliptical trajectory centered at O, and circular frequency ν_0, and to a negative root ν_0^2 there corresponds a motion with a hyperbolic trajectory about the same center, and of parameter $\sqrt{-\nu_0^2}$. In the first case ($\nu_0^2 > 0$), the inertial oscillation of A' around O is said to be stable, in the second case ($\nu_0^2 < 0$), unstable.

The sign of the roots ν_0^2 of (49) depends on the signs of a and b. But since the term a_{zz} in b is 10^4 times as large as either of the others, the sign of roots ν_0^2 of (49) depends in the final analysis on the signs of a and a_{zz}, that is, on the sign of the quadratic form Q. Consequently, *the inertial oscillations of A around O are stable or unstable depending on whether the hydrodynamic equilibrium at O is stable or unstable for transverse displacements* $\overrightarrow{OA'}$ *of the perturbed particle A.*

Let ν_S^2 and ν_D^2 be the roots of (49); we have

$$\nu_S^2 = \frac{b}{2}(1 + \sqrt{1 + 4a/b^2}),$$
$$\nu_D^2 = \frac{b}{2}(1 - \sqrt{1 + 4a/b^2}); \tag{50}$$

the first root is the larger when they are both real. Four cases can be distinguished, and they correspond to the four cases of Table II [31, 36, 37]:

Case I. When $\nu_S^2 > 0$ and $\nu_D^2 > 0$, the transverse orbital motion (48) is the resultant of two oscillations of circular frequencies ν_S and ν_D around elliptical trajectories with a common center O.

Case II. When $\nu_S^2 > 0$ and $\nu_D^2 < 0$, the motion is the resultant of a periodic motion on an elliptical trajectory of circular frequency ν_S and a motion on a hyperbolic trajectory of parameter $\sqrt{-\nu_D^2}$, with a common center O.

Case III. When $\nu_S^2 < 0$ and $\nu_D^2 < 0$, the motion is the resultant of two motions on hyperbolic trajectories with a common center O and parameters $\sqrt{-\nu_S^2}$ and $\sqrt{-\nu_D^2}$.

Case IV. When $\nu_S^2 < 0$ and $\nu_D^2 > 0$, the motion is the resultant of a periodic motion on an elliptical trajectory of circular frequency ν_D and a motion on a hyperbolic trajectory of parameter $\sqrt{-\nu_S^2}$, about the common center O.

As a special case when $\omega_x \equiv 0$, that is, when the current u is zonal ($\alpha \equiv 0$), these ellipses and hyperbolas degenerate into straight lines passing through O in the meridian plane from this point. The same situation holds under the approximation ($\omega_x = \omega_y = 0$) agreed to on page 438, but in this case the straight lines are transverse and no longer meridional. The straight lines which correspond to $\nu_S^2 \gtrless 0$ are quasi-vertical, while those related to $\nu_D^2 \gtrless 0$ lie in the neighborhood of the isentropic surface passing through O, which in general is very slightly inclined to the horizontal plane. In the latter case the straight line is a little more or a little less steeply inclined to the horizontal plane than the isentropic surface, depending on whether the inertial oscillation ν_D^2 is stable or unstable ($\nu_D^2 < 0$) [31].

When the current u deviates from the west-to-east direction, the longitudinal component ω_x of ω introduces an inertial influence which tends to substitute a transverse elliptical oscillation for the rectilinear meridional oscillation. The major axis of this very flattened ellipse is close to the vertical or to the isentropic surface depending on whether we are considering the oscillation of circular frequency ν_S or ν_D. In the case of instability these ellipses become hyperbolas elongated in the direction of the conjugate axis, which is quasi-vertical or quasi-isentropic depending on whether the motion is characterized by parameter $\sqrt{-\nu_S^2}$ or $\sqrt{-\nu_D^2}$.

Finally, the longitudinal motion (47) is superimposed on these transverse motions in such a way that the resultant orbital motion is on an elliptical or hyperbolic trajectory whose plane, to the approximation $\omega_x = \omega_y = 0$, is parallel to the current u and quasi-vertical or nearly tangent to the isentropic surface through O depending on whether one considers root ν_S^2 or ν_D^2 of equation (49) [31, 36, 37].

It is easy to see that $b \cong -\nabla\Theta\cdot\nabla\Pi \approx 10^{-4}$ sec^{-2} and $a \approx 10^{-12}$ sec^{-4}. The circular frequencies of the inertial oscillations are then approximated by the following values [31]:

$$\nu_S \cong \sqrt{b} \approx 10^{-2}\ \text{sec}^{-1} \text{ and } \nu_D \cong \sqrt{-a/b} \approx 10^{-4}\ \text{sec}^{-1},$$

and the corresponding periods approach the values $\tau_S \equiv 2\pi/\nu_S \approx 10$ min and $\tau_D \equiv 2\pi/\nu_D \approx 20$ hours ($\pi = 3.14\ldots$). From this there results the proposition [31]: *every geostrophic current allows two inertial oscillations, one of short, the other of long period.*

We shall now show that the inertial oscillation of short period is associated with the stability of hydrostatic equilibrium in the vertical direction z, and that the long-period oscillation is associated with the stability of hydrodynamic equilibrium in the isentropic surfaces [34, 37].

To the approximation $\omega_x = \omega_y = 0$, we have $\nu_S \cong \sqrt{b} \cong \sqrt{\frac{g}{\Theta}\frac{\partial\Theta}{\partial z}} \equiv \nu_s$, (critical circular frequency of Väisälä-Brunt). When hydrostatic stability prevails ($\partial\Theta/\partial z > 0$), ν_S is real, the corresponding displacement is a periodic function of time, and the short-period inertial oscillation is stable; on the other hand, when this equilibrium is unstable ($\partial\Theta/\partial z < 0$), ν_S is a pure imaginary, the corresponding displacement grows exponentially with time, and the oscillation is unstable.

Similarly we have

$$\nu_D \cong \sqrt{-\frac{a}{b}} \cong f\sqrt{1 - \frac{1}{f}\frac{\delta u}{\delta y}} \equiv \nu_d$$

(critical circular frequency of V. Bjerknes) and

$$\tau_D \cong \tau_0\left[1 - \frac{1}{f}\frac{\delta u}{\delta y}\right]^{-\frac{1}{2}},$$

where $\tau_0 \equiv 2\pi/f$ represents the Foucault pendulum half-day and the direction y is horizontal and perpendicular to current $\mathbf{u}$, pointing toward the left of this current (low-pressure region).

When the hydrodynamic equilibrium in the isentropic surfaces is stable ($\delta u/\delta y < f$), ν_D is real, the corresponding transverse displacement is a periodic function of time, and the inertial oscillation of long period is stable; on the other hand, when that equilibrium is unstable ($\delta u/\delta y > f$), ν_D is a pure imaginary, the corresponding displacement increases exponentially with time, and the oscillation is unstable. We also observe that the period τ_D is greater or less than the Foucault pendulum half-day depending on whether $\delta u/\delta y > 0$ or < 0. In the first case, which is the one usually encountered in the atmosphere, the period τ_D is of the same order of magnitude as the period of cyclonic waves. When, through an increase in the baroclinity, $\delta u/\delta y$ finally exceeds the threshold value f ($0 < f < \delta u/\delta y$) the quasi-isentropic inertial oscillations of period greater than the Foucault pendulum half-day and the hydrodynamic equilibrium for corresponding transverse displacements become unstable simultaneously. *It is worth noting that the inertial oscillation which may become unstable in the atmosphere is a quasi-isentropic oscillation whose period is precisely of the order of the period of atmospheric disturbances* [31], *and that this instability can appear only where these disturbances would normally occur* [34, 37].

In summary, there exists a unique and reciprocal correspondence between the stability and instability of inertial oscillations in the atmosphere on one hand, and the stability and instability of hydrostatic and hydrodynamic equilibrium on the other [34, 37].

Hydrodynamic Instability and Cyclogenesis in Extratropical Latitudes

Far from any perturbation, the vertical component of the velocity and the horizontal and vertical components of the acceleration of the air are practically zero. We can then reasonably assume that the state of geostrophic motion in these regions is that which precedes the initial stage of cyclogenesis. If we suppose that the geostrophic current $\mathbf{u}$ is zonal ($\alpha \equiv 0$, $\omega_x \equiv 0$), any particle displaced in the interior of this current from point (x, y, z) to point $(x + r_x, y + r_y, z + r_z)$ acquires an orbital motion according to equations (47) and (48)

$$\dot{r}_x = \left(f - \frac{\partial u}{\partial y}\right) r_y - \frac{\partial u}{\partial z} r_z, \qquad \ddot{r}_y = \psi_y, \qquad \ddot{r}_z = \psi_z. \tag{51}$$

Here ψ_y and ψ_z are the meridional and vertical components of the force ψ which acts on the displaced particle because of the nonuniformity of the field of flow $\mathbf{u}$ and of the field of potential temperature Θ associated with it [20]. These components are given by

$$\psi_y \cong -\nu_i^2 r_y + N r_z, \qquad \psi_z \cong N r_y - \nu_s^2 r_z, \tag{52}$$

where ν_s^2, ν_i^2, and N have been defined in (36), (37), and (34).

Let us first envisage a vertical displacement ($r_y \equiv 0$, $r_z \neq 0$). In this case the earth's rotation can be neglected ($f \cong 0$); therefore $\psi_y \cong 0$ and $\psi_z \cong -\nu_s^2 r_z$. Since the vertical distribution of mass is assumed to be stable, the vertical force ψ_z is a restoring force which tends to bring the displaced particle back to its point of departure, so that after several oscillations of short period (approximately 10 min) and of an amplitude which is smaller the greater the stability, the particle returns to its equilibrium position in the geostrophic current [39, 40]. Therefore, in an atmosphere possessing vertical hydrostatic equilibrium in which the air undergoes adiabatic transformations only, we can assume as a first approximation that on a synoptic scale the air particles are displaced along isentropic surfaces. The existence of these isentropic displacements, favored by vertical stability, led Rossby [29] to the concept of *lateral mixing* and Raethjen [27] to the similar idea of *Gleitaustausch* of the air.

To the approximation adopted above, the inertial oscillations of long period take place in the isentropic surfaces, and we are led to consider isentropic displacements. In this case we have $(\partial\Theta/\partial y)r_y + (\partial\Theta/\partial z)r_z \equiv 0$, and consequently the transverse force ψ which the surroundings apply to the displaced particle is horizontal to a first approximation; this gives $\psi_y \cong -\nu_d^2 r_y$ and $\psi_z \cong 0$. The horizontal orbital motion of the displaced particle is then defined by the differential system

$$\dot{r}_x - (\nu_d^2/f) r_y = 0, \qquad \ddot{r}_y + \nu_d^2 r_y = 0, \tag{53}$$

and the initial conditions $r_x = r_y = 0$ and $\dot{r}_x = 0$, $\dot{r}_y = v_0$, for $t = t_0 = 0$. Two cases can be distinguished depending on whether the meridional force is a restoring force ($\psi_y r_y < 0$, stability) or a force which carries the displaced particle always farther from its point of departure ($\psi_y r_y > 0$, instability) [39, 40].

First case: $\nu_d^2 \equiv f(f - \delta u/\delta y) > 0$. In this case the integral of (53) is an ellipse with parametric equations

$$r_x = \frac{v_0}{f}(1 - \cos \nu_d t), \qquad r_y = \frac{v_0}{\nu_d} \sin \nu_d t. \tag{54}$$

Having received at the initial instant a transverse impulse in the isentropic surface that contains it, the particle oscillates in this surface around an ellipse for which the longitudinal semiaxis is v_0/f and the horizontal projection of the transverse semiaxis is v_0/ν_d. The center of the ellipse lies on the longitudinal axis through the initial position, located with respect to that initial position in the same sense as current $\mathbf{u}$ or in the opposite sense depending on whether the particle rises ($v_0 > 0$) or descends ($v_0 < 0$) along the isentropic surface. The major axis of the ellipse is longitudinal or transverse depending on whether the vorticity, produced by the geostrophic current in the isentropic surface, is cyclonic ($\delta u/\delta y < 0$, great stability) or weakly anticyclonic ($0 < \delta u/\delta y < f$, slight stability). The ellipse reduces to an inertial circle when $\delta u/\delta y = 0$. The elliptical motion of the perturbed particle has a period $2\pi/\nu_d \geqq 20$ hours. Since $\dot{r}_x >$ or < 0 as $r_y >$ or < 0, the motion takes place in the anticyclonic sense. The surrounding medium, however, opposes this anticyclonic circulation, since in the stable case the dis-

placed particles do negative work at the expense of the energy of the neighboring air; displacement along the arc of the ellipse ceases as soon as the kinetic energy of the initial impulse is dissipated. Because of our original hypotheses, the displaced particles cannot describe their entire inertial trajectory, for they lose their individuality in a time less than $2\pi/\nu_d$. The particles descending the slope of an isentropic surface, following an elliptical arc, acquire a zonal velocity $u + \dot{r}_x$, less than that of the current u; those climbing the slope acquire a zonal velocity $u + \dot{r}_x$, greater than u. This transverse isentropic exchange, if it can be maintained, thus involves a weakening of the cyclonic vorticity ($-\delta u/\delta y > 0$) or an intensification of the anticyclonic vorticity ($0 > -\delta u/\delta y > -f$), that is, a decrease of the hydrodynamic stability in the isentropic surfaces or even the appearance of hydrodynamic instability.

Second case: $\nu_d^2 \equiv f(f - \delta u/\delta y) < 0$. In this case the integral of (53) is a hyperbola with the parametric equations

$$r_x = \frac{v_0}{f}(1 - \cosh|\nu_d|\, t), \qquad r_y = \frac{v_0}{|\nu_d|}\sinh|\nu_d|\, t. \quad (55)$$

Once the particle receives its velocity v_0, its distances r_x and r_y from initial position ($r_x = r_y = 0$) increase exponentially with time. Moving in the isentropic surface which passes through its point of departure, the particle traces a branch of a hyperbola, for which the longitudinal semiaxis is v_0/f and the horizontal projection of the transverse semiaxis is $v_0/|\nu_d|$. The center of the hyperbola lies on the longitudinal axis through the initial position, located with respect to that initial position in the same or in an opposite sense as the current u, depending on whether the particle rises ($v_0 > 0$) or descends ($v_0 < 0$) along the isentropic surface. The longitudinal axis is the transverse[1] axis of the hyperbola and is the major or minor axis depending on whether $\delta u/\delta y >$ or $< 2f$. The hyperbola is equilateral when $\delta u/\delta y = 2f$. We observe that the curvature of the branch of the hyperbola increases with the hydrodynamic instability. However, here $\dot{r}_x$ is $>$ or < 0 as $r_y <$ or > 0 and consequently the unstable isentropic inertial oscillation generates cyclonic circulation. Once set in motion, this circulation is maintained by the energy of instability released in the course of the isentropic displacement of the air particles. As before, it could be shown that the transverse isentropic exchange of air involves in this case a weakening of anticyclonic vorticity ($0 > -f > -\delta u/\delta y$), that is, a diminution of the hydrodynamic instability. *The cyclonic circulation which results from transverse isentropic displacement of the air in an unstable geostrophic flow tends to reestablish a state of stable dynamic equilibrium in the isentropic surfaces.* The result of this stabilizing action is that isentropic hydrodynamic instability is only a transitory state of short duration, which, according to what we have seen earlier, can develop only in regions of pronounced baroclinity.

1. As used here, the term "transverse" refers to the axis which passes through the vertices of the hyperbola.

This being established, it follows directly from the definition of the isentropic meridional gradient $\delta u/\delta y$ of a westerly geostrophic current u and from the criterion for hydrodynamic instability that this instability can appear only when the surfaces Θ = const and u = const are closely packed and intersect one another at a large angle. By referring to the vertical meridional cross sections of Palmén [22–24], we can verify that these conditions are established in only two regions [40]: (1) in the upper troposphere between 200 and 500 mb south of the belt of maximum westerlies, where $\delta u/\delta y - f$ is positive and of the order 10^{-5} sec^{-1}, and (2) in the lower troposphere between 600 and 900 mb in the zone of the polar front and immediately north of it. We observe that in the latter region the isopleths of Θ should be replaced by isopleths of the wet-bulb potential temperature, whose slopes are necessarily larger; actually therefore, the hydrodynamic instability in this portion of the lower troposphere is greater than Palmén's cross sections make it appear. In these two regions, hydrodynamic instability may occur.

On the other hand, (1) in the higher troposphere north of the belt of maximum westerlies the isopleths u = const and Θ = const are practically parallel, and (2) in the lower troposphere south of the belt of maximum westerlies the isopleths Θ = const are nearly horizontal. Hence these two regions are normally characterized by pronounced hydrodynamic stability [40]. In the higher troposphere, consequently, meridional isentropic displacements of air particles bring about the formation of cyclonic circulation at low latitudes in the temperate zone and anticyclonic circulation at higher latitudes [40]. The existence of these two circulations has been demonstrated by Rossby, Palmén, and their collaborators [32]. On the other hand, in the lower troposphere, meridional isentropic displacements of air particles bring about the formation of cyclonic circulation at high latitudes in the temperate zone and anticyclonic circulation at low latitudes in this zone. The existence of this latter circulation was shown by Wexler and Namias [45]. The cyclonic circulations at middle latitudes in the lower troposphere are simply those characterizing the normal activity of the polar front.

In summary, when the geostrophic motion becomes unstable for isentropic and quasi-isentropic transverse displacements in a baroclinic region of the atmosphere, the surrounding medium favors these displacements; the displaced particles set free energy of instability which augments the kinetic energy of the initial impulse [35–37]. Moreover, they acquire cyclonic circulation relative to the geostrophic current along branches of hyperbolas whose conjugate axes are perpendicular to the geostrophic flow [39].

Thus we perceive that an isentropic inertial oscillation of period greater than the Foucault pendulum half-day ($2\pi/f$), *which becomes unstable in a region of large baroclinity, may give rise to a cyclonic circulation.* A *nascent* cyclonic circulation hence seems to be no more than an unstable inertial oscillation of the atmosphere [31, 37]. But such a circulation inevitably involves major displacements of air and consequently a perturbation

of the pressure field. The motion then set up loses its inertial character. In short a theory of cyclogenesis cannot be established without consideration of a pressure perturbation [31, 42].

Hydrodynamic Instability and the Theory of Atmospheric Perturbations

In order to simplify the equations, we will designate by x^1 the horizontal meridional coordinate y and by x^2 the vertical coordinate z; hence the Eulerian variables x, y, z, and t will here be written x, x^1, x^2, and t. In this system of notation we have: $\omega_x \equiv 0$, $2\omega_y \equiv 2\omega \cos \phi = \partial W/\partial x^2$ and $2\omega_z \equiv f \equiv 2\omega \sin \phi = -\partial W/\partial x^1$, where $\mathbf{W} \equiv \boldsymbol{\omega} \times \mathbf{R}$ represents the linear velocity due to the earth's rotation. Equation (1) of zonal "simple motion" then assumes the form [42]

$$S\nabla P = \Theta\nabla\Pi = u\nabla W - \nabla\Phi, \tag{56}$$

while the equilibrium condition (2) can be written

$$\nabla u \times \nabla W = \nabla U \times \nabla W = \nabla\Theta \times \nabla\Pi = \nabla S \times \nabla P, \tag{57}$$

where $u \equiv u(x^1, x^2)$ represents the intensity of the westerly current, $U \equiv U(x^1, x^2) = W + u$ the absolute zonal current, and $S = S(x^1, x^2)$ the specific volume of air.

This being the case, the equations of adiabatic "small motions" [2] which can be superimposed on simple zonal motion (56) can be written [42]

$$Dv_x + \frac{\partial U}{\partial x^i} v^i + S\frac{\partial p}{\partial x} = 0, \tag{58}$$

$$Dv^i - \frac{\partial W}{\partial x^i} v_x + S\frac{\partial p}{\partial x^i} + s\frac{\partial P}{\partial x^i} = 0, \tag{59}$$

$$Ds + \frac{\partial S}{\partial x^i} v^i - S\frac{\partial v_x}{\partial x} - S\frac{\partial v^i}{\partial x^i} = 0, \tag{60}$$

$$Ds + \frac{S^2}{C^2} Dp + \frac{S}{\Theta}\frac{\partial\Theta}{\partial x^i} v^i = 0, \tag{61}$$

where $i = 1, 2$; p designates the perturbation pressure; s, the corresponding perturbation of specific volume; C, Laplace's speed of sound,

$$C^2 \equiv (c_p/c_v)PS \equiv (c_p/c_v)RT \approx 10^5 \text{ m}^2 \text{ sec}^{-2} \text{ [2]};$$

c_v, the specific heat at constant volume; T, the absolute air temperature; v_x, the zonal component; $v_y \equiv v^1$, the meridional component; and $v_z \equiv v^2$, the vertical component of the perturbation velocity v relative to the earth. We have also substituted $D \equiv \partial/\partial t + u\partial/\partial x$. The operator D represents the local derivative with respect to time t in a refereuce system embedded in the zonal current **u**. By eliminating v_x and s among the Eulerian equations (58), (59), and the equation of the adiabatic process (61), we obtain the equations of transverse perturbation motion (v^1, v^2) as a function of perturbation pressure p; these are

$$D^2 v^i + a_{ij} v^j = \frac{S^2}{C^2}\frac{\partial P}{\partial x^i} Dp - \frac{\partial W}{\partial x^i} S\frac{\partial p}{\partial x} - SD\frac{\partial p}{\partial x^i}, \tag{62}$$

where, (57),

$$a_{ij} \equiv \frac{\partial W}{\partial x^i}\frac{\partial U}{\partial x^j} - \frac{S}{\Theta}\frac{\partial P}{\partial x^i}\frac{\partial\Theta}{\partial x^j} \equiv \frac{\partial W}{\partial x^j}\frac{\partial U}{\partial x^i} - \frac{S}{\Theta}\frac{\partial P}{\partial x^j}\frac{\partial\Theta}{\partial x^i} \equiv a_{ji} \tag{63}$$

are simply the coefficients (12) of Kleinschmidt's quadratic form Q (13).

We next eliminate v_x and s between the equation (58) of zonal perturbation motion and the equations of continuity (60) and the adiabatic process (61),

$$D\frac{\partial v^i}{\partial x^i} + \frac{S}{C^2}\frac{\partial P}{\partial x^i} Dv^i - \frac{\partial U}{\partial x^i}\frac{\partial v^i}{\partial x} + S\left(\frac{D^2 p}{C^2} - \frac{\partial^2 p}{(\partial x)^2}\right) = 0. \tag{64}$$

System (62) and (64) of three equations containing partial derivatives of three unknown functions v^1, v^2, and p with respect to the independent variables x, x^1, x^2, and t, must be integrated under the following conditions at the external boundaries:

$$\left.\begin{aligned} &(a) \text{ at the surface of the globe:} \\ &\qquad x^2 \equiv z = 0; \qquad v^2 \equiv v_z = 0, \\ &(b) \text{ at the free surface:} \\ &\qquad P(x^1, x^2) + p(x, x^1, x^2, t) = \text{const}; \\ &\qquad Dp + (\partial P/\partial x^i)v^i = 0. \end{aligned}\right\} \tag{65}$$

We observe that setting $p \equiv 0$ in the differential system (62) reduces it to the differential system (48) for inertial oscillations of a zonal geostrophic current ($\omega_x \equiv 0$) and that the characteristic determinant

$$D^4 + (a_{11} + a_{22})D^2 + a_{11}a_{22} - (a_{12})^2$$

of (62) is formally identical with the left side (49) of Solberg's equation of circular frequencies. This circumstance demonstrates *the influence of the nature of the inertial oscillations or, what amounts to the same thing, the influence of the nature of the hydrodynamic equilibrium upon the solution of the differential system* (62) *and* (64) *for transverse perturbation motion.*

After having thus stated in all its generality the problem in mathematical analysis posed by the theory of perturbations, we make the customary restriction to a sinusoidal disturbance which is propagated by plane waves with constant velocity c parallel to the zonal axis x. In this case the quantities v_x, v^i, s, and p are proportional to $\exp[\sqrt{-1}\,\mu(x - ct)]$, with $\mu \equiv 2\pi/L$, where L is the wave length. We designate by ξ, ξ^i, σ, and ϖ the amplitudes of these quantities, which are functions of the transverse coordinates x^1 and x^2. The differential system (62) gives the amplitudes ξ^i directly as functions of ϖ; we find [42]

$$\xi^i = -\sqrt{-1}\,\mu(u - c)\,SA^{ij}\left(\frac{\partial\varpi}{\partial x^j} - K_j\varpi\right), \qquad (i, j = 1, 2) \tag{66}$$

where A^{ij} is the normalized algebraic minor of the element $A_{ij} \equiv -\mu^2(u - c)^2\gamma_{ij} + a_{ij}$ of the determinant $A \equiv \| A_{ij} \|$, with $\gamma_{ij} \equiv 1$ when $i \equiv j$ and $\equiv 0$ when

$i \neq j$. In (66) the K_j's are the components of the transverse vector

$$\mathbf{K} = \frac{S\nabla P}{C^2} - \frac{\nabla W}{u-c} = -\frac{\nabla W}{u-c}\left(1 - \frac{u(u-c)}{C^2}\right) - \frac{\nabla\Phi}{C^2}. \tag{67}$$

Equation (66) has meaning only if $A \neq 0$, that is, when the orbital circular frequency $\nu_0 \equiv \mu(c-u)$ of the perturbation is different from the circular frequencies ν_S and ν_D for inertial oscillations. When $\nu_0 \equiv \nu_S$ (or ν_D), and this may occur at certain points of the transverse plane, the amplitude relations deduced from (62) are compatible only if

$$\left(\frac{\partial\varpi}{\partial x^1} - K_1\varpi\right) \Big/ \left(\frac{\partial\varpi}{\partial x^2} - K_2\varpi\right) = A_{11}/A_{12} = A_{21}/A_{22}$$

at the above-mentioned points.

Continuing the analysis, we substitute (66) into (58) and (61) and obtain the amplitudes ξ and σ as functions of ϖ, that is,

$$\left.\begin{aligned} \xi &= SA^{ij}\frac{\partial U}{\partial x^i}\left(\frac{\partial\varpi}{\partial x^j} - K_j\varpi\right) - \frac{S\varpi}{u-c}, \\ \sigma &= \frac{S^2A^{ij}}{\Theta}\frac{\partial\Theta}{\partial x^i}\left(\frac{\partial\varpi}{\partial x^j} - K_j\varpi\right) - \frac{S^2\varpi}{C^2}. \end{aligned}\right\} \tag{68}$$

Finally we replace ξ^i, ξ, and σ in (60) by their values in (66) and (68); there results an equation in the partial derivatives of amplitude ϖ of the perturbation pressure p [31, 42],

$$\frac{1}{S}\frac{\partial}{\partial x^i}\left(SA^{ij}\frac{\partial\varpi}{\partial x^j}\right) - \left[\frac{1}{S}\frac{\partial}{\partial x^i}(SA^{ij}K_j) + A^{ij}K_iK_j - \frac{1}{(u-c)^2}\left(1 - \frac{(u-c)^2}{C^2}\right)\right]\varpi = 0. \tag{69}$$

We observe at once that equation (69) is of elliptic or hyperbolic type depending on whether $A >$ or < 0. The determinant $A \equiv \|A_{ij}\| \equiv \nu_0^4 - (a_{11} + a_{22})\nu_0^2 + a_{11}a_{22} - (a_{12})^2$ is identical with the first member of the equation for circular frequencies of inertial oscillation; therefore, we have $\|A_{ij}\| \equiv (\nu_0^2 - \nu_D^2)(\nu_0^2 - \nu_S^2)$. Since $|\nu_0| < |\nu_D| \ll |\nu_S|$ and $\nu_S^2 > 0$ in the atmosphere, $A >$ or < 0 as $\nu_D^2 >$ or < 0; there follows the proposition [42]: *the partial differential equation in the perturbation pressure is of elliptic or hyperbolic type depending on whether the geostrophic motion is stable or unstable.*

The amplitude ϖ which solves the present problem is that solution of the second order equation (69) which satisfies the boundary conditions [42]

$$\left.\begin{aligned} &A^{2i}\left(\frac{\partial\varpi}{\partial x^i} - K_i\varpi\right) = 0 \\ &\text{for } x^2 = 0 \text{ (or } z = 0), \text{ and} \\ &SA^{ij}\frac{\partial P}{\partial x^i}\left(\frac{\partial\varpi}{\partial x^j} - K_j\varpi\right) - \varpi = 0 \\ &\text{for } P(x_1, x_2) + \varpi(x^1, x^2)\exp\left[\sqrt{-1}\,\mu(x - ct)\right] \\ &= \text{const.} \end{aligned}\right\} \tag{70}$$

It is probable (1) that the existence of a solution ϖ of (69) satisfying (70) implies a dispersion equation among the quantities ν_s^2, ν_i^2, and N, characteristic of the simple motion, on one hand, and the wave length L and the speed of propagation c of the perturbation on the other hand; and (2) that the condition as to whether the dispersion equation has imaginary roots c depends on the sign of $\nu_i^2\nu_s^2 - N^2$, that is, the hydrodynamic stability. However, hydrodynamic instability is only a sufficient condition for dynamic instability of the perturbation, since selective instability of a perturbation is possible even in the case of a stable "simple motion."

Clearly the problem of these perturbations is so difficult that drastic assumptions are inevitable if one really wishes to undertake its solution. Generally the discussion is confined to homogeneous ($S \equiv$ constant), incompressible ($C = \infty$), or isothermal ($C \equiv$ constant) media; vertical accelerations are neglected (quasi-static hypothesis); the basic flow $\mathbf{u}$ is assumed to be constant or at most to be linearly dependent on altitude z [3], or it is assumed that the current $\mathbf{u}$ consists of two constant, parallel, adjacent flows; sometimes the meridional component of the perturbation velocity is even assumed to be geostrophic [3]. The effect of these simplifications is to make the partial differential equation (69) degenerate into an ordinary differential equation, integrable by means of simple functions.

Four fundamental factors are involved in the general equations (62) and (64): (1) *gravity* g (gravity waves), (2) *compressibility* C^{-2} (sound waves), (3) *the earth's rotation* (inertial waves) and the variation of the Coriolis parameter with latitude (planetary waves [28]) and (4) the *zonal current* u and its vertical and meridional gradients (shear waves). Examination, employing simple media, of each of these factors separately (the others being assumed to be zero) demonstrates that the atmospheric perturbations ($c \cong 10$ m sec^{-1}, $L \cong 10^6$ m) can be compared to inertial waves whose instability results from the gradient of u [2, 31].

On the whole, the qualitative results of the perturbation theory have been obtained under assumptions which are never even approximately realized in the atmosphere. To be convinced of this it is sufficient to refer to recently published vertical cross sections of the "jet stream" [22–24, 32]. The perturbations which interest us are those of a zonal baroclinic ($N \neq 0$) current whose intensity u depends not only on altitude z but also on latitude y. *The problem of atmospheric perturbations is thus one in partial differential equations* and not a problem of an ordinary differential equation. The synoptic weather charts show that the propagation velocity c and the wave length L are such that products and squares of the ratios u/C, c/C, and $(u-c)/C$, as well as the square of the ratio $[2\pi(u-c)/L]/f$, are negligible compared to unity [3, 4]. One is then justified in making these approximations, but they must be introduced *not* into the initial equations (58) to (61), which must retain all their generality, but *rather* into the final equation (69).

To this end we must calculate expressions for A^{ij} and K_i, substitute them into equations (69) and (70), carry out the derivations involving y and z, and finally estimate the order of magnitude of the terms in these equations, taking into account the above-mentioned approximations. By neglecting terms of order 10^{n-2} compared with those of order 10^n, we would obtain the approximate differential system of atmospheric perturbations, which would then have to be integrated. Unfortunately, it is impossible at present to make this choice of terms, for we have not drawn up tables giving the orders of magnitude of the meteorological elements and their first and second derivatives with respect to space and time as functions of altitude and latitude. It seems urgent to us to prepare such tables with the help of the sufficiently extensive aerological data which is already at our disposal.

Meanwhile, using Charney's theory of meteorological approximations [4], we can deduce from general equation (69) the approximate partial differential equation of the perturbation pressure associated with quasi-static, quasi-geostrophic long waves. There is obtained [43]

$$\frac{\nu_s^2}{f^2}\frac{\partial^2\varpi}{\partial y^2} + \frac{\partial^2\varpi}{\partial z^2} + \left[\frac{2}{S}\frac{\partial S}{\partial z} - \frac{g}{RT} - \frac{1}{T\nu_s^2}\frac{\partial(T\nu_s^2)}{\partial z}\right]\frac{\partial\varpi}{\partial z} - \left\{\frac{\nu_s^2\mu^2}{f^2}\left(1 - \frac{u_c}{u-c}\right) + \frac{1}{u-c}\left[\frac{\nu_s^2}{f^2}\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \left(\frac{g}{RT} + \frac{1}{T\nu_s^2}\frac{\partial(T\nu_s^2)}{\partial z}\right)\frac{\partial u}{\partial z}\right] + \frac{1}{S}\frac{\partial S}{\partial z}\frac{1}{T\nu_s^2}\frac{\partial(T\nu_s^2)}{\partial z}\right\}\varpi = 0, \tag{71}$$

where T designates the absolute air temperature and where $u_c \equiv (\partial f/\partial y)/\mu^2$ is the critical speed introduced by Bjerknes and Holmboe [1], that is, the intensity of zonal flow which corresponds to stationary, planetary waves of wave length $L \equiv 2\pi/\mu$ (Rossby [28]). Equations (66) and (68) reduce to

$$\xi^1 \equiv \eta = \sqrt{-1}\,\frac{\mu S}{f}\varpi,$$
$$\xi^2 \equiv \zeta = -\sqrt{-1}\,\frac{\mu(u-c)^2}{\nu_s^2}\frac{\partial}{\partial z}\left(\frac{S\varpi}{u-c}\right), \tag{72}$$

$$\xi \equiv -\frac{S}{f}\frac{\partial\varpi}{\partial y}, \qquad \sigma = \frac{S^2}{g}\frac{\partial\varpi}{\partial z}. \tag{73}$$

Equation (71) is identical with that of Charney [4] when T depends linearly on z (that is, $\partial(T\nu_s^2)/\partial z \equiv 0$, see eq. (36)) and u is a function of z only. Dividing the first term of (71) by ν_s^2, and letting ν_s^2 approach infinity, one obtains Kuo's equation [19].

Equation (71) is to be integrated with suitable boundary conditions. The boundary conditions other than those referring to the earth's surface ($\zeta = 0$ for $z = 0$) must be carefully chosen, for on them depend the form of the dispersion equation and the expression for the criterion of selective instability. The boundary conditions being linear in the unknown function and its first derivatives, the integration of equation (71) is a problem of the classic "mixed" type. This integration can be achieved only with computing machines.

It should be observed that in the case considered here the "simple motion" is a weakly baroclinic zonal current u ($N \approx 10^{-7}$) in which the isentropic surfaces have a slope of the order of 10^{-3}. Under these conditions equation (71) is always of elliptic type; for this reason any hydrodynamic instability is excluded and only the selective instability is possible. The constant c is then complex ($c = c_r + \sqrt{-1}c_i$), as is the function ϖ (that is, $\varpi \equiv \varpi' + \sqrt{-1}\,\varpi''$), so that the perturbation pressure assumes the form

$$p = e^{c_i\mu t}[\varpi'(y, z)\cos\mu(x - c_r t) - \varpi''(y, z)\sin\mu(x - c_r t)]. \tag{74}$$

Here the real functions ϖ' and ϖ'' satisfy a system of partial differential equations of elliptic type which is easy to deduce from (71). In this case the form of the perturbation pressure p demonstrates the phase lag, both in altitude and latitude, which the troughs and ridges of unstable perturbations undergo. The meridional tilt of the axes of troughs and ridges is an essential characteristic of atmospheric perturbations. Their tilt relative to the meridians assures the transport of zonal momentum and kinetic energy along them.

The Stability of Permanent, Horizontal, Isobaric Motion

When the air is in permanent, horizontal, isobaric motion, the atmosphere is said to be in a state of hydrodynamic equilibrium. The dynamic method which we employed (pp. 434–437) for the case of rectilinear isobars can be applied to the case of curvilinear isobars, provided we use a system of orthogonal curvilinear coordinates σ^1, σ^2, σ^3, fixed relative to the earth and having one (σ^1) of the three families of coordinate lines coincident with the lines of intersection between the isobaric surfaces and the equipotential surfaces of the gravity field. Let $Oxyz$ be a right-handed, rectangular Cartesian coordinate system of origin O, in which the axes are tangent to the coordinate lines passing through O; axis Ox points in the direction of the isobaric motion. We designate by $u \equiv u(\sigma^1, \sigma^2)$ the velocity of air along the isobars obtained by varying σ^1 (gradient wind), by R_y the radius of geodesic curvature, and by R_z the radius of normal curvature at O of isobar σ^1 considered as a line in the surface $\sigma^3 = \sigma_0^3$ passing through O. By then adopting without change the reasoning presented on pages 434–437, we find that the equilibrium between gravity, the pressure gradient, the centrifugal force, and the Coriolis force at point O, for a transverse displacement (r_y, r_z) from this point, is stable or unstable depending on whether [41]

$$Q(r_y, r_z) \equiv a_{yy}r_y^2 + 2a_{yz}r_y r_z + a_{zz}r_z^2 > \text{ or } < 0, \tag{75}$$

where

$$
\begin{aligned}
a_{yy} &\equiv 2\left(\omega_z + \frac{u}{R_y}\right)\left(2\omega_z - \frac{\partial u}{\partial y} + \frac{u}{R_y}\right) - \frac{\partial \Pi}{\partial y}\frac{\partial \Theta}{\partial y}, \\
a_{yz} &\equiv -2\left(\omega_z + \frac{u}{R_y}\right)\left(2\omega_y + \frac{\partial u}{\partial z} - \frac{u}{R_z}\right) - \frac{\partial \Pi}{\partial y}\frac{\partial \Theta}{\partial z}, \\
a_{zy} &\equiv -2\left(\omega_y - \frac{u}{R_z}\right)\left(2\omega_z - \frac{\partial u}{\partial y} + \frac{u}{R_y}\right) - \frac{\partial \Pi}{\partial z}\frac{\partial \Theta}{\partial y}, \\
a_{zz} &\equiv 2\left(\omega_y - \frac{u}{R_z}\right)\left(2\omega_y + \frac{\partial u}{\partial z} - \frac{u}{R_z}\right) - \frac{\partial \Pi}{\partial z}\frac{\partial \Theta}{\partial z},
\end{aligned}
\tag{76}
$$

with $a_{yz} \equiv a_{zy}$. We note that in the differential quotients on the right-hand sides of (76), dy and dz represent elements of arc along the coordinate lines σ^2 and σ^3 and that these elements of arc are not in general exact differentials [41].

There is an obvious analogy between the quadratic form (75) and equation (13) which holds in the case of geostrophic motion. In the case of isobaric motion, this form retains the dynamical significance and the interpretation on an energy basis of Kleinschmidt's form (see pp. 434–437). Moreover, we can apply the discussion of Kleinschmidt's quadratic form to the sign of this form. If we assume the vertical distribution of mass to be stable in the field of gravity, the state of isobaric motion is then stable at any point, whatever the transverse displacements from the point, with the exception of displacements close to the isentropic surface when

$$\nabla\Theta \cdot \operatorname{curl} \mathbf{U} \equiv \frac{\partial \Theta}{\partial y} \operatorname{curl}_y \mathbf{U} + \frac{\partial \Theta}{\partial z} \operatorname{curl}_z \mathbf{U} < 0,$$

where

$$\operatorname{curl}_x \mathbf{U} \equiv 2\omega_x, \qquad \operatorname{curl}_y \mathbf{U} \equiv 2\omega_y + \frac{\partial u}{\partial z} - \frac{u}{R_z},$$

$$\operatorname{curl}_z \mathbf{U} \equiv 2\omega_z - \frac{\partial u}{\partial y} + \frac{u}{R_y};$$

$\mathbf{U} \equiv \mathbf{W} + \mathbf{u}$ being the absolute velocity of the air.

When the coordinate surfaces $\sigma^3 = \text{const}$ are equipotential surfaces in the field of gravity, the z-axis coincides with the zenith direction at the origin. If we assume that the equigeopotential surfaces are spherical and concentric with the earth, $R_z = -r$, where r stands for the distance from the earth's center to the point under consideration. In this case, R_y represents the radius of curvature at O of the horizontal projection of the isobar obtained by varying σ^1; in synoptic meteorology R_y is "the radius of curvature R_{is} of the isobar" at this point, ($R_y \equiv R_{is}$). We note that $R_{is} > 0$ or < 0, depending on whether the circulation along the isobar is cyclonic or anticyclonic at the point considered. Since $2\omega_y + u/r$ is negligible compared to $\partial u/\partial z$, it is evident that the isobaric motion is unstable in the neighborhood of the isentropic surfaces when

$$\frac{\delta u}{\delta y} > f + \frac{u}{R_{is}}, \tag{77}$$

where $\delta u/\delta y$ has been defined by (35). This formula shows that the threshold value $f + u/R_{is}$ of the instability of isobaric motion is higher or lower than the threshold value f of the instability of geostrophic motion ($R_{is} = \infty$) depending on whether the curvature of the isobar is positive or negative. *Anticyclonic circulation thus favors hydrodynamic instability* [41].

When the isobar at O coincides with the parallel through this point, $R_{is} \equiv r \operatorname{ctn} \phi$, where ϕ is the latitude of O; here the limit of hydrodynamic instability is given by the approximate expression [22, 41]

$$\frac{\delta u}{\delta y} \cong 2\omega \sin \phi + \frac{u \tan \phi}{r}. \tag{78}$$

By virtue of the instability criterion (77) for a permanent, isobaric, horizontal current, cyclonic currents as a whole are more stable than anticyclonic currents. Let us assign an arbitrary value ($\delta u/\delta y$) to the transverse isentropic gradient $\delta u/\delta y$ of the isobaric current u. Let us suppose that it satisfies the condition most generally encountered, that is, $\delta u/\delta y < f$. The domain of stable currents will then be separated from the domain of unstable currents by an anticyclonic current of the curvature $1/R^*_{is} = (1/u)[(\delta u/\delta y) - f] < 0$. In this case, the geostrophic current ($R_{is} = \infty$), the limit between the domains of cyclonic and anticyclonic currents, will be located within the domain of stable currents composed of all cyclonic currents and of weakly anticyclonic currents ($|R_{is}| > |R^*_{is}|$). The domain of unstable flow contains only anticyclonic currents with a radius of curvature less than $|R^*_{is}|$.

Let us now admit, following Wippermann [46], the *existence of a transverse isentropic exchange* of air particles. When the current u is stable, this exchange can maintain itself only at the expense of the surrounding medium, and the displaced particles acquire an anticyclonic circulation with respect to the current u (see p. 442), thereby attenuating the curvature of the current u when it is cyclonic or reinforcing it when it is weakly anticyclonic ($|R_{is}| > |R^*_{is}|$). In this case, the current u, whether cyclonic ($R_{is} > 0$) or weakly anticyclonic ($R_{is} < 0$ with $|R_{is}| > |R^*_{is}|$) will be transformed progressively into an anticyclonic current whose radius of curvature will approach the limit $|R^*_{is}| \approx 10^6$ m. On the other hand, if the current u were unstable ($R_{is} < 0$ and $|R_{is}| < |R^*_{is}|$), the transversely displaced particles in the isentropic surfaces would acquire a cyclonic circulation (see p. 443) with respect to the current u, releasing energy of instability. The anticyclonic curvature of the unstable current u would thereupon diminish, that is, the current u would spontaneously become a current within the domain of stability. Summarizing, we may say that an anticyclonic current whose radius of curvature $|R_{is}| > |R^*_{is}|$ approaches the limit $|R^*_{is}|$ would thus appear to be a current of some persistence [46].

Inasmuch as hydrodynamic instability is a transitory state of ephemeral duration, it will not manifest itself unless its cause is of even less duration (*i.e.*, produced instantaneously); according to Kleinschmidt

[18], large-scale condensation is such a cause. The gradient $\delta u/\delta y$ in a cloud layer should be determined in surfaces of equal wet-bulb potential temperature rather than in surfaces of equal dry-bulb potential temperature. Inasmuch as the slope of the former is much more marked than that of the latter, it follows that $\delta u/\delta y$ undergoes, at the moment of condensation, a sudden increase which is capable of bringing about hydrodynamic instability ($\delta u/\delta y > f$). We then have $R_{is}^* > 0$ and, as a result, the current u will spontaneously become a cyclonic current of the stable domain ($0 < R_{is} \lesseqgtr R_{is}^*$) (cyclogenesis according to Wippermann [46]).

The Permanent Circular Vortex. Application to Tropical Cyclones

The problem of the stability of this state of motion has been treated in several different ways [2, 5, 6, 8, 9, 13–18, 21, 31]; we will consider it here as a special case of permanent, horizontal, isobaric motion. To this end we compare the atmosphere to a circular vortex around the polar axis and at some point O we adopt Kleinschmidt's coordinate system $OXYZ$, having axis OX tangent to the parallel through O and directed eastward, and axis OY perpendicular to the vortex axis and directed toward the earth's interior. Under these conditions, $dY = -dR$, $R_y = R$, $R_z = \infty$; where R is the radius of rotation from O; there follows from (76),

$$\left.\begin{aligned} a_{YY} &\equiv 2\left(\omega + \frac{u}{R}\right)\left(2\omega - \frac{\partial u}{\partial Y} + \frac{u}{R}\right) - \frac{\partial \Pi}{\partial Y}\frac{\partial \Theta}{\partial Y}, \\ a_{YZ} &\equiv -2\left(\omega + \frac{u}{R}\right)\frac{\partial u}{\partial Z} - \frac{\partial \Pi}{\partial Y}\frac{\partial \Theta}{\partial Z}, \\ a_{ZY} &\equiv -\frac{\partial \Pi}{\partial Z}\frac{\partial \Theta}{\partial Y}, \qquad a_{ZZ} \equiv -\frac{\partial \Pi}{\partial Z}\frac{\partial \Theta}{\partial Z}, \end{aligned}\right\} \tag{79}$$

where $a_{YZ} \equiv a_{ZY}$, and $u \equiv u(R, Z)$ represents the zonal velocity of the air. The discriminant a reduces here to

$$\begin{aligned} a &\equiv a_{YZ}^2 - a_{YY}a_{ZZ} \\ &\equiv 2\left(\omega + \frac{u}{R}\right)\frac{g_z}{\Theta}\frac{\partial \Theta}{\partial Z}\left[\frac{\delta u}{\delta Y} - 2\omega - \frac{u}{R}\right], \end{aligned} \tag{80}$$

where g_z is the absolute value of the component of gravity parallel to the earth's axis.

Since the inequality $\partial\Theta/\partial Z > 0$ generally holds in the atmosphere, the permanent circular vortex is stable or unstable depending on whether

$$\frac{\delta u}{\delta Y} < \text{or} > 2\omega + \frac{u}{R}. \tag{81}$$

The circulation along the parallels being cyclonic or anticyclonic as $u >$ or < 0, we observe that anticyclonic circulation favors the instability of the vortex. Consequently an easterly zonal circulation will become unstable more easily than a westerly one. Since the zonal circulation is easterly in tropical latitudes and westerly in temperate latitudes, and the Coriolis parameter increases with latitude, it is evident from (78) that the zonal circulation in tropical regions can reach the threshold of instability more easily than that of temperate regions. However, inasmuch as the isentropic surfaces in tropical latitudes are nearly horizontal, stabilization of the zonal circulation is favored there.

Finally it is easy to show, using criteria (81), that the permanent circular vortex is stable or unstable depending on whether its absolute angular momentum $\mathfrak{M} \equiv R(u + R\omega)$ increases ($\delta\mathfrak{M}/\delta R > 0$) or decreases ($\delta\mathfrak{M}/\delta R < 0$) toward the exterior of the vortex in the isentropic surfaces [2].

The state of neutral equilibrium ($\delta\mathfrak{M}/\delta R = 0$), characterized by the invariance of absolute angular momentum $\mathfrak{M}$ in the isentropic surfaces, would be the state of exchange equilibrium (*Austauschgleichgewicht*) according to Raethjen [27] and Kleinschmidt [18]. In other words it is that equilibrium condition toward which atmospheric air would tend if subjected to isentropic lateral mixing, Raethjen's *Gleitaustausch*, or Rossby's *lateral mixing*. If this condition prevailed, the lateral Reynolds stress T_x^y due to isentropic turbulence would be proportional to $\delta u/\delta y - f$ (neglecting the curvature term in (78)), where x represents the eastward tangent to the parallel and y the northward tangent to the meridian. However, this is true only if the isentropic turbulence is purely transverse. In reality the turbulent particles move in the isentropic surfaces parallel to the mean flow as well as perpendicular to it. Priebsch [25] has been able to show that the stress T_x^y is actually proportional to $\delta u/\delta y$ and *not* to $\delta u/\delta y - f$. As a result, the absolute angular momentum relative to the earth's axis of rotation is not an invariant property of isentropic exchange of air [44].

We observe that the permanent circular vortex satisfies an extremum principle [38] identical to that stated on page 436.

Moreover, the method we followed in studying the inertial oscillations of a geostrophic current (pp. 032–040) will naturally extend itself to cover the case of a stationary circular vortex [31]. We thus obtain the results which we had established earlier for the case of zonal geostrophic motion ($\omega_x \equiv 0$ or $\alpha \equiv 0, \pi$). Similarly, the equations of adiabatic perturbations of a permanent circular vortex can be derived in the same way as those for a zonal geostrophic current (pp. 444–446) provided cylindrical or spherical coordinates are used. This problem has recently been treated in terms of cylindrical coordinates, in a slightly different fashion, by Queney [26], who has in addition proposed a method for the simplification of the equations which differs from Charney's [4].

Let us further note that Fjørtoft [10] has recently taken up the study of the stability of stationary circular vortices with the aid of a new method based on the primary integral of the equations of motion (integral of the equation of mechanical energy) and utilizing the calculus of variations.

Finally, we consider a permanent circular vortex whose vertical axis z is located at latitude ϕ. If we designate by u the linear velocity of the vortex, $u > 0$

or < 0 depending on whether the vortex is cyclonic or anticyclonic relative to the earth. We shall assume that the vertical distribution of mass is stable. Then the vortex is unstable at any point for a transverse isentropic displacement from this point when, (80) and (81),

$$\frac{\delta u}{\delta R} + \frac{u}{R} + f < 0, \tag{82}$$

since the vorticity in the atmosphere is always cyclonic in a fixed, geocentric reference system ($u/R + f/2 > 0$, see equation (80) where $f/2$ must be substituted for ω). This being the case, to an isentropic displacement of radial component ΔR there corresponds an azimuthal motion of the displaced particle, whose velocity is given by, (7),

$$V_x = u - \left(f + \frac{\delta u}{\delta R} + \frac{u}{R}\right)\Delta R.$$

Here V_x denotes the particle's velocity along the parallel associated with the vortex. Consequently the cyclonic circulation, relative to the earth, of air particles subjected to a radial isentropic displacement toward the exterior of the vortex ($\Delta R > 0$), decreases or increases depending on whether the vortex is stable or unstable. It follows that *when the air particles in a vertically oriented, unstable, circular vortex undergo an outward radial, isentropic displacement, they release instability energy which reinforces their cyclonic circulation around the vertical axis of the vortex.* According to Sawyer [30], this mechanism plays an important role in the formation of tropical cyclones. The probable truth of this opinion is enhanced by the fact that at low latitudes the Coriolis parameter f is relatively small (10^{-5} sec^{-1}) so that condition (82) is more easily fulfilled there than in middle latitudes.

Inertial Oscillations of an Arbitrary Flow

Ertel [7] has given the equation of circular frequencies of the inertial oscillations of an air parcel in arbitrary motion. This more general problem has recently been taken up by W. L. Godson [11].

If the atmosphere is referred to a right-handed Cartesian system, $x^1x^2x^3$, chosen at random but at rest with respect to the earth, the equation of motion and the equation of continuity are, respectively,

$$\frac{d^2x^i}{dt^2} + 2\omega_{ki}\frac{dx^k}{dt} = -\Theta\frac{\partial\Pi}{\partial x^i} - \frac{\partial\Phi}{\partial x^i} = -S\frac{\partial P}{\partial x^i} - \frac{\partial\Phi}{\partial x^i} \quad (i, k = 1, 2, 3) \tag{83}$$

and

$$\frac{d}{dt}\left(\frac{\delta' x^1 \delta' x^2 \delta' x^3}{S}\right) = 0 \qquad \text{with } \delta' t \equiv 0,$$

in which $\omega_{ki} \equiv -\omega_{ik}$, are the antisymmetrical components of the rotation ω of the earth [41], and the repeated index k should be considered as a summation index (dummy index convention). In the general case considered here, S, Θ, P, Π are functions of x^i and of t, and Φ is a function of x^i. As in the case of geostrophic motion (see p. 435), let us associate to the equilibrium position $O(x_0^1, x_0^2, x_0^3)$ which a particle occupies at time t_0, the perturbed position

$$A(x^1 + \Delta x^1, x^2 + \Delta x^2, x^3 + \Delta x^3)$$

which this particle occupies at the later instant $t > t_0$ after receiving an impulse at O at time t_0. The unperturbed particle will occupy position $P(x^1, x^2, x^3)$ at time t.

The coordinate components Δx^i of the displacement $\mathbf{r} \equiv \overrightarrow{PA}$ should satisfy the variation equations deduced from (83), that is,

$$\frac{d^2\Delta x^i}{dt^2} + 2\omega_{ki}\frac{d\Delta x^k}{dt} = -\Delta\Theta\frac{\partial\Pi}{\partial x^i} - \Theta\,\Delta\frac{\partial\Pi}{\partial x^i} - \Delta\frac{\partial\Phi}{\partial x^i}, \tag{84}$$

with

$$\frac{\Delta S}{S} = \frac{\delta'\Delta x^i}{\delta' x^i} \equiv \operatorname{div} \mathbf{r}.$$

Let us recall that d/dt represents the individual total derivative following the motion of air with respect to the earth. The variation Δ applied to any scalar quantity can be broken down into a local variation Δ_l at P and a convective variation $\Delta x^k \frac{\partial}{\partial x^k}$ along $\overrightarrow{PA}$; we thus have $\Delta \equiv \Delta_l + \Delta x^k \frac{\partial}{\partial x^k}$. Assuming the perturbation to be adiabatic ($\Delta\Theta \equiv 0$) and assuming an absence of local variation in the geopotential Φ, equation (84) reduces to

$$L_{ik}(\Delta x^k) = -\Theta\frac{\partial\Delta_l\Pi}{\partial x^i}, \tag{85}$$

provided we set

$$L_{ik} \equiv \gamma_{ik}\frac{d^2}{dt^2} - 2\omega_{ik}\frac{d}{dt} + \sigma_{ik},$$

in which γ_{ik} ($\equiv 0$ or 1 depending on whether $i \neq k$ or $i \equiv k$) are the coefficients of the metric form in rectangular coordinates x^1, x^2, x^3 and

$$\sigma_{ik} \equiv \Theta\frac{\partial^2\Pi}{\partial x^i\partial x^k} + \frac{\partial^2\Phi}{\partial x^i\partial x^k} \equiv S\frac{\partial^2 P}{\partial x^i\partial x^k} - \frac{S^2}{C^2}\frac{\partial P}{\partial x^i}\frac{\partial P}{\partial x^k} + \frac{\partial^2\Phi}{\partial x^i\partial x^k} \equiv \sigma_{ki}$$

are the components of Ertel's stability tensor [7]. If we consider the perturbation to be small, we may consider the coefficients of the differential system (85) as being constants whose numerical values are those which these coefficients take at the point P at the instant t.

In general, we will designate by α_{ij} the elements of the determinant of the third order $\alpha \equiv \| \alpha_{ij} \|$ and by α^{ij} the algebraic minor of α relative to the element α_{ij}; we will then have $\alpha^{ij}\alpha_{ik} \equiv 0$ or α depending on whether $j \neq k$ or $j \equiv k$. With this in mind, let us now multiply the two terms of (85) by L^{ik} and perform the

summation with respect to the index i; we then have the normal differential system

$$L(\Delta x^k) = -L^{ik}\left(\Theta \frac{\partial \Delta_l \Pi}{\partial x^i}\right). \tag{86}$$

The general integral of the nonhomogeneous linear system (86) is composed additively of a particular integral of this system and the general integral of the homogeneous differential system

$$L(\Delta x^k) = 0, \qquad (k = 1, 2, 3) \tag{87}$$

in which

$$L \equiv \frac{d^6}{dt^6} + (\gamma^{ij}\sigma_{ij} + \gamma_{ij}f^i f^j)\frac{d^4}{dt^4} + (\gamma_{ij}\sigma^{ij} + \sigma_{ij}f^i f^j)\frac{d^2}{dt^2} + \sigma \qquad (i, j = 1, 2, 3)$$

with $f^k \equiv 2\omega_{ij}$, the index series (k, i, j) being derived from the series (1, 2, 3) by a rotating permutation. The f^k terms are nothing but the Coriolis parameters associated with the coordinate system $x^1x^2x^3$. Let us note that the elongations Δx^i of an inertial oscillation ($\Delta_l \Pi \equiv 0$) of the air particle under consideration satisfy the differential system (87). The characteristic equation

$$\nu_0^6 - \beta\nu_0^4 + \epsilon\nu_0^2 - \sigma = 0 \tag{88}$$

of this system is obtained in the same way that the characteristic equation (49) is obtained from the differential system (48). In equation (88) we have set

$$\beta \equiv \gamma^{ij}\sigma_{ij} + \gamma_{ij}f^i f^j \quad \text{and} \quad \epsilon \equiv \gamma_{ij}\sigma^{ij} + \sigma_{ij}f^i f^j.$$

Since equation (88) is cubic with respect to ν_0^2, it follows that in the atmosphere *a particle set into any type of motion can undergo, in general, three inertial oscillations of different periods* [7].

Let us now attempt to determine the three roots, ν_0^2, of equation (88). In order to do this let us substitute for the coordinate system $x^1x^2x^3$ the system xyz whose z-axis coincides with the zenith direction at the point P; let us then adopt the approximation allowed (p. 438) and, finally, let us introduce the geostrophic flow $\mathbf{u}(u_x, u_y, 0)$ associated with the arbitrary motion (83); then, by definition, we have

$$u_x \equiv -\frac{\Theta}{f}\frac{\partial \Pi}{\partial y}, \qquad u_y \equiv \frac{\Theta}{f}\frac{\partial \Pi}{\partial x}, \qquad u_z \equiv 0. \tag{89}$$

We will admit moreover that

$$\Theta \frac{\partial \Pi}{\partial z} + g = 0. \tag{90}$$

In (89) and (90) we have set $x \equiv x^1$, $y \equiv x^2$, $z \equiv x^3$, and $f^3 \equiv f \equiv 2\omega \sin\phi$ and $f^2 \equiv f^1 \cong 0$. To the approximation to which we have just consented, the equation of motion (83, $i \equiv 3$) in the vertical, $z \equiv x^3$, has been replaced by the equation of hydrostatic balance (90). This simplification is amply justified in the atmosphere. Let us note that in the system xyz, it is possible to consider the second derivatives of the geopotential Φ as negligible. Substituting (89) and (90) into the σ_{ij} expression, we obtain

$$\sigma \equiv \| \sigma_{ik} \| \equiv \begin{vmatrix} f\Theta \frac{\partial}{\partial x}\left(\frac{u_y}{\Theta}\right) & -f\Theta \frac{\partial}{\partial x}\left(\frac{u_x}{\Theta}\right) & \frac{g}{\Theta}\frac{\partial \Theta}{\partial x} \\ f\Theta \frac{\partial}{\partial y}\left(\frac{u_y}{\Theta}\right) & -f\Theta \frac{\partial}{\partial y}\left(\frac{u_x}{\Theta}\right) & \frac{g}{\Theta}\frac{\partial \Theta}{\partial y} \\ f\Theta \frac{\partial}{\partial z}\left(\frac{u_y}{\Theta}\right) & -f\Theta \frac{\partial}{\partial z}\left(\frac{u_x}{\Theta}\right) & \frac{g}{\Theta}\frac{\partial \Theta}{\partial z} \end{vmatrix}$$

with the symmetry relations

$$\frac{\partial}{\partial x}\left(\frac{u_x}{\Theta}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\Theta}\right) = 0, \qquad f\Theta \frac{\partial}{\partial z}\left(\frac{u_x}{\Theta}\right) + \frac{g}{\Theta}\frac{\partial \Theta}{\partial y} = 0,$$

$$f\Theta \frac{\partial}{\partial z}\left(\frac{u_y}{\Theta}\right) - \frac{g}{\Theta}\frac{\partial \Theta}{\partial x} = 0,$$

and

$$\frac{\delta u_x}{\delta x} + \frac{\delta u_y}{\delta y} = 0, \tag{91}$$

in which we have set, (35),

$$\frac{\delta}{\delta x} \equiv \frac{\partial}{\partial x} - \left(\frac{\partial \Theta/\partial x}{\partial \Theta/\partial z}\right)\frac{\partial}{\partial z}, \qquad \frac{\delta}{\delta y} \equiv \frac{\partial}{\partial y} - \left(\frac{\partial \Theta/\partial y}{\partial \Theta/\partial z}\right)\frac{\partial}{\partial z}.$$

From the relation (91) we are able to deduce the existence of a stream function for geostrophic flow within the isentropic surfaces (Θ = const).

It is easy to show that, (36),

$$\sigma \equiv f^2\nu_s^2\left(\frac{\delta u_x}{\delta x}\frac{\delta u_y}{\delta y} - \frac{\delta u_x}{\delta y}\frac{\delta u_y}{\delta x}\right)$$

$$\equiv -f^2\nu_s^2\left[\left(\frac{\delta u_x}{\delta x}\right)^2 + \frac{\delta u_x}{\delta y}\frac{\delta u_y}{\delta x}\right] \approx 10^{-22} \text{ to } 10^{-21} \text{ sec}^{-6}$$

and

$$\epsilon \equiv \nu_s^2\nu_d^2 + f^2\Theta^2\frac{\partial(u_x/\Theta, u_y/\Theta)}{\partial(x, y)} \cong \nu_s^2\nu_d^2 \approx 10^{-12} \text{ sec}^{-4}$$

with

$$\nu_d^2 \equiv f\left[f + \left(\frac{\delta u_y}{\delta x} - \frac{\delta u_x}{\delta y}\right)\right] \approx 10^{-8} \text{ sec}^{-2},$$

in which $\delta u_y/\delta x - \delta u_x/\delta y$ represents the vertical component of vorticity determined in the isentropic surfaces (Θ = const) by the geostrophic current $\mathbf{u}(u_x, u_y, 0)$. Finally, we obtain

$$\beta \equiv \nu_s^2 + \nu_i^2 - \frac{f}{\Theta}\left(\frac{\partial \Theta}{\partial x}u_y - \frac{\partial \Theta}{\partial y}u_x\right) \cong \nu_s^2 \approx 10^{-4} \text{ sec}^{-2},$$

in which

$$\nu_i^2 \equiv f\left[f + \left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y}\right)\right] \approx 10^{-8} \text{ sec}^{-2}.$$

The expressions ν_i^2 and ν_d^2 defined above generalize the corresponding expressions (37) and (40), respectively, which are valid in the case of a particle in geostrophic motion.

In the general case, as in the geostrophic case, we also have

$$\nu_d^2 \cong \nu_i^2 - \frac{N^2}{\nu_s^2}$$

in which

$$N^2 \equiv N_x^2 + N_y^2,$$

with

$$N_x \equiv f\frac{\partial u_x}{\partial z} \cong -\frac{g}{\Theta}\frac{\partial \Theta}{\partial y},$$

$$N_y \equiv f\frac{\partial u_y}{\partial z} \cong +\frac{g}{\Theta}\frac{\partial \Theta}{\partial x},$$

together define the baroclinity at point P.

In order to enable us to interpret the formulas conveniently, let us orient the horizontal axis x of the xyz system along the tangent to the isobar at P in the direction of the geostrophic flow $\mathbf{u}(u_x, u_y, 0)$; we then have [11]:

$\frac{\delta u_x}{\delta y}$, the transversely isentropic geostrophic wind shear, (perpendicular to the streamlines of geostrophic flow within the isentropic surfaces); this wind shear is cyclonic or anticyclonic depending on whether $\delta u_x/\delta y < 0$ or > 0;

$\frac{\delta u_y}{\delta x}$ $>$ or $<$ 0 depending on whether the curvature of streamlines of geostrophic motion in the isentropic surfaces is cyclonic or anticyclonic; and

$\frac{\delta u_x}{\delta x}$ $>$ or $<$ 0 depending on whether these streamlines are converging or diverging.

It follows from the order of magnitude and approximate values of the coefficients β, ϵ, and σ of equation (88), as well as from their classical expression as functions of the roots, that the largest root of this equation has a value in the neighborhood of ν_s^2. Dividing the first member of equation (88) by $\nu_0^2 - \nu_s^2$, we obtain Godson's approximate equation [11]

$$\nu_0^4 - \nu_d^2\nu_0^2 + C = 0 \qquad (92)$$

in which

$$C \equiv \sigma/\nu_s^2 \equiv -f^2\left[\left(\frac{\delta u_x}{\delta x}\right)^2 + \frac{\delta u_x}{\delta y}\frac{\delta u_y}{\delta x}\right] \approx 10^{-18} \text{ to } 10^{-17} \text{ sec}^{-4}.$$

If we refer to the equations of the inertial oscillations, derived from (85), it is apparent that the inertial oscillation corresponding to the root ν_s^2 is quasi-vertical (static stability) and that those corresponding to the two roots ν_0^2 of (92) are quasi-isentropic [11].

The nature of the roots of the biquadratic equation (92) depends on the sign of the discriminant

$$\nu_d^4 - 4C \equiv f^2\left[f^2 + \left(\frac{\delta u_y}{\delta x} + \frac{\delta u_x}{\delta y}\right)^2 + 4\left(\frac{\delta u_x}{\delta x}\right)^2 + 2f\left(\frac{\delta u_y}{\delta x} - \frac{\delta u_x}{\delta y}\right)\right].$$

In general, equation (92) admits of two real roots ($4C < \nu_d^4$), whose approximate values are ν_d^2 and C/ν_d^2. There are four cases to be considered:

I. $\nu_d^2 > 0$, $C > 0$: hydrodynamic stability regardless of the particular isentropic displacement at point P.

II. $\nu_d^2 < 0$, $C > 0$: hydrodynamic instability regardless of the particular displacement at point P.

III. $\left.\begin{matrix}\nu_d^2 < 0,\ C < 0\\ \nu_d^2 > 0,\ C < 0\end{matrix}\right\}$: hydrodynamic instability at P.

It should be noted that C can be positive only if $(\delta u_x/\delta y)(\delta u_y/\delta x) < 0$. In this case, a transversally isentropic, cyclonic shear (anticyclonic shear) of the geostrophic wind will necessarily correspond to a cyclonic (anticyclonic) curvature of the streamlines of geostrophic motion within the isentropic surface at P. The case of cyclonic geostrophic wind shear invariably corresponds to the stable type I (marked stability); the case of anticyclonic geostrophic wind shear corresponds to the unstable type II or to the stable type I (weak stability), but only if the absolute value of the anticyclonic vorticity $\delta u_y/\delta x - \delta u_x/\delta y < 0$ is smaller than the Coriolis parameter f.

Furthermore, C is necessarily negative if

$$(\delta u_x/\delta y)(\delta u_y/\delta x) > 0.$$

In this case, a transversally isentropic, anticyclonic (cyclonic) shear of the geostrophic wind will correspond to a cyclonic (anticyclonic) curvature of the streamlines of geostrophic motion in the isentropic surface at P.

The sign of the transversally isentropic geostrophic wind shear and the sign of the curvature of streamlines of geostrophic motion in isentropic surfaces both play preponderant roles; the divergence or convergence of these lines of motion, on the other hand, are without influence over the conditions leading to hydrodynamic stability or instability—only the absolute value $|\delta u_x/\delta x| \equiv |\delta u_y/\delta y|$ is decisive.

The biquadratic equation (92) admits of two conjugate imaginary roots when:

IV. $4C - \nu_d^4 > 0$: hydrodynamic instability regardless of the isentropic displacement which takes place at point P. This case can arise only when ν_d^2 is of an order of magnitude less than 10^{-8} sec^{-2}, that is, when the vorticity $\delta u_y/\delta x - \delta u_x/\delta y$ is anticyclonic, and when its absolute value is of the same order of magnitude as the Coriolis parameter f. The existence in the general case of a second type of total instability in the isentropic surfaces is thus not excluded. Let us note that the first type (case II) is none other than the one we met previously in our discussion of geostrophic motion (case II of Table II) and as a consequence the results obtained in the geostrophic case can be extended to the general case of any given motion.

Thus, the condition $\nu_d^2 < 0$ is a sufficient condition for instability, but is no longer (as in the geostrophic case) a necessary and sufficient condition. On the other hand, the condition $\nu_d^2 > 0$ is a necessary condition of stability but is no longer (as in the geostrophic case) a necessary and sufficient condition [11]. Nevertheless, it

can be said, in a general way, that the hydrodynamic stability increases or decreases depending on whether ν_a^2 increases or decreases. It is the instability corresponding to the weak positive values and to the negative values of ν_a^2 which seems to be of the greatest practical importance [12].

Godson has not only calculated the inertial orbits in the four cases discussed above [11] but has also made a statistical and synoptic study of the four corresponding types of stability and instability [12].

The hydrodynamic instability which accompanies the deepening of frontal cyclones appears in the atmosphere wherever the baroclinity is strongest and the static stability, simultaneously, is weakest (see pp. 438–440), that is, in the tropical air above the polar front zone. It is particularly with reference to the importance of hydrodynamic instability in the formation of cyclonic circulation that new synoptic studies, in greater number and on a larger scale, appear to be desirable.

REFERENCES

1. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1:1–22 (1944).
2. Bjerknes, V., and others, *Hydrodynamique physique.* Paris, Presses Universitaires de France, 1934.
3. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4:135–162 (1947).
4. —— "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2, 17 pp. (1948).
5. Ertel, H., "Über die Stabilität der zonalen atmosphärischen Zirkulation." *Meteor. Z.*, 57:397–400 (1940).
6. —— "Die Westwindgebiete der Troposphäre als Instabilitätszonen." *Meteor. Z.*, 60:397–400 (1943).
7. —— Jaw, J.-J., und Li, S.-z., "Tensorielle Theorie der Stabilität." *Meteor. Z.*, 58:389–392 (1941).
8. Fjørtoft, R., "On the Deepening of a Polar Front Cyclone." *Meteor. Ann.*, 1:30–44 (1942).
9. —— "On the Frontogenesis and Cyclogenesis in the Atmosphere. Part I—On the Stability of the Stationary Circular Vortex." *Geofys. Publ.*, Vol. 16, No. 5, 28 pp. (1946).
10. —— "Application of Integral Theorems in Deriving Criteria of Stability for Laminar Flows and for the Baroclinic Circular Vortex." *Geofys. Publ.*, Vol. 17, No. 6, 52 pp. (1950).
11. Godson, W. L., "Generalized Criteria for Dynamic Instability." *J. Meteor.*, 7:268–278 (1950).
12. —— "Synoptic Significance of Dynamic Instability." *J. Meteor.*, 7:333–342 (1950).
13. Helmholtz, H. v., "Über atmosphärische Bewegungen." *Meteor. Z.*, 5:329–340 (1888).
14. Høiland, E., "On the Interpretation and Application of the Circulation Theorems of V. Bjerknes." *Arch. Math. Naturv.*, 42: 25–57 (1939).
15. —— "On the Stability of the Circular Vortex." *Avh. norske VidenskAkad.*, Ser. 1, No. 11, 24 pp. (1941).
16. Holmboe, J., "On Dynamic Stability of Zonal Currents." *J. mar. Res.*, 7:163–174 (1948).
17. Kleinschmidt, E., "Zur Theorie der labilen Anordnung." *Meteor. Z.*, 58:157–163 (1941).
18. —— "Stabilitätstheorie des geostrophischen Windfeldes." *Ann. Hydrogr., Berl.*, 69:305–325 (1941).
19. Kuo, H.-L., "Dynamic Instability of Two-Dimensional Nondivergent Flow in a Barotropic Atmosphere." *J. Meteor.*, 6:105–122 (1949).
20. Möller, F., "Der Zusammenhang statischer und dynamischer Labilität nach E. Kleinschmidt." *Meteor. Z.*, 60:269–273 (1943).
21. Moltschanow, P., "Bedingungen des Gleichgewichts und der Stabilität der Luftmassen nach der Horizontalen und der Vertikalen." *Petermanns Mitt.*, Ergänzungsband 47, Heft 216, SS. 62–67 (1933).
22. Palmén, E., "On the Distribution of Temperature and Wind in the Upper Westerlies." *J. Meteor.*, 5:20–27 (1948).
23. —— and Nagler, K. M., "An Analysis of the Wind and Temperature Distribution in the Free Atmosphere over North America in a Case of Approximately Westerly Flow." *J. Meteor.*, 5:58–64 (1948).
24. Palmén, E., and Newton, C. W., "A Study of the Mean Wind and Temperature Distribution in the Vicinity of the Polar Front in Winter." *J. Meteor.*, 5:220–226 (1948).
25. Priebsch, J., "Eine Bemerkung zur Dynamik turbulenter Strömungen unter dem Einfluss der Erdrotation." *Ann. Hydrogr., Berl.*, 71:169–171 (1943).
26. Queney, P., "Adiabatic Perturbation Equations for a Zonal Atmospheric Current." *Tellus*, Vol. 2, No. 1, pp. 35–51 (1950).
27. Raethjen, P., "Labile Gleitumlagerungen." *Ann. Hydrogr., Berl.*, 69:325–331 (1941).
28. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
29. Rossby, C.-G., Namias, J., and Simmers, R. G., "Fluid Mechanics Applied to the Study of Atmospheric Circulations." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 1, pp. 1–125 (1938).
30. Sawyer, J. S., "Notes on the Theory of Tropical Cyclones." *Quart. J. R. meteor. Soc.*, 73:101–126 (1947).
31. Solberg, H., "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones." *P. V. Météor. Un. géod. géophys. int.* Edimbourg, septembre 1936, II:66–82 (1939).
32. University of Chicago, Dept. Meteor., "On the General Circulation of the Atmosphere in Middle Latitudes." *Bull. Amer. meteor. Soc.*, 28:255–280 (1947).
33. Van Mieghem, J., "Forme intrinsèque du critère d'instabilité dynamique de E. Kleinschmidt." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 30:19–33 (1944).
34. —— "Relation d'identité entre la stabilité de l'équilibre dynamique de E. Kleinschmidt et la stabilité des oscillations d'inertie de l'atmosphère terrestre." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 30:134–143 (1944).
35. —— "Interprétations énergétiques du critère d'instabilité de E. Kleinschmidt." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 31:345–352 (1945).
36. —— "Les oscillations d'inertie du courant géostrophique." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 31:547–555 (1945).
37. —— "Sur la stabilité du courant géostrophique." *La Météor.*, pp. 9–33 (1946).
38. —— "Le principe d'extremum et la stabilité de certains états de mouvement de l'air atmosphérique." *Arch. Meteor. Geophys. Biokl.*, (A) 1:347–357 (1949).
39. —— "Contribution à l'étude de la cyclogénèse." *Inst. R. météor. Belg., Mém.*, 23:1–23 (1946).

40. —— "L'instabilité hydrodynamique et les perturbations du courant zonal d'Ouest." *Arch. Meteor. Geophys. Biokl.*, (A) 1:143–148 (1949).
41. —— "La stabilité du mouvement permanent, horizontal et isobare de l'air atmosphérique." *Inst. R. météor. Belg., Mém.*, 28:38–60 (1948).
42. —— "Perturbations d'un courant atmosphérique permanent zonal." *Inst. R. météor. Belg., Mém.*, 18:1–23 (1944).
43. —— "L'équation aux dérivées partielles de la pression de perturbation associée aux ondulations de grande longueur d'onde du courant géostrophique zonal." *Inst. R. météor. Belg., Mém.*, Vol. 39, 45 pp. (1950).
44. —— "Zijdelingse turbulentie in de atmosfeer." *Med. K. Vl. Acad. België*, XII, No. 14, 16 pp. (1950).
45. Wexler, H., and Namias, J., "Mean Monthly Isentropic Charts and Their Relation to Departures of Summer Rainfall." *Trans. Amer. geophys. Un.*, 19:164–170 (1938).
46. Wippermann, F., "Über die Rolle der dynamischen Labilität bei der Zyklogenese." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 12, SS. 180–182 (1950).

STABILITY PROPERTIES OF LARGE-SCALE ATMOSPHERIC DISTURBANCES

By R. FJØRTOFT

*The Institute for Advanced Study**

Introduction

The large-scale motion of the earth's atmosphere is to the first approximation a solid rotation from west to east. Upon this are superimposed a more or less orderly zonal flow in the relative motion and the large-scale disturbances familiar in meteorology. In this article causes for the creation and maintenance of these disturbances will be discussed, excluding, however, the more or less permanent disturbances forced upon the atmosphere because of the earth's topographic inhomogeneity.

By the use of the phrase "disturbances in a zonal flow," a separation is implied between two components of the flow. This may seem artificial since the hydrodynamic equations are directly applicable to the total component of the flow. There are, however, several reasons for doing this: The immediate impression obtained by studying hemispherical weather maps is one of a more or less orderly zonal flow upon which are superimposed disturbances that behave to some degree as physical entities themselves. Further, the zonal flow and the disturbances undergo somewhat systematic changes, seemingly of great importance to weather. Therefore, by a separation of the atmospheric flow into some kind of orderly zonal flow and disturbances superimposed thereupon, one isolates at the outset certain phenomena which appear to be related to actual weather. Besides, this separation will enable one to deal satisfactorily with problems in which a detailed knowledge of the motion is unnecessary, as exemplified by the stability investigations carried out in the discussion of barotropic disturbances (pp. 460–463).

Granted that such a separation of the atmospheric flow may prove useful, it becomes important for quantitative treatment to express this separation in mathematical terms. How this should be done is to some extent arbitrary because there may be several ways of defining the orderly flow, and thereby the disturbances. In this article, the orderly flow will be characterized, for an arbitrary hydrodynamic element α, by the mean value of α at a fixed time along latitude circles:

$$\bar{\alpha} = \frac{1}{2\pi}\int_0^{2\pi} \alpha \, d\psi,$$

while the corresponding element in the flow of disturbances will be defined by

$$\alpha' = \alpha - \bar{\alpha}.$$

The flow therefore is composed of an orderly, axially symmetric motion and an irregular flow vanishing in the mean along zonal circles. The degree of irregularity may, however, vary widely. In this article it is assumed that all irregularities considerably smaller than the smallest-scale cyclones have already been smoothed out in some way. The corresponding turbulent stresses will be entirely neglected throughout this article.

As already pointed out above, the sum of the two components of flow must obey the hydrodynamic equations of motion. There must therefore exist a coupling between these two components. Actually, in many cases this coupling is so strong that a full understanding of what happens with one component cannot be achieved without taking into account simultaneous changes of the other.

It is an established procedure to separate the hydrodynamic equations into one set valid for the orderly motion and one for the irregular flow. To implement this process the equations will first be simplified by suppressing certain terms of minor importance. With the conventional simplification in the Coriolis acceleration, the equation of motion is

$$\rho\left[\frac{D\mathbf{v}}{dt} + f\mathbf{k}\times\mathbf{v} - \mathbf{g}\right] = -\nabla p,$$

where ρ is the density, f is the Coriolis parameter, $\mathbf{g}$ is the acceleration of gravity, $\mathbf{v}$ is the velocity, and p is the pressure. The coordinate system has been selected so that the x-axis is directed east, the y-axis north, and the z-axis upward; $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the x, y, and z directions, respectively. By elimination of ∇p and substitution of $\varkappa = \ln\vartheta$ (where ϑ is the potential temperature) by means of the relationship $\nabla\rho - \Gamma\nabla p = -\rho\nabla\varkappa$, one obtains

$$\nabla\times\left[\frac{D\mathbf{v}}{dt} + f\mathbf{k}\times\mathbf{v}\right] = -\nabla\times\varkappa\mathbf{g}. \qquad (1)$$

The neglected term $-\nabla\varkappa\times[D\mathbf{v}/dt + f\mathbf{k}\times\mathbf{v}]$ is small compared with the others. Equation (1) is equivalent to

$$\frac{D\mathbf{v}}{dt} = -\nabla\gamma - f\mathbf{k}\times\mathbf{v} - \varkappa\mathbf{g}, \qquad (2)$$

where $\nabla\gamma$ is a certain laminar vector. It is easily understood that by introducing the simplification above, the effects from solenoids in horizontal planes have been neglected. Now let $Q(z)$ represent some standard distribution of density with height. The following approximate equation will then constitute the continuity equation:

$$\nabla\cdot Q\mathbf{v} = Q\nabla_h\cdot\mathbf{v}_h + \frac{\partial Qw}{\partial z} = 0. \qquad (3)$$

* This article is to some extent based upon work performed under contract N6-ori-139, Task Order I between the Office of Naval Research and The Institute for Advanced Study.

In this article only adiabatic processes will be considered. The physical equation is therefore

$$\frac{\partial \kappa}{\partial t} = -\mathbf{v}\cdot\nabla\kappa. \tag{4}$$

From the foregoing equations one obtains by separation into the mean and irregular flows:

$$\frac{\partial \bar{\mathbf{v}}}{\partial t} + \bar{\mathbf{v}}\cdot\nabla\bar{\mathbf{v}} = -\nabla\bar{\gamma} - f\mathbf{k}\times\bar{\mathbf{v}} - \bar{\kappa}\mathbf{g} - \overline{\mathbf{v}'\cdot\nabla\mathbf{v}'}, \tag{5a}$$

$$\nabla\cdot Q\bar{\mathbf{v}} = 0, \tag{5b}$$

$$\frac{\partial \bar{\kappa}}{\partial t} = -\bar{\mathbf{v}}\cdot\nabla\bar{\kappa} - \overline{\mathbf{v}'\cdot\nabla\kappa'}, \tag{5c}$$

and

$$\frac{\partial \mathbf{v}'}{\partial t} + \mathbf{v}'\cdot\nabla\mathbf{v}' = -\nabla\gamma' - f\mathbf{k}\times\mathbf{v}' - \kappa'\mathbf{g} - \bar{\mathbf{v}}\cdot\nabla\mathbf{v}' - \mathbf{v}'\cdot\nabla\bar{\mathbf{v}} + \overline{\mathbf{v}'\cdot\nabla\mathbf{v}'}, \tag{6a}$$

$$\nabla\cdot Q\mathbf{v}' = 0, \tag{6b}$$

$$\frac{\partial \kappa'}{\partial t} = -\bar{\mathbf{v}}\cdot\nabla\kappa' - \mathbf{v}'\cdot\nabla\bar{\kappa} + \overline{\mathbf{v}'\cdot\nabla\kappa'}. \tag{6c}$$

These equations reveal the coupling which must exist between the two components of the flow, as both components occur in each set of equations.

The Circular Vortex

According to the definitions above, the orderly flow may be considered as a pure zonal flow with velocity component $\bar{u}\mathbf{i}$, and a meridional flow with velocity $\bar{\mathbf{v}}_m = \bar{v}\mathbf{j} + \bar{w}\mathbf{k}$, which is identically the same in all meridional planes. The meridional component of (5) may be written

$$\frac{\overline{D\mathbf{v}_m}}{dt} = -\nabla\bar{\gamma} - \bar{\kappa}\mathbf{g} - f\mathbf{k}\times\bar{u}\mathbf{i}.$$

The approximate balance existing in the large-scale relative motion in the atmosphere gives, when $\nabla\bar{\gamma}$ is eliminated from this equation:

$$\frac{\partial \bar{u}}{\partial z} = -\frac{g}{\bar{f}}\frac{\partial \bar{\kappa}}{\partial y}. \tag{7}$$

This is the so-called thermal wind equation applied to the mean flow.

One may study the axially symmetric meridional motions in a qualitative way by the method of velocity circulation used primarily by V. Bjerknes [4] and Høiland [16]. If $\nabla\bar{\gamma}$ is eliminated from the foregoing equation by taking the circulation along some arbitrary closed curve in a meridional plane, and the resulting equation differentiated once with respect to time, one gets

$$\frac{d}{dt}\oint \frac{\overline{D\mathbf{v}_m}}{dt}\cdot\delta\mathbf{r} = -\oint \frac{\partial \bar{\kappa}}{\partial t}\,\mathbf{g}\cdot\delta\mathbf{r} - \oint f\frac{\partial \bar{u}}{\partial t}\,\mathbf{j}\cdot\delta\mathbf{r}. \tag{8}$$

The expression for $\partial\bar{u}/\partial t$ is obtained from the zonal component of (5a):

$$\frac{\partial \bar{u}}{\partial t} = -\bar{\mathbf{v}}_m\cdot(\nabla\bar{u} - f\mathbf{j}) - \overline{\mathbf{v}'\cdot\nabla u'}. \tag{9}$$

Substituting this expression for $\partial\bar{u}/\partial t$ in (8) and likewise for $\partial\bar{\kappa}/\partial t$ from (5c), one gets

$$\frac{d}{dt}\oint \frac{\overline{D\mathbf{v}_m}}{dt}\cdot\delta\mathbf{r} = \oint \bar{\mathbf{v}}_m\cdot[f\nabla\bar{u}\mathbf{j} - f^2\mathbf{j}\mathbf{j} + \nabla\bar{\kappa}\mathbf{g}]\cdot\delta\mathbf{r} + \oint\overline{\mathbf{v}'\cdot\nabla\kappa'}\mathbf{g}\cdot\delta\mathbf{r} + \oint f\overline{\mathbf{v}'\cdot\nabla u'}\mathbf{j}\cdot\delta\mathbf{r}. \tag{10}$$

A study of the first integral on the right-hand side of this equation leads to the conditions which must exist if in a pure, axially symmetric motion the meridional circulations should accelerate or decelerate, in other words to the now well-known stability criteria for a circular vortex for vortex-ring perturbations. It will be assumed in this article that all orderly flows which are treated are stable in this sense. Most likely this is usually the case in the atmosphere.

The remaining terms on the right-hand side of (10) represent the effects upon the acceleration of the meridional circulations which are due to the disturbances. The character of the resulting forced circulations will now also depend essentially upon the stability properties of the circular vortex for vortex-ring perturbations. Briefly, one may say that the presence of effects changing the fields of mass and velocity in the orderly zonal flow, other than effects of the meridional circulations themselves, will generally tend steadily to destroy the balance in the meridional motions. Because of the stability of the circular vortex the resulting added meridional circulations will act to restore the equilibrium, an equilibrium which will, however, be different from the original one. If the stability is large enough, the whole development may be thought of as one which goes through different equilibrium stages by smoothing out over sufficiently large periods the relatively high frequency oscillations superimposed upon this trend. The simplifications following from such a procedure are essentially the same as those introduced by the systematic use of the condition of quasi-geostrophic motion [7, 12]. With the simplifications above, A. Eliassen [11] has studied the forced meridional circulations produced by given sources of heat and angular momentum.

To see in a qualitative fashion how the irregular flow affects the mean meridional circulations one may use the simplifications mentioned above in connection with the circulation integrals in (10). One then obtains

$$\oint \bar{\mathbf{v}}_m\cdot[f\nabla\bar{u}\mathbf{j} - f^2\mathbf{j}\mathbf{j} + \nabla\bar{\kappa}\mathbf{g}]\cdot\delta\mathbf{r} = -\oint\overline{\mathbf{v}'\cdot\nabla\kappa'}\mathbf{g}\cdot\delta\mathbf{r} - \oint f\overline{\mathbf{v}'\cdot\nabla u'}\mathbf{j}\cdot\delta\mathbf{r}. \tag{11}$$

It will now be assumed that $\nabla\bar{u}$ and $\partial\bar{\kappa}/\partial y$ are small enough to be neglected where they occur in (9) and (11). It is further assumed that $\partial\bar{\kappa}/\partial z \geqq 0$, which is

necessary to insure the stability of the circular vortex in the present case. Equations (9) and (11) now reduce to

$$\frac{\partial \bar{u}}{\partial t} = f\bar{v} - \overline{\mathbf{v}'\cdot\nabla u'}, \tag{12}$$

$$-\oint f^2\bar{v}\,\delta y - \oint g\frac{\partial \bar{\kappa}}{\partial z}\bar{w}\,\delta z = \oint \overline{\mathbf{v}'\cdot\nabla\kappa'}g\,\delta z - \oint f\overline{\mathbf{v}'\cdot\nabla u'}\,\delta y. \tag{13}$$

Let it now be assumed that the path of integration consists of the sides of a "rectangle" bounded below by the earth's surface and with the top at about tropopause height (Fig. 1). The direction of integration is indicated by arrows. With subscripts from 1 to 4 denoting average values along the sides of the rectangle correspondingly labelled, (13) may be written

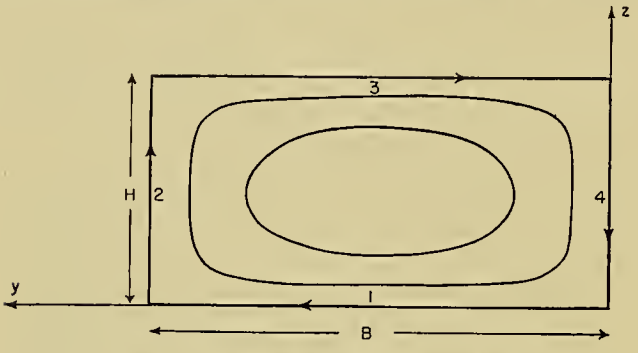


Fig. 1.—Idealized meridional circulation.

$$f^2(\bar{v}_3 - \bar{v}_1)B + g\frac{\partial \bar{\kappa}}{\partial z}(\bar{w}_2 - \bar{w}_4)H = g(\overline{\mathbf{v}'\cdot\nabla\kappa_2'} - \overline{\mathbf{v}'\cdot\nabla\kappa_4'})H + f(\overline{\mathbf{v}'\cdot\nabla u_3'} - \overline{\mathbf{v}'\cdot\nabla u_1'})B. \tag{14}$$

Here, B and H are the lengths of the sides of the rectangle. Suppose further that the rectangular-shaped boundary is a streamline in a corresponding simple cellular meridional motion, in which $\bar{v}$ and $\bar{w}$ are derived from a stream function

$$\psi_a \sim \sin\frac{\pi y}{B}\sin\frac{\pi z}{H}. \tag{15}$$

This implies that for the present the atmosphere is treated as incompressible. It may be anticipated, however, that the results to be obtained will also roughly apply to cellular motions whose kinematics are essentially the same as in this most simple cellular motion.

By deriving $\bar{v}$ and $\bar{w}$ from (15) and forming the averages $\bar{v}_1$, $\bar{v}_3$, $\bar{w}_2$, and $\bar{w}_4$, one obtains

$$(\bar{w}_4 - \bar{w}_2)B = (\bar{v}_3 - \bar{v}_1)H. \tag{16}$$

If one does not think of the velocities in this formula as averages along the sides of the rectangular boundary, but rather as averages over the whole region of the velocities in the ascending and descending motions, and in the north and south motions, the formula simply states the well-known continuity principle that the ratio between the magnitude of the vertical and horizontal velocities is proportional to the ratio between the vertical and horizontal scale of the motion. By a suitable choice of the integration curve in the circulation integrals given above, one could therefore also apply (16) to cellular motions of a much more general character than the one determined from (15). In view of (16), (14) may now be written

$$f(\bar{v}_3 - \bar{v}_1)\left(f^2 + g\frac{\partial \bar{\kappa}}{\partial z}\frac{H^2}{B^2}\right) = fg(\overline{\mathbf{v}'\cdot\nabla\kappa_2'} - \overline{\mathbf{v}'\cdot\nabla\kappa_4'})\frac{H}{B} + f^2(\overline{\mathbf{v}'\cdot\nabla u_3'} - \overline{\mathbf{v}'\cdot\nabla u_1'}). \tag{17}$$

Taking likewise an average of (12) along the horizontal sides of the rectangle, one obtains by subtraction

$$\frac{\partial}{\partial t}(\bar{u}_3 - \bar{u}_1) = f(\bar{v}_3 - \bar{v}_1) - (\overline{\mathbf{v}'\cdot\nabla u_3'} - \overline{\mathbf{v}'\cdot\nabla u_1'}).$$

Substituting here for $f(\bar{v}_3 - \bar{v}_1)$ from (17), one obtains

$$\frac{\partial}{\partial t}\left(\frac{\bar{u}_3 - \bar{u}_1}{H}\right) = \left[\frac{g\dfrac{\partial \bar{\kappa}}{\partial z}\dfrac{H^2}{B^2}}{f^2 + g\dfrac{\partial \bar{\kappa}}{\partial z}\dfrac{H^2}{B^2}}\right]\left[\frac{\overline{\mathbf{v}'\cdot\nabla u_1'} - \overline{\mathbf{v}'\cdot\nabla u_3'}}{H}\right] + \left[\frac{f^2}{f^2 + g\dfrac{\partial \bar{\kappa}}{\partial z}\dfrac{H^2}{B^2}}\right]\left[\frac{g}{Bf}(\overline{\mathbf{v}'\cdot\nabla\kappa_2'} - \overline{\mathbf{v}'\cdot\nabla\kappa_4'})\right]. \tag{18}$$

This formula now determines, as a function of two terms, the time rate of change in the vertical wind shear of the mean zonal flow: The first term involves the wind shear that would directly result from the dynamic effects of the irregular motion; the second, the mean meridional temperature gradient resulting directly from the thermal effects of the disturbances. However, it is seen that only fractions of these quantities are effective in building up the resulting shear since they are multiplied by factors smaller than unity. This is clearly a result of the interference with the effects resulting from the forced meridional circulations, and could have been obtained by more direct considerations. The formula given above, however, may serve as a rough indicator of how this interference depends upon the stability of the circular vortex and the horizontal and vertical scales of the motion. In the present discussion it will only be pointed out how the dynamically conditioned increase in the wind shear becomes more and more compensated when the vertical stability goes to zero, or when H/B becomes smaller, while at the same time the thermally conditioned increase in wind shear becomes correspondingly more important. The relative importance, in regard to the general circulation, of the dynamic effects of the disturbances on the one hand, and the thermal effects on the other, naturally depends also upon the relative magnitudes of the terms $\overline{\mathbf{v}'\cdot\nabla u'}$ and $\overline{\mathbf{v}'\cdot\nabla\kappa'}$ and their distribution. This is intimately connected with the problems to be treated in the following sections.

Baroclinic Disturbances

One way of classifying atmospheric disturbances is with respect to the sources of energy which are at their disposal. The energy equation, when integrated over an isolated volume τ of the atmosphere, is obtained from (2) and (3) and becomes

$$\int Q\tfrac{1}{2}\mathbf{v}^2\,d\tau = \int Q\varkappa gz\,d\tau + \text{const.}$$

Substituting here $\mathbf{v} = \bar{u}\mathbf{i} + \bar{\mathbf{v}}_m + \mathbf{v}'$, one arrives at

$$\int Q\tfrac{1}{2}\mathbf{v}'^2\,d\tau = \int Q\varkappa gz\,d\tau - \int Q\tfrac{1}{2}\bar{u}^2\,d\tau + \text{const},$$

having neglected the relatively much smaller kinetic energy contained in the mean meridional circulations. Applying the same approximation, the time rate of change of this equation becomes, in view of (5) and (6),

$$\frac{d}{dt}\int Q\tfrac{1}{2}\mathbf{v}'^2\,d\tau = \int Q\varkappa' w' g\,d\tau - \int Q\frac{\partial\bar{u}}{\partial y}u'v'\,d\tau - \int Q\frac{\partial\bar{u}}{\partial z}u'w'\,d\tau. \tag{19}$$

Consequently, one may classify disturbances into three categories according to whether the main source of energy is potential, kinetic, or both. In this section the first one will be considered.

Clearly, it is the correlation between the fluctuations in temperature and vertical velocity which will be decisive in determining whether potential energy shall be a source or sink for the disturbances. It is in accordance with synoptic experience that in most cases cold air masses sink relative to the warmer ones. It is therefore to be expected that potential energy is, at least partly, an important factor in creating and maintaining the large-scale disturbances. In order to understand how a positive correlation between the fluctuations of temperature and vertical velocity may be brought about, one may write

$$\varkappa' = -\mathbf{l}\cdot\nabla\bar{\varkappa} = -K\mathbf{v}'\cdot\nabla\bar{\varkappa}, \qquad (K > 0).$$

Here, $\mathbf{l}$ is a kind of mixing length, and a positive correlation has been assumed between $\mathbf{l}$ and $\mathbf{v}'$. Consequently, one has

$$\int Q\varkappa' w' g\,d\tau = -K\int Q(\mathbf{v}'\cdot\nabla\bar{\varkappa})w' g\,d\tau = -K\int Qv'w'\frac{\partial\bar{\varkappa}}{\partial y}g\,d\tau - K\int Qw'^2\frac{\partial\bar{\varkappa}}{\partial z}g\,d\tau.$$

Having assumed vertical stability, it is therefore seen that as requirements for potential energy to be fed into the disturbances (*i*) horizontal temperature gradients must exist, and (*ii*) the slope of the streamlines in the meridional planes, w'/v', must be of the same sign as the slope $(-\partial\bar{\varkappa}/\partial y)/(\partial\bar{\varkappa}/\partial z)$ of the isentropic surfaces of the mean flow, but have a smaller magnitude. This last requirement may also be stated as saying that in the identity $\mathbf{v}'\cdot\nabla\bar{\varkappa} = v'\partial\bar{\varkappa}/\partial y + w'\partial\bar{\varkappa}/\partial z$ there must be a tendency for the vertical transport of entropy to compensate the horizontal transport. The latter, however, has to be the dominating effect, so that approximately $\mathbf{v}'\cdot\nabla\bar{\varkappa} = v'\partial\bar{\varkappa}/\partial y$.

For reasons of continuity it is to be expected that for decreasing horizontal dimensions of the disturbances the magnitude of the vertical velocities will increase relative to the horizontal velocities. It would therefore not be surprising if, for sufficiently small horizontal dimensions of the disturbances, w'/v' would exceed the values for which a conversion of potential energy into kinetic energy of the disturbances could take place. This effect will now be studied in more detail. To obtain results comparable with others which will be referred to at the end of this section, an incompressible atmosphere will be considered, bounded by two horizontal rigid planes at distance h apart. Also it will be assumed that

$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial y} = \frac{df}{dy} = 0; \qquad \frac{d^2u}{dz^2} = 0.$$

By eliminating $\nabla_h\gamma$ from the horizontal component of (2), one now obtains

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x}\right)\frac{\partial v}{\partial x} = -f\frac{\partial u}{\partial x} = f\frac{\partial w}{\partial z}. \tag{20}$$

Let it be assumed that instantaneously v and its derivatives may be obtained from the identity

$$v = \text{const}\cdot\sin\frac{2\pi}{L}\left[x - u\left(\frac{h}{2}\right)t\right],$$

representing instantaneously a simple wave propagating with a speed given by the value of u halfway between the boundaries. By substituting into (20), one obtains

$$-\frac{4\pi^2}{L^2}\frac{du}{dz}\left(z - \frac{h}{2}\right)v = f\frac{\partial w}{\partial z}.$$

Applying the boundary condition $w = 0$ for $z = 0$, one obtains by integration between $z = 0$ and the height $h/2$ where w reaches its maximum value:

$$\frac{w_{z=h/2}}{v} = \frac{\pi^2 h^2}{2L^2 f}\frac{du}{dz}.$$

Applying now the condition (*ii*) above to $w_{z=h/2}/v$, one obtains

$$\frac{\pi^2 h^2}{2L^2 f}\frac{du}{dz} < -\frac{\partial\varkappa/\partial y}{\partial\varkappa/\partial z},$$

or, by substitution from the thermal wind relationship $du/dz = -(g/f)(\partial\varkappa/\partial y)$:

$$\frac{L^2}{h^2} > \frac{\pi^2}{2f^2}\left(g\frac{\partial\varkappa}{\partial z}\right). \tag{21}$$

This now constitutes approximately the restriction on the horizontal scale of the disturbances if potential energy is to be converted into kinetic energy of the disturbances.

It will now be shown how one may express the conditions for a positive correlation between $\varkappa'$ and w'

in terms of the horizontal fields of velocity and temperature only. The conclusions arrived at are very much like those of Bjerknes and Holmboe [3]. The line of argument followed below is in some respects similar to that of Sutcliffe [26, 27]. (See also [8] and the article by J. G. Charney in this Compendium.[1])

From (2) one finds the vorticity equation in the vertical component to be

$$\frac{\partial \zeta}{\partial t} + \mathbf{v}_h \cdot \nabla \zeta + \beta v + f \nabla_h \cdot \mathbf{v} = 0,$$

when the presumably small terms

$$w(\partial \zeta / \partial z) + \zeta \nabla_h \cdot \mathbf{v} + \nabla_h w \times (\partial \mathbf{v} / \partial z) \cdot \mathbf{k}$$

are neglected. In this equation ζ is the vorticity and β is the variation of the Coriolis parameter with latitude. Let α^* be defined from $\alpha^* \int_0^\infty Q\, dz = \int_0^\infty Q\alpha\, dz$, where Q is now supposed to be the standard density in some isothermal atmosphere.[2] The term α^* will represent the mean, in the vertical, of α with respect to mass. Taking this mean of the vorticity equation, one obtains, by virtue of the continuity equation (3) and the boundary conditions $Qw = 0$ for $z = 0, z = \infty$,

$$\frac{\partial \zeta^*}{\partial t} + \mathbf{v}_h^* \cdot \nabla \zeta^* + \beta v^* + [(\mathbf{v}_h - \mathbf{v}_h^*) \cdot \nabla (\zeta - \zeta^*)]^* = 0. \tag{22}$$

In this equation the last term depends upon the existence of vertical wind shears, or in consequence of the thermal wind equation, upon the horizontal temperature gradients. It must therefore be this term that provides for the effects responsible for conversion of potential into kinetic energy. For a rough estimate of this term one may assume $d^2\mathbf{v}_h/dz^2 = 0$. From the thermal wind equation this implies that $\nabla_h \varkappa = \nabla_h \varkappa^*$. Equation (22) now takes the form

$$\frac{\partial \zeta^*}{\partial t} = -\mathbf{v}_h^* \cdot \nabla \zeta^* - \beta v^* - \mathbf{v}_T \cdot \nabla \zeta_T, \tag{23}$$

where

$$\mathbf{v}_h^* = \mathbf{v}_{z=0} + \mathbf{v}_T,$$

$$\mathbf{v}_T = H \frac{d\mathbf{v}_h}{dz} = -\frac{gH}{f} \nabla \varkappa^* \times \mathbf{k}, \tag{24}$$

and H = height of the homogeneous atmosphere. By taking the mean in the vertical of the physical equation (4), one further obtains

$$\frac{\partial \varkappa^*}{\partial t} = -\mathbf{v}_h^* \cdot \nabla \varkappa^* - \left(w \frac{\partial \varkappa}{\partial z} \right)^*. \tag{25}$$

To these equations one may add

$$\nabla \cdot \mathbf{v}_h^* = 0, \qquad \nabla \cdot \mathbf{v}_T = 0, \tag{26}$$

of which the first is exactly true owing to the definition of $\mathbf{v}_h^*$, and the second approximately true because of the identity (24).

It may be inferred from (22) that, under the foregoing assumptions, vorticity in the vertical-mean motion can vary individually only as a result of an advection of the vorticity of the thermal wind by the thermal wind $\mathbf{v}_T$. It is also apparent that if in the thermal wind field there is a transport of cyclonic thermal wind vorticity into regions of high vorticities in the actual motion, these vorticities will intensify. This corresponds to one of the rules developed by Sutcliffe for the sea-level motion [26, p. 205]. Applied to troughs which are symmetrical with respect to meridians, this leads, as Fig. 2 illustrates, to the synoptically well-known rule that troughs in the upper-air flow pattern intensify if troughs in the temperature patterns lag behind the troughs in the streamline patterns. A similar rule applies to ridges. It should be noted that these intensifications cannot be a result of the solenoids in horizontal planes, since these were neglected in the derivation of (2). The corresponding increase in the kinetic energy must have resulted from a conversion of potential energy. With reference to the discussion earlier in this section, it may therefore be concluded that, under the conditions mentioned above, the cold air must be subsiding relative to the warmer air, and further that the temperature changes due to the vertical motions can only partly compensate the advective changes. This rather definite knowledge of the three-dimensional flow structure based only upon a knowledge of the structure of the horizontal temperature and flow patterns is noteworthy. Clearly, nothing could in principle prevent a horizontal flow as illustrated in Fig. 2 from having any distribution of vertical velocities initially. This seeming discrepancy is due to the specific use which has been made of the conditions for quasi-geostrophic motion, which actually may be interpreted as effecting a smoothing similar to the one mentioned under the study of the mean meridional motions. As there, the success of such a smoothing, and therefore of the specific use of the conditions for quasi-geostrophic motion, depends upon the stability of the noise motion which is superimposed upon the "geostrophically" conditioned trend.

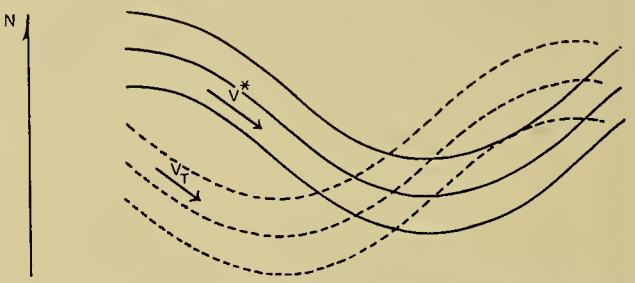


FIG. 2.—Flow pattern (solid lines) and temperature pattern (dashed lines) in an intensifying trough.

The question may now be raised as to what are the different mechanisms leading to the conditions mentioned above, under which potential energy can be converted into kinetic energy of the disturbances. Ad-

1. "Dynamic Forecasting by Numerical Process" by J. G. Charney, pp. 470–482.

2. The assumption of isothermalcy is not necessary, but convenient.

mitting that several such mechanisms may exist, the discussion here will be confined to the self-exciting one, by means of which small disturbances may amplify because of an instability of the underlying basic zonal flow. A theoretical attack on the problem of waves in a baroclinic atmosphere was first undertaken in the Norwegian polar front school of meteorology, principally by H. Solberg [25] (in addition, see [2, 5]). Waves were examined on an inclined surface of discontinuity separating two barotropic layers. Thus, all the baroclinity was considered as concentrated in a surface of discontinuity. In principle, however, with regard to the possibility of converting potential into kinetic energy, there is no difference between such a basic flow and one with continuously distributed baroclinity. In contrast to the pure, polar front waves, which may possibly feed also upon the kinetic energy of the basic current, are the baroclinic waves for a horizontally uniform basic flow, first examined by Charney [6]. The essential stability properties of these waves may be found easily by means of equations (23)–(26), if one neglects the vertical transport of potential temperature. As remarked in an earlier connection, and confirmed by the more rigorous solutions [10; 14, pp. 46–51], this approximation will be justified for the relatively large wave lengths. The adiabatic equation (3) may then be written

$$\left(\frac{\partial}{\partial t} + u^* \frac{\partial}{\partial x}\right) \varkappa^* + \frac{\partial \varkappa^*}{\partial y} v^* = 0, \tag{27}$$

or, by differentiation with respect to x and use of (24),

$$\left(\frac{\partial}{\partial t} + u^* \frac{\partial}{\partial x}\right) v_T - u_T \frac{\partial v^*}{\partial x} + \frac{\partial v_T}{\partial y} v^* + \frac{\partial u^*}{\partial x} v_T = 0.$$

Let it now be assumed by way of example that

$$\frac{\partial v^*}{\partial y} = \frac{\partial v_T}{\partial y} = 0,$$

so that the results arrived at may be said to apply approximately to disturbances whose scale in the y-direction is large compared with the scale in the x-direction. It follows then from (26) that $\partial u^*/\partial x = \partial u_T/\partial x = 0$. With the assumption of no horizontal shear for the zonal flow, u^* and u_T will now have to be independent of x and y. It is also easily understood that neither can they depend upon time. Equations (23) and (25) now reduce to

$$\left[\left(\frac{\partial}{\partial t} + u^* \frac{\partial}{\partial x}\right) \frac{\partial}{\partial x} + \beta\right] v^* + \left(u_T \frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}\right) v_T = 0, \tag{28}$$

$$u_T \frac{\partial}{\partial x} v^* - \left(\frac{\partial}{\partial t} + u^* \frac{\partial}{\partial x}\right) v_T = 0.$$

Suppose

$$v^* \sim \exp\left[i(\mu x + \omega t)\right]$$

and

$$v_T \sim \exp\left[i(\mu x + \omega t)\right]$$

to be solutions of (28). By substitution into (28) one gets as the condition that v^* and v_T do not vanish identically,

$$\omega = -\mu u^* + \frac{\beta}{2\mu} \pm \sqrt{\left(\frac{\beta}{2\mu}\right)^2 - \mu^2 u_T^2}.$$

When the square root becomes imaginary, and the negative sign is taken, v^* and v_T will increase exponentially. When the substitution for u_T is made from (24), the criterion for stability and instability therefore becomes

$$\left(\frac{\beta}{2\mu}\right)^2 - \mu^2 H^2 \left(\frac{du}{dz}\right)^2 \begin{matrix} < 0 & \text{(unstable)} \\ > 0 & \text{(stable).} \end{matrix} \tag{29}$$

Formula (29) reveals the high degree of instability of a horizontally uniform current in a baroclinic atmosphere. This result is common to all the different studies of baroclinic waves, which otherwise differ widely both with respect to the manner of formulating the problem of instability, and with respect to some of the conclusions obtained [1; 6; 10; 14, pp. 35–51]. Below is a summary of some of the assumptions and conclusions.

Charney

Assumptions:

1. Geostrophic approximation.
2. Compressibility.
3. Infinite atmosphere.
4. Vertical stability.

Conclusion:

All waves, at least down to about 1200 km, are unstable for a sufficiently large meridional temperature gradient.

Eady

Assumptions:

1. Geostrophic assumption.
2. Incompressibility.
3. Vertical stability.
4. Inertia effects of inhomogeneity neglected:

$$Q\left(\frac{D\mathbf{v}}{dt} + f\mathbf{k} \times \mathbf{v}\right) = \text{const}\left(\frac{D\mathbf{v}}{dt} + f\mathbf{k} \times \mathbf{v}\right).$$

5. Constant Coriolis parameter: $df/dy = 0$.
6. (*a*) Finite atmosphere bounded by two rigid walls at distance h apart.
   (*b*) Infinite atmosphere.

Conclusions from assumption (6) are:

(*a*) All waves are unstable[3] if and only if

$$\left(\frac{L}{h}\right)^2 > \frac{\pi^2}{2f^2}\left(g\frac{\partial \bar{\kappa}}{\partial z}\right).$$

(*b*) No waves are unstable unless a layer of smaller vertical stability is underlying one of greater stability.

3. Compare with the result obtained on p. 457.

FJØRTOFT

Assumptions:

1. Incompressibility.
2. Vertical stability zero.
3. Constant Coriolis parameter: $df/dy = 0$.
4. Finite atmosphere bounded by horizontal walls at distance h apart.

Conclusion:

All waves are unstable if and only if

$$\left(\frac{L}{h}\right)^2 > \frac{\pi^2}{2f^2}\left(\frac{du}{dz}\right)^2.$$

It may be remarked that Eady's second conclusion will no longer hold if the simplification mentioned in his fourth assumption is not made, in other words, even the infinite atmosphere with uniform vertical stability will be unstable if account is taken of the fact that density diminishes to zero for increasing heights. The appearance of a stabilizing influence from the vertical shear for short wave lengths in Fjørtoft's conclusion is due to effects to be discussed at the end of the following section. This stabilizing effect could not possibly appear in the other studies because there $(\omega + 2\pi u/L)^2$ was neglected compared with f^2 [14, pp. 41–42].

Barotropic Disturbances

In the case of barotropy there can be no vertical wind shear in the state of quasi-balance which characterizes the motions with which this article is concerned. Equations (18) and (21) therefore reduce to

$$\frac{D\zeta_{abs}}{dt} = 0; \qquad \nabla\cdot\mathbf{v} = 0, \tag{30}$$

with the asterisks now dropped as superfluous. These equations represent the classical equations for conservation of vorticity in a two-dimensional nondivergent flow of a nonviscous fluid. The fundamental investigation of wave motions in such flows was carried out by Rayleigh [22]. Although in the studies of polar front waves the destabilizing effects of a gliding discontinuity were considered to be important, it has not been until recently that barotropic phenomena in their full and complex generality have been taken up for systematic investigations, primarily by the Chicago school of meteorology. Starting with Rossby's work on planetary waves [24], in which the specific importance of the variability of the Coriolis parameter was discovered, a series of papers have followed in which different barotropic phenomena have been discussed. Briefly, one may say that while some of them, apart from the modifications following from the spherical shape of the earth, are analogous to the classical works by Rayleigh, others represent original investigations as exemplified by those treating stationary solutions of the nonlinear vorticity equation [9, 13, 17, 20]. The mathematical solution of (30) involves, of course, all the difficulties connected with the solution of nonlinear equations. The difficulties may even be very great for the linearized equations, particularly when there is a variable zonal current [18]. It is, however, possible and also useful to obtain an understanding of several of the most important barotropic phenomena by direct physical considerations [14, pp. 15–35]. In the following discussion this kind of argument will be used. We will make the purely formal simplification of assuming that the horizontal motion takes place as if on a circular disk; however, f will be kept a variable parameter.

One of the physical principles which may be used for a general discussion of some barotropic phenomena is the conservation of total angular momentum:

$$\int uR\, dF = \text{const.} \tag{31}$$

Here R is the distance from the pole, and dF a surface element in the plane of motion. By introducing the velocity circulation $c = \int_0^{2\pi} uR\, d\psi$ along zonal circles, (31) becomes equivalent to

$$\int c\, dF = \text{const.} \tag{32}$$

Another equivalent expression can easily be shown to be [14, p. 21]

$$G = \int \zeta_{abs}(R_0, \psi_0)R^2\, dF = \text{const.} \tag{33}$$

Here, ζ_{abs} is represented as a function of Lagrangian coordinates, and is therefore independent of time because the absolute vorticity is conserved, and R represents the generally time-variable radial positions of the fluid particles. Equation (33) is now a condition which restricts all future radial positions. In Fig. 3

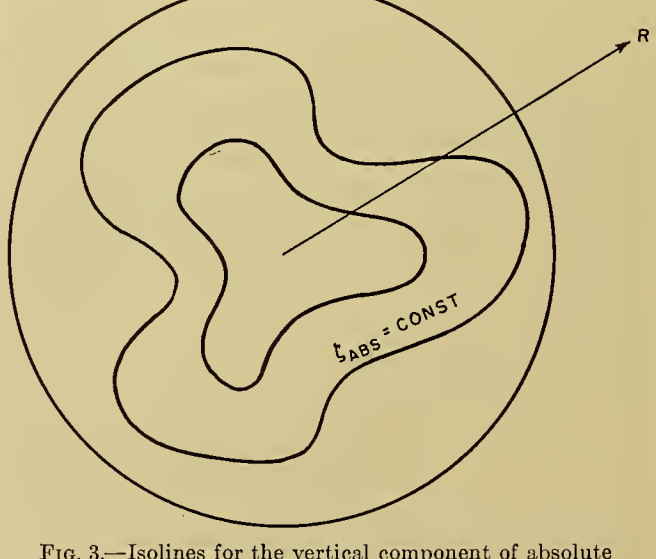


FIG. 3.—Isolines for the vertical component of absolute vorticity.

the irregular lines are curves of equal absolute vorticity for the case that ζ_{abs} varies monotonically from one isoline to the other. Clearly, a necessary and sufficient condition for having a pure zonal flow is that the lines

of equal vorticity have a zonal distribution. The motion corresponding to the vorticity distribution in Fig. 3 is therefore certainly not zonal. It may now be inferred from (33) that neither can it be so at any later time. The reason for this is that by varying the positions of the fluid particles, G assumes an extreme value if and only if the lines of equal vorticity are zonal. In the present case this extremum would be either an absolute minimum or maximum. Therefore, the isolines of ζ_{abs} could not possibly become zonal unless at the same time G varied in contradiction to (33). Thus, at all times one must have

$$\int \zeta'^2 \, dF \geqq m, \qquad (m > 0), \tag{34}$$

where a prime again has been used to indicate deviations from zonal means. With possible exceptions for some singular cases, it can be shown that (34) must be true for a quite general distribution of vorticities.

A problem of considerable interest is involved in the question as to what will happen in a barotropic atmosphere left with a certain distribution of vorticities that do not correspond to a motion which is stationary, either absolutely or with respect to a coordinate system rotating at a constant speed. Will the structure of the subsequent motion tend to approach a more or less definite limit, or will different structures be repeated more or less periodically? Synoptic experience is probably most in favor of the first point of view, though there have been attempts to interpret some cycles in the general circulation on a strictly barotropic basis. In favor of the first point of view one can at least say with certainty that theoretically no oscillations in the strict sense are possible. This is because in a purely oscillatory motion the fluid particles would simultaneously have to reassume earlier positions, but with different velocities. However, this is impossible because, according to the conservation of vorticity, the positions of the fluid particles determine uniquely the vorticity distribution which in turn determines the velocities.

From the condition $\int \zeta'^2 \, dF + \int \bar{\zeta}_{abs}^2 \, dF = \text{const}$, there is also an upper limit to $\int \zeta'^2 dF$, so that (34) may be extended to

$$M \geqq \int \zeta'^2 \, dF \geqq m, \qquad (m > 0). \tag{35}$$

Returning now to the kind of vorticity distributions which are illustrated in Fig. 3, it will be assumed as a particular case that the vorticity distribution deviates only slightly from a zonal one. The corresponding flow is a zonal flow upon which there are superimposed small disturbances. The time-invariant value of G will now be but slightly different from a maximum or minimum value, which implies that no change in the vorticity pattern from the nearly zonal one to one characterized as in Fig. 3 can take place without necessarily changing the value of G. The proof for this is simply the reversal of the one stating that it was not possible to come arbitrarily near to a zonal distribution. The corresponding zonal flow is, therefore, stable in this case. Let $\bar{\zeta}_0 + f$ and ζ_0' be the initial vorticities in the mean zonal flow and in the disturbances. An expression and criterion for the present stability is

$$M \geqq \int \zeta'^2 \, dF \geqq m, \qquad (m > 0), \tag{36a}$$

$$M \to 0 \quad \text{when} \quad \int \zeta_0'^2 \, dF \to 0, \tag{36b}$$

if

$$\bar{\zeta}_0 + f \quad \text{varies monotonically with latitude.} \tag{36c}$$

It should be mentioned that the result which has been found that polar anticyclones are always unstable in a barotropic atmosphere [23] cannot be true if (36) is true. That is because anticyclones may exist for which condition (36c) still may be true.

Suppose now again that the absolute vorticity for the mean zonal flow varies monotonically with latitude and consider the possible changes in the mean zonal flow which will result from the meridional exchange of air. It may be assumed, in accordance with the most frequent conditions, that $d(\bar{\zeta}_0 + f)/dy > 0$. One has

$$\bar{u} = \frac{c}{2\pi R} = \frac{1}{2\pi R} \int_F \zeta \, dF,$$

where F is the area north of the latitude circle which is being considered. Since equal areas of fluid are going in and out of a latitude circle, $\bar{u}$ will have to decrease or increase according as the vorticities leaving the circle are replaced by vorticities of lower or higher magnitudes, respectively. Now, each fluid particle is assigned a certain absolute vorticity $\bar{\zeta}_0 + f + \zeta_0'$ which is moving with the particle. The effect of this transport may be considered as a separate effect from those of the transport of $\bar{\zeta}_0 + f$ and of ζ_0'. As to the first effect it is easily understood that because $\bar{\zeta}_0 + f$ is increasing northwards, air streaming out of a zonal circle will have to be replaced by air with lower value of the absolute vorticity. This effect, when considered alone, will therefore amount to a decrease in c at all latitudes. This is in conflict with the principle of the conservation of total angular momentum, equation (32). It can, therefore, immediately be inferred that the effect from the transport of the initial irregular vorticities must be to compensate exactly this loss in angular momentum, that is, to create in the mean a compensating westerly flow [14, p. 28]. In consequence of this it may be concluded that, in the mean at least, the air with positive ζ_0' has to be transported to the north, and that with negative ζ_0' to the south. In the special case of a stationary flow pattern, as for instance for the Rossby waves, the two effects compensate each other exactly at each latitude. In other cases, however, one can only expect that the compensation is accomplished after integration over all latitudes. It has been assumed by Kuo [19] that the effect of the transport

of the irregular vorticities will be the dominating one at middle latitudes, resulting there in an increase of the westerlies. However, the arguments used are not convincing, and without carrying out numerical calculations nothing certain can be said with respect to the resulting changes in the mean zonal flow. Preliminary calculations [15] seem to indicate that rather than increasing the westerlies at middle latitudes, the tendency under certain conditions may as well be to decrease the westerlies at middle latitudes and increase them in belts to the north and south, particularly to the north.

In a barotropic atmosphere the energy equation reduces to

$$\int \tfrac{1}{2}\mathrm{v}'^2\,dF = -\int \tfrac{1}{2}\bar{u}^2\,dF + \text{const.}$$

Consequently, there is an upper bound to the kinetic energy of the disturbances. In the case where ζ_{abs} varied monotonically from one isoline of vorticity to the next it was found that there must be a lower bound to $\int \zeta'^2\,dF$ which is different from zero. The same must, therefore, be the case for the kinetic energy of the irregular flow. Thus, in this case

$$A \geqq \int \tfrac{1}{2}\mathrm{v}'^2\,dF \geqq a, \qquad (a > 0). \tag{37}$$

The stability under condition (36c) therefore does not imply that the disturbances are entirely damped out. Whether the total kinetic energy of the disturbances will increase or decrease is usually difficult to ascertain and will depend upon the character of the changes in $\bar{u}$. Writing

$$\begin{aligned}\tfrac{1}{2}\int (\mathrm{v}'^2 - \mathrm{v}_0'^2)\,dF &= -\tfrac{1}{2}\int (\bar{u}^2 - \bar{u}_0^2)\,dF \\ &= -\int \frac{\bar{u}_0}{R}(\bar{u}R - \bar{u}_0 R)\,dF - \tfrac{1}{2}\int (\bar{u} - \bar{u}_0)^2\,dF,\end{aligned} \tag{38}$$

one finds easily [14, p. 23] by using condition (31) in the form

$$\int (\bar{u}R - \bar{u}_0 R)\,dF = 0 \tag{39}$$

that $\bar{u}$ has to decrease where $\bar{u}_0/R$ is large and increase where $\bar{u}_0/R$ is small, if the kinetic energy of the disturbances is to increase. Particularly if one has to do with small disturbances in a zonal flow, the upper bound to the changes in $\bar{u}$ which can be caused by a transport of the initial irregular vorticities is also a small quantity. In order to find necessary conditions for real instability one has therefore to investigate the effect from the transport of $\bar{\zeta}_0 + f$. It was found earlier that $\bar{u}$ would decrease if $d(\bar{\zeta}_0 + f)/dy > 0$. By similar arguments one will find that $\bar{u}$ on the other hand has to increase where $d(\bar{\zeta}_0 + f)/dy < 0$. So, the necessary conditions for real instability will be that

$$\begin{aligned}&\frac{d}{dy}(\bar{\zeta}_0 + f) > 0 \quad \text{where } \frac{\bar{u}_0}{R} \text{ is large,} \\ &\frac{d}{dy}(\bar{\zeta}_0 + f) < 0 \quad \text{where } \frac{\bar{u}_0}{R} \text{ is small.}\end{aligned} \tag{40}$$

The trivial case with solid rotation in the mean flow, $\bar{u}_0/R = \text{const}$, implies, of course, according to (38) and (39), that $\int \tfrac{1}{2}\mathrm{v}'^2\,dF$ at most can remain constant in time, but will decrease if some changes in $\bar{u}$ result.

Hitherto, the most complete mathematical treatment of barotropic waves in a basic flow which is unstable in the above sense has been undertaken by Kuo [18]. This instability is fundamentally the same as the one occurring when a gliding discontinuity exists. However, by treating the realistic case with a continuous shear and including the variation of the Coriolis parameter, two important modifications result:

1. As results of an assumed continuous shear under average atmospheric conditions:

a. All waves below approximately 300 km are stable.

b. An intermediate wave length of maximum instability exists.

2. As a result of the inclusion of $df/dy \neq 0$, the longest waves become stable.

It is important to notice that the most unstable waves of the type discussed above are relatively long waves compared with the most unstable baroclinic waves [10; 14, p. 50].

How can the stability occurring for short waves mentioned above be understood? It was previously seen that provided conditions (40) were fulfilled the kinetic energy of the disturbances would necessarily increase as a result of a transport of $\bar{\zeta}_0 + f$. Therefore, when the shortest waves become stable this can only be a result of the transport of the initial irregular vorticities, ζ_0', which furthermore must tend to be the dominating effect for the shortest wave lengths. An understanding of the stabilizing influence arising from the transport of irregular vorticities may be obtained in the following way: Suppose a wavelike disturbance with untilted troughs and ridges to exist initially in a nonuniform zonal current. The instantaneous transport of the irregular vorticities is accomplished by a component $\bar{u}_0\mathbf{i}$ of the mean flow and a component $\mathbf{v}_0'$ of the irregular flow. When small disturbances are considered, or disturbances in which $\mathbf{v}_0'$ is essentially parallel to the lines $\zeta_0' = \text{const}$, only the transport by the first component has to be considered. It is now obvious that if the angular velocity $\bar{u}_0/R$ varies with latitude, the lengths of the lines $\zeta' = \text{const}$ have to increase as a result of this transport. On the other hand, the areas enclosed by these lines are conserved as are also the values of ζ' because the effects of the transverse displacements of the vorticities $\bar{\zeta}_0 + f$ are disregarded in this connection. Consequently, it follows from Stokes' theorem that the velocity circulation for the disturbances taken along the closed curves $v\zeta' = \text{const}$, which approximately are also streamlines for $\mathbf{v}'$, must remain

constant. But since the lengths of the streamlines for v' increase, the average intensity of $|v'|$ must decrease correspondingly. When this stabilizing influence dominates the destabilizing influence from the transverse transport of vorticities of the basic flow, which can be shown to be the case for the shortest waves [14, p. 31], kinetic energy must flow from the disturbances to the mean flow so that in accordance with what was said above, $\bar{u}$ will have to increase where $\bar{u}/R$ is large and decrease where $\bar{u}/R$ is small.

Combined Baroclinic and Barotropic Disturbances

While there is enough evidence for the importance both of baroclinic and barotropic effects, there must be a limit to the extent to which phenomena can be explained purely barotropically or baroclinically. In general, one must expect that barotropic and baroclinic effects either add together in a more or less simple fashion, or they may be coupled to such a degree that the consideration of both effects simultaneously may give rise to entirely new types of phenomena.

The only studies until now on waves for which both potential and kinetic energy are possible sources for the growth of the disturbance are those on polar front waves. Because of the complexity in the solutions for these waves one does not know whether the one or the other of these two possible sources is the more important, although the shearing instability has been interpreted as the decisive one, seemingly, however, without any convincing justification. The unstable baroclinic waves treated in this article have accordingly been looked upon as physically entirely different waves. It is the writer's opinion that this probably is not true. Further investigations on this subject can be carried out with relative ease when the quasi-geostrophic approximation is made. Relatively simple equations appropriate for the most simple polar front model have been worked out by Phillips [21].

It is not unlikely that the study of an atmosphere where typical barotropic and baroclinic effects are operating in full generality may contribute considerably to a further understanding of the behavior of the atmosphere. For this purpose equations (23) and (25), or essentially similar equations, may prove useful, at least for theoretical investigations, because of their great simplicity and generality.

REFERENCES

1. Berson, F. A., "Summary of a Theoretical Investigation into the Factors Controlling the Instability of Long Waves in Zonal Currents." *Tellus*, Vol. 1, No. 4, pp. 44–52 (1949).
2. Bjerknes, J., and Godske, C. L., "On the Theory of Cyclone Formation at Extra-tropical Fronts." *Astrophys. norveg.*, Vol. 1, No. 6 (1936).
3. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1:1–22 (1944).
4. Bjerknes, V., "Application of Line Integral Theorems to the Hydrodynamics of Terrestrial and Cosmic Vortices." *Astrophys. norveg.*, 2:263–339 (1937).
5. —— and others, *Hydrodynamique physique*. Paris, Presses Universitaires de France, 1934.
6. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4:135–162 (1947).
7. —— "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2, 17 pp. (1948).
8. —— Fjørtoft, R., and Neumann, J. v., "Numerical Integration of the Barotropic Vorticity Equation." *Tellus*, 2:237–254 (1950).
9. Craig, R. A., "A Solution of the Nonlinear Vorticity Equation for Atmospheric Motion." *J. Meteor.*, 2:175–178 (1945).
10. Eady, E. T., "Long Waves and Cyclone Waves." *Tellus*, Vol. 1, No. 3, pp. 33–52 (1949).
11. Eliassen, A., "Slow Thermally or Frictionally Controlled Meridional Circulation in a Circular Vortex." Unpublished manuscript (1951).
12. —— "The Quasi-static Equations of Motion with Pressure as Independent Variable." *Geofys. Publ.*, Vol. 17, No. 3, 44 pp. (1949).
13. Ertel, H., "Die Westwindgebiete der Troposphäre als Instabilitätszonen." *Meteor. Z.*, 60:397–400 (1943).
14. Fjørtoft, R., "Application of Integral Theorems in Deriving Criteria of Stability for Laminar Flows and for the Baroclinic Circular Vortex." *Geofys. Publ.*, Vol. 17, No. 6 (1950).
15. —— Unpublished manuscript, 1950.
16. Høiland, E., "On the Interpretation and Application of the Circulation Theorems of V. Bjerknes." *Arch. Math. Naturv.*, Vol. 42, No. 5, pp. 25–57 (1939).
17. —— "On Horizontal Motion in a Rotating Fluid." *Geofys. Publ.*, Vol. 17, No. 10 (1950).
18. Kuo, H.-L., "Dynamic Instability of Two-Dimensional Nondivergent Flow in a Barotropic Atmosphere." *J. Meteor.*, 6:105–122 (1949).
19. —— "The Motion of Atmospheric Vortices and the General Circulation." *J. Meteor.*, 7:247–258 (1950).
20. Neamtan, S. M., "The Motion of Harmonic Waves in the Atmosphere." *J. Meteor.*, 3:53–56 (1946).
21. Phillips, N. A., Unpublished manuscript, 1950.
22. Rayleigh, Lord, *Scientific Papers*, 6 Vols. Cambridge, University Press, 1899–1920. (See Vol. I, pp. 474–490; Vol. III, pp. 17–23, 575–584; Vol. VI, pp. 197–204.)
23. Rossby, C.-G., "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6:163–180 (1949).
24. —— and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
25. Solberg, H., "Das Zyklonenproblem." *Verh. III intern. Kongress für techn. Mechanik* (1930).
26. Sutcliffe, R. C., "A Contribution to the Problem of Development." *Quart. J. R. meteor. Soc.*, 73:370–383 (1947).
27. —— and Forsdyke, A. G., "The Theory and Use of Upper Air Thickness Patterns in Forecasting." *Quart. J. R. meteor. Soc.*, 76:189–217 (1950).

THE QUANTITATIVE THEORY OF CYCLONE DEVELOPMENT

By E. T. EADY

Imperial College of Science and Technology

Introduction

The title of this article indicates its principal theme but not its full scope, for although we shall explain the method of formulating and solving certain theoretical problems and shall interpret the answers in terms of the initial stages of development of extratropical cyclones and anticyclones, our analysis has also a wider significance. Not only does a fundamentally similar theoretical analysis apply to a wide variety of development problems (including, for example, the development of "long" waves as well as the shorter "frontal" waves and even phenomena due primarily to ordinary convective instability) so that a comprehensive analysis is desirable, but this analysis gives results of primary importance in the theory of development *in general*, that is, from the point of view of the general forecasting problem. We shall infer from our results that there exist, in general, certain ultimate limitations to the possibilities of weather forecasting. Certain apparently sensible questions, such as the question of weather conditions at a given time in the comparatively distant future, say several days ahead, are *in principle* unanswerable and the most we can hope to do is to determine the relative *probabilities* of different outcomes. The full significance of our theoretical problems becomes apparent only when it is clear what *kind* of question we should attempt to answer.

The science of meteorology is a branch of mathematical physics; it can be fully understood only in a quantitative manner. Moreover, all the practical questions we should like to answer are of a quantitative character. Having discovered the relevant equations of motion, we ought to aim at obtaining significant integrals (more precisely, solutions of significant boundary-value problems) which may be applied directly to practical problems. In order to obtain tractable problems, and at the same time to see clearly what we are doing (*i.e.*, "to see the wood for the trees"), we may, for a first analysis, simplify the equations by omitting all factors not vitally affecting the nature of the answer. Later we may refine our solutions by taking into account factors previously omitted (*e.g.*, by the method of successive approximations), thereby testing whether we have in fact included all the vital factors. This is a procedure with which we are familiar and, however laborious it may be in practice, it introduces no new difficulty in principle. The really serious difficulty is to discover what kind of problem ought to be solved, for this difficulty arises as soon as we consider the question of the stability of atmospheric motion. Observation suggests that the motion may, at least sometimes, be unstable, and we shall infer from subsequent analysis that instability (to a greater or less degree) is a *normal* feature of atmospheric motion.

It is important to be quite clear as to the meaning of the term "unstable" when applied to a system of fluid motion. If we suppose the initial field of motion to be given, the final field of motion, after a given interval of time, is determined precisely by the equations of motion, continuity, radiation, etc., together with the appropriate boundary conditions. If we consider a slightly different (perturbed) initial state, the new final state, after the same interval of time, will be determined in a similar manner. The stability or instability of the motion depends on the behaviour of the resulting change (perturbation) in the final state as the time interval is increased. If the final perturbation remains small for all time for *all* possible initial perturbations, the motion is stable. If, on the other hand, the perturbation in some or all regions grows (initially) at an exponential rate for *any* possible initial perturbation, the motion is unstable. There is an intermediate case, conveniently described as *neutral* stability, when the perturbations grow linearly or according to a low-degree power law, but this need not concern us here.

The practical significance of a demonstration that the motion is unstable is clear, for in practice, however good our network of observations may be, the initial state of motion is never given precisely and we never know what small perturbations may exist below a certain margin of error. Since the perturbation may grow at an exponential rate, the margin of error in the forecast (final) state will grow exponentially as the period of the forecast is increased, and this possible error is unavoidable whatever our method of forecasting. After a limited time interval, which, as we shall see, can be roughly estimated, the possible error will become so large as to make the forecast valueless. In other words, the set of all possible future developments consistent with our initial data is a divergent set and any direct computation will simply pick out, arbitrarily, one member of the set. Clearly, if we are to glean any information at all about developments beyond the limited time interval, we must extend our analysis and consider the properties of the set or "ensemble" (corresponding to the Gibbs-ensemble of statistical mechanics) of all possible developments. Thus long-range forecasting is necessarily a branch of statistical physics in its widest sense: both our questions and answers must be expressed in terms of *probabilities*.

There are two important connections between these general considerations and subsequent analysis. Firstly, this analysis will show the existence of at least one type of large-scale unstable disturbance in a simplified but typical system, and we shall infer that instability is a

normal feature of atmospheric motion. Although the unstable disturbances are continually tending to establish a new stable state, radiative processes are continuously tending to restore the initial system, which therefore remains *permanently* unstable. Secondly, for such a system the study of the ensemble of all possible perturbations is relatively simple. In each system there exists a disturbance of maximum growth-rate (so that we can determine the growth of the margin of error and estimate the limited time interval referred to above) which eventually becomes dominant in subsequent developments by a process analogous to Darwinian natural selection. Almost any initial disturbance tends eventually to resemble the dominant, which is therefore the most probable development. The "ensemble" possesses at least *some* strongly marked statistical properties which may be utilised to extend the range of forecasts. In spite of inaccuracies due to oversimplification this result is practically significant. The disturbances referred to are approximations to nascent cyclones, long waves, etc.; were it not that "natural selection" is a very real process, weather systems would be much more variable in size, structure, and behaviour.

The Basic Equations

We shall regard as basic equations the three dynamical equations, the thermal equation, and the equation of continuity; others, such as the gas laws and the laws of radiation, will be regarded as subsidiary. The number of dependent variables we need, or that we find it convenient to use, depends on the nature of the problem and the degree of accuracy aimed at. In the problems with which we shall be concerned it is possible to express the basic equations in terms of the three components of velocity, pressure, and entropy (or density) alone so that the five basic equations, together with appropriate boundary conditions, form a complete set. Clearly these equations can be appropriate only for a limited range of problems when certain approximations are justified; we shall in fact make further approximations, our aim being to retain only those terms which are of prime importance in the range in which we are interested.

A completely realistic theory of the stability of atmospheric motion should deal with nonsteady initial conditions, but for simplicity we shall confine our attention to the case in which the initial motion is steady, and in fact we shall be concerned mainly with rectilinear horizontal motion. Our analysis will be approximately true even when the very-large-scale distribution is slowly changing.

The relative importance of the terms in our equations depends partly on the scale of the phenomena with which we are concerned. Here we are interested in disturbances of the order of magnitude of nascent cyclones, say 1000 km in horizontal extent and occupying a large part (or the whole depth) of the troposphere. From our point of view ordinary or gravitational convection, originating from static instability (*i.e.*, superadiabatic lapse rate), and ordinary turbulence of frictional or convective origin are small-scale phenomena. The epithet "ordinary" is appropriate in each case because the disturbances whose nascent form we are studying may be regarded as elements of a large-scale convective process and this process, regarded statistically, is a kind of large-scale turbulence. From our point of view the significance of small-scale turbulence, including ordinary convection, lies in its statistical properties, such as ability to transport heat, momentum, etc. Now frictionally induced turbulence is most effective near the earth's surface, and rough calculations (which we have not space to describe) using empirical estimates of skin friction indicate that frictional dissipation of energy usually has a relatively small effect on the development of large-scale disturbances (especially over a sea surface) in their *nascent* stage, provided the unstabilising factors are not too weak. Since we are most interested in those regions where the unstabilising factors are relatively strong, we may obtain a useful first approximation during the nascent stage if we neglect the frictional terms in the equations of motion. It is not possible to neglect frictional terms throughout the whole life-history of a disturbance because in the long run the kinetic energy destroyed by friction must equal that generated as a result of instability.

Surface friction transports heat vertically through a shallow layer, but since we are interested in the behaviour of deep layers this effect will be neglected. Moreover, surface turbulence is partly convective in origin, and we may regard shallow convection as included in this argument. But sometimes (*e.g.*, in strong polar outbreaks) deep and widespread convection transports heat to great heights at a great rate. We shall ignore this possible complication and concentrate our attention on systems in middle and high latitudes which are statically stable in their initial stages.

Just as, in the long run, we cannot ignore skin friction so, in the long run, we cannot ignore radiative processes. Large-scale turbulence (the statistical aspect of our disturbances) appears to be a major factor in transporting heat poleward to compensate the unbalanced radiation flux. But during the nascent stage, development (measured by the time for growth of the disturbance by a given factor) is relatively rapid and it is precisely for this reason that we are able to neglect frictional terms. Hence it is reasonable to suppose that in the nascent stage we may, for a first approximation, neglect the change in the radiation balance caused by the disturbance and use for our thermal equation the adiabatic equation.

Consider first the case of unsaturated air. To a close enough approximation the entropy of dry air is measured, in suitable units, by $\Phi \equiv (1/\gamma) \ln p - \ln \rho$, where p = pressure, ρ = density, γ = specific heat ratio, and Φ is conserved during the motion. In this case we shall define the static stability, which measures the restoring force due to gravity on a particle displaced vertically, as $\partial\Phi/\partial z$ (z = vertical coordinate). Now consider the case of saturated air in contact with a cloud. The static stability is now measured by the difference

between the actual entropy lapse and that of the appropriate wet-adiabatic. Thus there is a sharp, and usually a large, reduction in static stability when air becomes saturated. The effective horizontal entropy gradients are also modified as a result of saturation, but to a much smaller extent. Normally a cloud mass behaves, to a sufficiently close approximation, as if the air were unsaturated except for the appropriate modification in static stability.

Our equations are complicated by the fact that air is a compressible fluid, but it is clear that this feature is not, for our purposes, a significant one. We are concerned essentially with a particular type of "vibration" problem though our disturbances have mathematically complex wave velocities. The moduli of these wave velocities are in all cases, as our calculations verify, small compared with the velocity of sound. It is not surprising therefore that the forces associated with compressibility are negligible, that is, that the air behaves *dynamically* as if it were incompressible. The static effect of compressibility, involving a large change of density with height, is a complication which prevents atmospheric motion from being quite the same as that of an incompressible fluid. Nevertheless, even for deep disturbances, the behaviour differs little from that of an incompressible fluid of similar mean density: the modifications are essentially of the nature of distortions of wave structure without much change in more significant features like growth rate. We shall therefore confine our attention to "equivalent" incompressible fluid systems. Our results are, of course, more directly applicable to analogous oceanographic problems.

Lack of space prevents a discussion of these points in mathematical terms. It has been shown elsewhere [3] that the basic equations may be further simplified by the elimination of the pressure field, so that we have finally four equations connecting the four dependent variables V_x, V_y, V_z (velocity components), and Φ (entropy). But our present concern is the physical interpretation and practical significance of certain calculations rather than the calculations themselves.

The General Theory of the Instability of Fluid Motion

The various types of instability occurring in dynamical meteorology merge into one another so that most systems encountered in practice are, to a greater or less degree, hybrid. Nevertheless it is not only simpler but theoretically more instructive to consider certain ideal limiting cases where one or another unstabilising factor acts alone. Four simple types of instability will interest us:

1*a*. Gravitational instability (ordinary convection or static instability).

1*b*. Centrifugal instability (dynamic instability).

2*a*. Baroclinic instability (with thermal wind).

2*b*. Helmholtz instability (at a velocity discontinuity).

Instability of type 1*a* is, in middle and high latitudes, nearly always a small-scale phenomenon, but in low latitudes a modified form, taking into account the rotation of the earth, is intimately concerned with the development of tropical cyclones.

Instability of type 1*b* has been the subject of much recent investigation, usually under the heading "dynamic instability," but despite its theoretical importance it is probably rare for large-scale motion. The name "centrifugal" has been preferred to "dynamic" because it is more descriptive and less confusing—other types of instability may reasonably be called "dynamic."

Instability of type 2*a* is probably the most important, on a large scale, in middle and high latitudes. It is to this type that our earlier remarks regarding the normality of instability and the existence of "natural selection" directly apply.

Instability of type 2*b* was investigated by Helmholtz and Rayleigh for nonrotating barotropic fluids. The Norwegian wave theory of cyclones was a partially successful attempt to extend the theory to rotating barotropic fluids.

Although we shall choose our initial systems so that only one type of instability is in question, the same general method of analysis applies in every case. Using the method of small perturbations, we obtain a set of simultaneous, linear, partial differential equations involving the perturbations as dependent variables. By elimination we obtain a partial differential equation with only one dependent variable and look for simple solutions satisfying appropriate boundary conditions. Usually these solutions involve only circular or exponential functions in the horizontal (x and y directions) and all contain the factor $e^{\theta_1 t}$ where t represents time and θ_1 is a constant called the growth rate. For θ_1 to be real we usually have to use a moving coordinate system. Fortunately these solutions for unstable waves are, practically, the most important ones and the disturbance of maximum growth rate, when it exists, is probably dominant relative to one of *arbitrary* initial structure. In any case a study of these particular solutions enables us to *understand* the process of breakdown of the initial system and to estimate the relative importance of various factors.

The method of analysis outlined above is necessary if we require precise results and is the only one which is completely unequivocal. But it is mathematical in form and usually rather involved so that significant physical principles, which give us insight into our problems and immediately suggest generalisations, tend to be obscured. Now, except that our interest is centred in the unstable region, we are concerned with what are essentially vibration problems and we may expect to find that energy considerations are of paramount importance. For, by the law of conservation of energy, the kinetic energy associated with any perturbation must be equal to the decrease in "potential" energy of the system, and a necessary condition for instability is that it should be possible to find displacements which will decrease "potential" energy; the condition will be sufficient only if these displacements are consistent with all the equations of motion and boundary conditions. More precisely, using Rayleigh's method, we

may express separately the changes in kinetic and "potential" energy in terms of arbitrary displacements of the form $\delta x = e^{\theta_1 t}\delta x_0(x, y, z)$, etc. The kinetic energy change contains the factor θ_1^2, while the "potential" energy change does not, so that the law of conservation of energy gives an expression for θ_1^2 in terms of the displacements. The possible simultaneous values of the displacements are restricted (or constrained) by the equations of motion and we can to some extent delimit possible values of θ_1^2 by considering only some of the constraints, as in somewhat analogous problems in dynamics. In the present instance we consider only the equation of continuity and one momentum equation (and suitable boundary conditions) and then apply algebraic inequality theory to our undetermined displacements. We thereby obtain an *upper* bound to the possible value of θ_1^2 (corresponding to *negative* square of frequency) just as in dynamics we obtain a lower bound to frequency (squared) by considering a less constrained system.

The value of the foregoing, described in more detail elsewhere [3], derives partly from the fact that we can treat a wider variety of problems than we can by complete solution, while the value of θ_1^2 (maximum) thereby obtained is usually not much greater than the true value of θ_1^2. But its greatest usefulness is that it makes the process of breakdown of unstable systems immediately intelligible. If we take into account only our limited set of constraints, it is immediately evident what kind of displacement field is necessary for a release of "potential" energy. It is of course essential that the term "potential" energy be correctly interpreted. For our purpose it comprises *all* forms of energy other than the kinetic energy of the perturbation and therefore includes, besides gravitational potential energy, the organised kinetic energy of the mean flow (smoothed of harmonic variation). This distinction between two kinds of kinetic energy change is justified by their different roles in the turbulent motion which is the ultimate state in practice: turbulent energy may arise either from a decrease in gravitational potential energy or from a decrease in kinetic energy of the mean motion. We may classify our systems according to whether the "potential" energy source is (*a*) static (gravitational) or (*b*) dynamic (kinetic). Now the displacement field is merely the nascent form of a process of overturning and we shall need to consider only two possibilities: (1) overturning in a *vertical* plane; (2) overturning in a *quasi-horizontal* plane. Thus we may also classify our systems according to the kind of overturning associated with instability. In our list of four simple types of instability we anticipated their classification from both points of view. Let us consider the characteristics of these systems.

1*a*. *Gravitational Instability.* We consider barotropic conditions, so that initially there is no wind change with height, and for simplicity we suppose that $d\Phi/dz = B$, where B is constant. If, to begin with, we neglect the rotation of the earth, then "potential" energy exists only in gravitational potential form. Suppose that two small parcels of air of equal potential volume were slowly interchanged. Then since potential density would depend only on entropy, we should obtain, if B were negative, a net release of energy for *any* two parcels at different levels, while if B were positive no interchange could release potential energy. Clearly the constraints associated with the continuity equation cannot alter this result—the overturning process is equivalent to a set of such interchanges of different amplitudes. Since horizontal motion does not affect potential energy we need consider only vertical overturning; calculations by the energy method give $\theta_1^2 \leq -gB$, where g = gravitational acceleration. Of course in this simple case it is easy to obtain complete solutions, representing the nascent stage of Bénard cells, and calculations show that θ_1^2 (maximum) is nearly attained for narrow deep cells, where little energy is wasted in horizontal motion. We shall postpone the extension to large-scale convection, where the rotation of the earth is considered, since this is really a combination of types 1*a* and 1*b*.

1*b*. *Centrifugal Instability.* We shall suppose the motion to be barotropic and horizontal with the initial velocity V_x a function of y only. For simplicity we take dV_x/dy constant and, to begin with, we put $B = 0$ (isentropic conditions). Then "potential" energy exists only in the "kinetic" form. Let us consider the change due to overturning in the (vertical) y, z plane. Filaments of air in the x-direction move as a whole and we easily derive from the equations of motion that, during displacement, $\delta V_x = f\delta y$, where f is the Coriolis parameter. If the x-axis is directed toward the east, this corresponds to constancy of absolute angular momentum. But for our purposes, where a mean value of f is used, the orientation of the x-axis is arbitrary. A simple calculation shows that potential energy is released only if $dV_x/dy > f$, corresponding to negative absolute vorticity, and the energy method gives $\theta_1^2 \leq f(dV_x/dy - f)$. Although values of dV_x/dy *near* the critical value are sometimes observed in narrow bands, it is doubtful whether centrifugal instability ever occurs on a large scale except perhaps in low latitudes. The rotation of the earth, normally at least, has a stabilising effect so far as *vertical* overturning is concerned.

Similar results are obtained if, instead of rectilinear motion, we consider a barotropic circular vortex (with no motion relative to the earth as a special case). The condition for instability is again negative absolute vorticity.

We may note that in both the foregoing cases maximum instability occurs for shallow, flat, cells since "potential" energy changes depend only on horizontal motion (no energy is wasted in vertical motion).

1*ab*. *Gravitational-Centrifugal Instability.* It is easy to combine the results of the previous sections for a system in which neither B nor dV_x/dy vanishes. There is instability if either B or $(f - dV_x/dy)$ is negative, for the cells may be either so deep that centrifugal stability is negligible or so shallow that static stability is negligible. The important practical case is that of no motion (special case of circular vortex) with $B < 0$. Instability occurs for disturbances which are sufficiently

deep relative to their breadth, the condition being that there should be a *net* release of "potential" energy. In low latitudes not only is $-B$ sometimes (temporarily) relatively large but the stabilising effect of the earth's rotation is small, so that convection cells of relatively enormous diameter (nascent hurricanes) can develop.

In general, maximum growth rate corresponds either to very deep or to very shallow cells but this result is not of great significance because practical systems are very inhomogeneous.

2*a*. *Baroclinic Instability.* We suppose the initial motion to be rectilinear and take the initial velocity V_x to be a function of z only. For equilibrium this implies a horizontal gradient of entropy $A \equiv \partial\Phi/\partial y$, where, approximately ($A$ not too small), $dV_x/dz = -gA/f$. For simplicity we suppose A to be constant. Pure baroclinic instability should correspond to $B = 0$, but it will be convenient to consider directly the more general case $B \neq 0$. In practice we usually have (at least in the mean) $B > 0$, and this is the only case we need examine, for when $B < 0$ the system is obviously unstable. The isentropic surfaces have an angle of slope $\alpha(<\pi/2)$ given by $\tan\alpha = |A/B|$ and in practice we normally have $\tan\alpha \ll 1$.

Consider first the change in gravitational potential energy resulting from interchange of parcels of air in the manner of subsection 1*a*. The result is no longer independent of the y-displacement. If the direction of displacement lies outside the acute angle α, there is an increase of energy, but if it lies inside, energy is released. There is zero change for displacement either along the isentropic surfaces or horizontally, and calculation shows that maximum release of energy occurs (approximately, assuming $\alpha \ll \pi/4$) for displacement in the direction of the bisector of α (in the y, z plane), which we shall call the s-axis. Now consider the change in kinetic "potential" energy. If the overturning were in the vertical plane, this would occur in the manner of subsection 1*b*, with increase of energy. But if overturning occurs in the x, s plane (*i.e.*, quasi-horizontally), with the perturbations varying harmonically in the x-direction, there is no change in the mean motion and no change of energy (correct to the appropriate order of small quantities). Hence our energy method gives instability in *all* cases and on calculation:

$$\theta_1^2 \leq -g_s B_s \approx \frac{1}{4}\frac{f^2}{h^2} \qquad (h^2 >> 1),$$

where g_s and B_s are the components of gravity and entropy gradient, respectively, along the s-axis (note analogy to 1*a*) and h^2 is the Richardson number defined by $h^2 = gB/(dV_x/dz)^2$.

This result is of course provisional but, as in other cases, is verified by complete solution. With artificial (but physically possible) boundary conditions we can obtain nearly the maximum value of θ_1, but a more realistic model gives $\theta_1 \approx 0.31\, f/h$. The reduction in the coefficient from 0.5 to 0.31 is due to the additional constraints imposed by the boundary conditions, as a result of which displacements cannot everywhere be in the optimum s-direction: for one particular wave length (more precisely, for one ratio of horizontal to vertical "wave length"), the displacements are as near optimum as possible and it is to this dominant wave that the coefficient 0.31 applies. Longer waves grow more slowly while very short waves are stable. Thus there is "natural selection" for one particular wave structure. It has been shown elsewhere how, by considering compound systems containing a region where h^2 is a minimum, realistic models of both nascent wave cyclones and long waves may be constructed. Briefly, the smaller disturbances develop in frontal regions, where cloud masses reduce the effective static stability and therefore also h^2. The long waves occupy the whole troposphere, and secondary modifications, caused by constraints associated with the variability of the Coriolis parameter, are then significant.

1*ab*. *Generalised Vertical-Overturning Instability.* We may consider from the point of view of *vertical* overturning an initial system similar to that of subsection 2*a*, but it will be convenient to generalise by supposing V_x to vary with y as well as with z. Then using the same general method as before, we obtain

$$2\theta_1^2 \leq -\left[gB + f\left(f - \frac{dV_x}{dy}\right)\right] + \sqrt{\left[gB + f\left(f - \frac{dV_x}{dy}\right)\right]^2 + 4\left[(gA)^2 - gBf\left(f - \frac{dV_x}{dy}\right)\right]},$$

where the surd has always to be taken as positive. This is the general formula for vertical overturning, including the examples previously given as special cases. If either $B < 0$ or $dV_x/dy > f$, the system is certainly unstable. If neither condition is satisfied, we require for instability $(gA)^2 > gBf(f - dV_x/dy)$, which is equivalent to $1/h^2 > [1 - (1/f)(dV_x/dy)]$. In the important special case when dV_x/dy vanishes, this condition becomes simply $h^2 < 1$, equivalent to the well-known condition of negative absolute vorticity in the isentropic surfaces.

2*b*. *Helmholtz Instability.* Consider the system of two barotropic air masses with uniform horizontal motion $V_x = U_1$ in one, and $V_x = U_2$ in the other, separated by a vertical "front" at the x, z plane. If the earth were not rotating, we could apply the well-known results of Helmholtz (complete solutions) which show this system to be unstable for *all* perturbation wave lengths, growth-rate being inversely proportional to wave length, but it is instructive to apply the energy method. Helmholtz's solutions involve only horizontal motion, associated with corrugation of the "front," so we need consider only *horizontal* overturning. The "potential" energy is entirely "kinetic" and the manner of its release is clear from the flow pattern, obtained by considerations of continuity and boundary conditions alone. Outside the y-limits to which the corrugations of the "front" extend, the mean motion is unaltered, but inside these limits there is a change in the

mean flow of each air mass (in opposite directions) resulting in a decrease in organised kinetic energy (see Fig. 1).

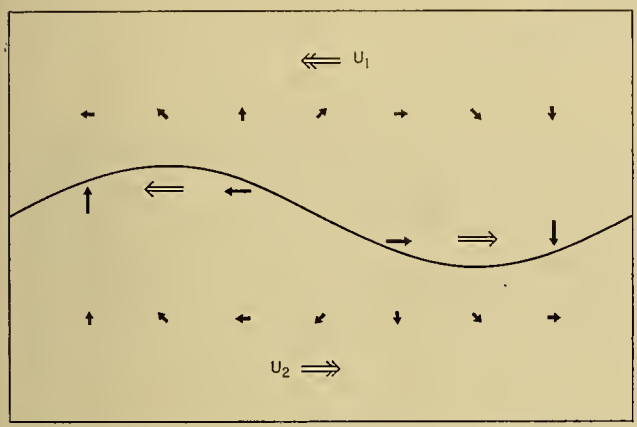


FIG. 1.—Helmholtz waves. Arrows with a single head and a single shaft denote perturbation velocities; those with a single head and double shaft, perturbation mean velocities; those with a double head and a double shaft, initial velocities.

Now, as we have seen, the earth's rotation has a stabilising effect only when there is *vertical* overturning. Hence we should expect results similar to those mentioned above when the earth's rotation is taken into account. Complete solution of the problem confirms this, and $\theta_1^2 = (\lambda^2/4)(U_1 - U_2)^2$, where $2\pi/\lambda$ is the wave length for a rotating, as well as for a nonrotating, system.

In practice, B is usually strongly positive so that any vertical motion would decrease the net release of "potential" energy. Moreover, frontal surfaces are not usually vertical and in general the boundary conditions cannot be satisfied by purely horizontal motion. Hence a sloping front is less unstable than a vertical one. Since the unstabilising effect of a velocity discontinuity is inversely proportional to the wave length, whereas the stabilising effect of static stability is independent of the scale of motion, it follows that only waves shorter than a critical wave length are unstable. Very short waves are always unstable because static stability may be neglected.

Future Developments

The discussion above is merely an outline of elementary principles. Although our analysis shows that the equations of dynamical meteorology are by no means intractable from the point of view of computing future developments, the results so far obtained are only of limited applicability. Our calculations give only the initial form of the most probable (dominant) new development and we need to compute the further development when the perturbations are no longer small. As the period of the forecast is extended, analytical methods become increasingly involved and clumsy and sooner or later we have to resort, at least partly, to numerical methods. An adequate degree of accuracy is practically attainable only with the use of computing machines, and electronic large-memory computers will play an important part in extending and generalising the elementary theory.

The development of numerical methods, even to the extent of a direct attack using observed data, does not absolve us from the necessity of understanding the precise significance of our solutions. Not only do we have to know how and where to approximate, but the reliability of our solutions varies with time, place, and forecast period. In fact for long forecast periods what is significant is not the detail, which is usually partially, perhaps entirely, accidental (*i.e.*, dependent on minutiae below the margin of error), but the general nature (*e.g.*, persistently settled or unsettled) of the majority of possible solutions. We need to develop the statistical theory referred to earlier. Not all the questions about future weather we should like to answer are in fact answerable, but it may well be that the growth of uncertainty in some directions is compensated by statistical regularity in others.

REFERENCES

1. BJERKNES, V., and others, *Hydrodynamique physique*. Paris, Presses Universitaires de France, 1934.
2. CHARNEY, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4:135–162 (1947).
3. EADY, E. T., "Long Waves and Cyclone Waves." *Tellus*, Vol. 1, No. 3, pp. 33–52 (1949).
4. LAMB, H., *Hydrodynamics*, 6th ed. Cambridge, University Press, 1932.

DYNAMIC FORECASTING BY NUMERICAL PROCESS

By J. G. CHARNEY

*The Institute for Advanced Study**

Introduction

As meteorologists have long known, the atmosphere exhibits no periodicities of the kind that enable one to predict the weather in the same way one predicts the tides. No simple set of causal relationships can be found which relate the state of the atmosphere at one instant of time to its state at another. It was this realization that led V. Bjerknes in 1904 [1] to define the problem of prognosis as nothing less than the integration of the equations of motion of the atmosphere. But it remained for Richardson [12] to suggest in 1922 the practical means for the solution of this problem. He proposed to integrate the equations of motion numerically and showed exactly how this might be done. That the actual forecast used to test his method was unsuccessful was in no sense a measure of the value of his work. In retrospect, it becomes obvious that the inadequacies of observation alone would have doomed any attempt however well conceived, a circumstance of which Richardson was aware. The real value of his work lay in the fact that it crystallized once and for all the essential problems that would have to be faced by future workers in the field and that it laid down a thorough groundwork for their solution.

For a long time no one ventured to follow in Richardson's footsteps. The paucity of the observational network and the enormity of the computational task stood as apparently unsurmountable barriers to the realization of his dream that one day it might be possible to advance the computation faster than the weather. But with the increase in the density and extent of the surface and upper-air observational network on the one hand, and the development of large-capacity high-speed computing machines on the other, interest has revived in Richardson's problem and attempts have been made to attack it anew.

These efforts have been characterized by a devotion to objectives more limited than Richardson's. Instead of attempting to deal with the atmosphere in all its complexity, one tries to be satisfied with simplified models approximating the actual motions to a greater or lesser degree. By starting with models incorporating only what are thought to be the most important of the atmospheric influences, and by gradually bringing in others, one is able to proceed inductively and thereby to avoid the pitfalls inevitably encountered when a great many poorly understood factors are introduced all at once.

A necessary condition for the success of this stepwise method is, of course, that the first approximations bear a recognizable resemblance to the actual motions. Fortunately, the science of meteorology has progressed to the point where one feels that at least the main factors governing the large-scale atmospheric motion are known. Thus integrations of even the linearized barotropic and thermally inactive baroclinic equations have yielded solutions bearing a marked resemblance to reality. At any rate, it seems clear that the models embodying in mathematical form the collective experience and the positive skill of the forecaster cannot fail utterly. This conviction has served as the guiding principle in the work of the meteorology project at The Institute for Advanced Study with which the writer has been connected.

The Geostrophic Approximation

In the selection of a suitable first approximation, Richardson's discovery that the horizontal divergence was an unmeasurable quantity had to be taken into account. Here a consideration of forecasting practice gave rise to the belief that this difficulty could be surmounted: forecasts were made by means of geostrophic reasoning from the pressure field alone—forecasts in which the concept of horizontal divergence played no part. And indeed, this belief was substantiated when it was shown by Charney [2] and Eliassen [7] that the geostrophic approximation could be used to reduce the equations of motion to a single dynamically consistent equation in which pressure appears as the sole dependent variable.

If, in addition to the geostrophic approximation, one makes the assumptions that frictional and nonadiabatic effects may be ignored, the equations of motion are found to be contained in the statement that the potential temperature and the potential vorticity, as defined by Rossby [13], are conserved, and that both the quasi-hydrostatic and the geostrophic approximations may be used for evaluating the terms involving density and horizontal velocity in the conservation equations. A partial justification for this procedure was given by the author [2], but the ultimate test must depend upon comparison of theory with observation.

The potential vorticity is defined as specific volume times the scalar product of the absolute vorticity and the entropy gradient. For the large-scale atmospheric motions the isentropic surfaces are quasi-horizontal, and we may to a first approximation replace the absolute vorticity and entropy gradient by their vertical components. The pressure equation then becomes [3]

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{f(\zeta_g+f)}{g\partial \ln\theta/\partial z}\left(\frac{\partial^2}{\partial z^2}+\alpha\frac{\partial}{\partial z}+\beta\right)\right]\frac{\partial p}{\partial t}=\gamma \quad (1)$$

in a rectangular coordinate system with x pointing

* Most of the work described in this article was performed under contract N6-ori-139, Task Order I, between the Office of Naval Research and The Institute for Advanced Study.

east, y north, and z upwards. The quantity f is the Coriolis parameter $2\Omega \sin \phi$, with Ω the angular speed of the earth's rotation and ϕ the latitude; p is the pressure; α, β, and γ are functions of p and its space derivatives; θ is the potential temperature, and ζ_g is the geostrophic vorticity:

$$\zeta_g \approx \frac{1}{\rho f}\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right) \equiv \frac{1}{\rho f}\nabla_h^2 p, \tag{2}$$

where ∇_h is the horizontal operator, ρ the density, and the symbol "$\approx$" denotes approximate equality. For simplicity of presentation, the curvature of the earth is here ignored.

Equation (1) is elliptic or hyperbolic in the pressure tendency according as the factor multiplying the z-derivative terms is positive or negative. Since this factor has the same sign as the approximate potential vorticity, $\rho^{-1}(\zeta_g + f)\partial \ln \theta/\partial z$, we may infer that the elliptic or hyperbolic character of (1) is preserved at a material point. Thus we are assured that if (1) is everywhere elliptic initially it will remain so. On the other hand, if (1) is elliptic at some points and hyperbolic at others it will retain this property. In the latter case the integration problem becomes exceedingly complicated. Moreover, it is not known if, or to what extent, the geostrophic approximation applies when the potential vorticity is negative. Fortunately, however, negative potential vorticities occur rarely, and then only locally, and we may therefore be justified in assuming at the start that equation (1) is elliptic everywhere and for all time. (In the more general case, where the isentropic surfaces are not assumed quasi-horizontal, it can also be proved that the pressure-tendency equation is and remains a second order elliptic differential equation if the potential vorticity is initially positive everywhere.)

We may envisage a solution of equation (1) by the following procedure. Assuming that p is known at time t, we evaluate its coefficients and nonhomogeneous terms at time t and solve for $\partial p/\partial t$ by a relaxation method (Southwell [14]). The boundary conditions are that there shall be no influx or efflux of mass through the bottom or top of the atmosphere, translated into equivalent conditions on the pressure tendency. Having obtained $\partial p/\partial t$ we calculate the pressure at time $t + \Delta t$ from

$$p(t + \Delta t) = p(t) + \Delta t\, \partial p/\partial t.$$

The sequence of steps is then iterated to give $p(t + 2\Delta t)$, $p(t + 3\Delta t)$, etc.

As computing machinery for calculations of such great complexity have not yet been available, more simplified atmospheric models have been devised to test the validity of some of the basic assumptions. These will now be described.

Advective Models

The General Case. A first simplification is obtained by ignoring the effect of vertical motion on the change in potential temperature. Although this assumption is more questionable than those already made, one cannot ignore the fact that the advective hypothesis has been used with a degree of success in present-day synoptic practice. Hence, in accordance with the plan of utilizing the results of synoptic experience in the construction of models, we shall investigate the form taken by the equations of motion under this hypothesis.

First, by way of further justification, we shall show by means of scale considerations that the term $u\partial \ln \theta/\partial x + v\partial \ln \theta/\partial y$ is usually larger in order of magnitude than $w\partial \ln \theta/\partial z$ in the adiabatic equation,

$$\frac{d \ln \theta}{dt} = \frac{\partial \ln \theta}{\partial t} + u\frac{\partial \ln \theta}{\partial x} + v\frac{\partial \ln \theta}{\partial y} + w\frac{\partial \ln \theta}{\partial z} = 0, \tag{3}$$

in which u, v, and w are the x, y, and z velocity components, respectively.

From the geostrophic and thermal wind equations we have

$$u\frac{\partial \ln \theta}{\partial x} + v\frac{\partial \ln \theta}{\partial y} \approx \frac{f}{g}\left(u\frac{\partial v}{\partial z} - v\frac{\partial u}{\partial z}\right),$$

and from the vertical vorticity and continuity equations

$$\begin{aligned}\frac{1}{\zeta + f}\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)(\zeta + f) \\ \approx -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \approx \frac{1}{\rho}\frac{\partial(\rho w)}{\partial z} \sim \frac{\partial w}{\partial z}.\end{aligned} \tag{4}$$

The symbol "$\sim$" denotes equality in order of magnitude only.

Assuming a sinusoidal dependency, with wave lengths l_x, l_y, and l_z,

$$\begin{aligned} u &= U \exp\,[2\pi i(x/l_x + y/l_y + z/l_z)] \\ v &= V \exp\,[2\pi i(x/l_x + y/l_y + z/l_z)] \\ w &= W \exp\,[2\pi i(x/l_x + y/l_y + z/l_z)], \end{aligned}$$

where $U \sim V$ and $l_x \sim l_y$, and noting that each term on the left-hand side of (4) has the same order of magnitude, we obtain

$$\left|\frac{w\partial \ln \theta/\partial z}{u\partial \ln \theta/\partial x + v\partial \ln \theta/\partial y}\right| \sim \frac{l_z^2}{l_x^2}\frac{\nu_b^2}{\nu_i^2},$$

where $\nu_i \equiv f$ is the frequency of a horizontal inertial oscillation and $\nu_b^2 = g\partial \ln \theta/\partial z$ is the frequency of a buoyancy oscillation. For the large-scale motions in middle latitudes l_x, $l_y = 2 \times 10^6$ to 6×10^6 m; $l_z = 10^4$ m to 2×10^4 m; $\nu_i \approx 10^{-4}$ sec^{-1}; and $\nu_b \approx 10^{-2}$ sec^{-1}. Hence the above ratio varies between 0.03 and 1. For a typical large-scale system (l_x, $l_y = 4 \times 10^6$ m; $l_z = 1.5 \times 10^4$ m) its value is about 0.14. We note, however, that the advective hypothesis is less valid in the stratosphere where the buoyancy frequency ν_b is several times greater than 10^{-2} sec^{-1}.

The following derivation of the geostrophic-advective equations is due substantially to Fjørtoft [5], although it is much in the spirit of Sutcliffe's work [15].

If the thermal wind equation

$$\frac{\partial \mathbf{v}_g}{\partial z} \approx \frac{g}{f}\mathbf{k} \times \nabla_h \ln \theta, \tag{5}$$

in which $\mathbf{v}_g$ is the geostrophic wind and $\mathbf{k}$ is the vertical

unit vector, is integrated from the ground to the level z, fixed with respect to time, we obtain

$$\mathbf{v}_g = \mathbf{v}_{g0} + \frac{g}{f}\int_0^z \mathbf{k} \times \nabla_h \ln\theta \, dz \tag{6}$$

and, by differentiating with respect to t and taking the curl,

$$\frac{\partial \zeta_g}{\partial t} = \frac{\partial \zeta_{g0}}{\partial t} + \frac{g}{f}\nabla_h^2 \int_0^z \frac{\partial \ln\theta}{\partial t} dz, \tag{7}$$

since the variability of f may be ignored in the differentiation. To eliminate $\partial\zeta_{g0}/\partial t$ we proceed as follows: If we define

$$\bar{\alpha} \equiv \frac{\int_0^\infty \frac{\rho}{\zeta_g + f}\alpha \, dz}{\int_0^\infty \frac{\rho}{\zeta_g + f} dz}, \tag{8}$$

equation (7) becomes, by integration with respect to z,

$$\overline{\frac{\partial \zeta_g}{\partial t}} = \frac{\partial \zeta_{g0}}{\partial t} - \frac{g}{f}\overline{\nabla_h^2 A}, \tag{9}$$

where

$$A \equiv -\int_0^z \frac{\partial \ln\theta}{\partial t} dz. \tag{10}$$

Utilizing the relationship [2, eq. 68]

$$\overline{\frac{\partial \zeta_g}{\partial t}} = -\overline{\mathbf{v}_g \cdot \nabla_h(\zeta_g + f)}, \tag{11}$$

which may also be derived directly by vertical integration of (4), we eliminate $\partial\zeta_{g0}/\partial t$ between (7) and (9) to obtain

$$\frac{\partial \zeta_g}{\partial t} = -\overline{\mathbf{v}_g \cdot \nabla_h(\zeta_g + f)} + \frac{g}{f}(\overline{\nabla_h^2 A} - \nabla_h^2 A). \tag{12}$$

It now becomes convenient to replace z by p as the vertical coordinate. The geopotential $\Phi = gz$ may then be introduced as a dependent variable in place of p, and, with slight approximation, the local and horizontal derivatives may be replaced by derivatives in an isobaric surface. The latter are denoted by the subscript p. For the large-scale systems in which we are primarily interested (since only for these is the geostrophic approximation valid), f is usually several times greater than ζ_g, and the mean defined by (8) may be replaced by the simple average with respect to pressure. Then, since

$$\zeta_g \approx f^{-1}\nabla_p^2 \Phi,$$

equation (12) becomes

$$\nabla_p^2\left(\frac{\partial \Phi}{\partial t} + A - \bar{A}\right) = \bar{J}_p\left(\frac{1}{f}\nabla_p^2\Phi + f, \Phi\right), \tag{13}$$

where J_p is the Jacobian operator defined by

$$J_p(\alpha, \beta) \equiv \frac{\partial\alpha}{\partial x}\frac{\partial\beta}{\partial y} - \frac{\partial\alpha}{\partial y}\frac{\partial\beta}{\partial x}, \tag{14}$$

and

$$A \equiv -\frac{1}{g}\int_p^{p_0} \frac{\partial \ln\theta}{\partial t}\frac{dp}{\rho} \tag{15}$$

with p_0 the surface pressure. We now introduce the advective hypothesis,

$$\frac{\partial \ln\theta}{\partial t} \approx -\mathbf{v}_g \cdot \nabla_p \ln\theta, \tag{16}$$

so that

$$\begin{aligned} A &\approx \frac{1}{g}\int_p^{p_0} \mathbf{v}_g \cdot \nabla_p \ln\theta \frac{dp}{\rho} \\ &\approx \frac{1}{fg}\int_p^{p_0} J_p\left(\frac{\partial\Phi}{\partial p}, \Phi\right) dp, \end{aligned} \tag{17}$$

with the result that all quantities in (13) are expressed in terms of Φ and its derivatives.

The numerical integration of (13) is simpler than that of the general quasi-geostrophic equation. One solves the two-dimensional Poisson's equation

$$\nabla_p^2 S = \bar{J}_p \tag{18}$$

and determines the field of $\partial\Phi/\partial t$ from

$$\frac{\partial \Phi}{\partial t} = S - A + \bar{A}. \tag{19}$$

There are a number of methods, both analytic and numerical, by which (18) can be solved. The best for hand calculation is probably the relaxation method of Southwell [14] as it is rapid and can be used with any type of boundary. However this method is not always suitable for automatic machine computation as it makes large memory demands. It has been found best for a machine with a small internal and large external memory to use the analytic solution expressed in terms of a Green's function G:

$$S = \iint G(x, y, x', y')\bar{J}\, dx'dy', \tag{20}$$

where the integral extends over the forecast area and $\bar{J}$ is regarded as a function of x' and y'. The function G is a solution of the homogeneous equation $\nabla_p^2 G = 0$, satisfying the same condition as S on the boundary. Finally we mention that one may also use an "analogy" device, consisting of a physical system whose equilibrium state is governed by an equation having the same form as (18). The physical magnitude $\bar{J}$ appearing in (18) is varied at will and the magnitude S determined by measurement from the resulting equilibrium configuration.

A Simplified Version. While (13) is already a considerable simplification of the quasi-geostrophic equation, it still presents appreciable difficulties for numerical integration. Further simplifications will therefore be considered. The first is based on the observation that the isolines of temperature and potential temperature in large-scale systems are approximately parallel at all levels, that is,

$$\frac{1}{\rho}\nabla_p \ln\theta = \frac{\pi(p)}{\bar{\rho}}\nabla_p \ln\bar{\theta} \tag{21}$$

where the bar refers to a fixed isobaric surface $\bar{p}$. If $\bar{\mathbf{v}}_g$, $\bar{\rho}$, $\bar{T}$, and $\bar{\theta}$ are the velocity, density, temperature,

and potential temperature respectively at this level,

$$\mathbf{v}_g = \bar{\mathbf{v}}_g + \int_{\bar{p}}^{p} \frac{\partial \mathbf{v}_g}{\partial p}\, dp = \bar{\mathbf{v}}_g + \sigma \mathbf{v}', \tag{22}$$

where

$$\mathbf{v}' = -R\bar{T}f^{-1}\nabla_p \ln \bar{\theta} \times \mathbf{k}, \tag{23}$$

and

$$\sigma = [\mu(p) - \mu(\bar{p})]/\bar{p}; \qquad \mu(p) \equiv \int_{p}^{p_0} \pi(p)\, dp. \tag{24}$$

Similarly we find

$$A = (g\bar{\rho})^{-1}\mu(p)\mathbf{v}_g \cdot \nabla_p \ln \bar{\theta},$$
$$\bar{A} = (g\bar{\rho})^{-1}\bar{\mu}\,\bar{\mathbf{v}}_g \cdot \nabla_p \ln \bar{\theta},$$

where $\bar{\mu}$ is the pressure average of $\mu(p)$. If the level $\bar{p}$ is chosen so that $\mu(\bar{p}) = \bar{\mu}$, it is easily found that (13) becomes

$$\nabla_p^2 \frac{\partial \Phi}{\partial t} = J_p\left(\frac{1}{f}\nabla_p^2\Phi + f, \Phi\right) + (R\bar{T})^2\,\overline{\sigma^2} J_p\left(\frac{1}{f}\nabla_p^2 \ln\theta, \ln\theta\right) \tag{25}$$

at this level. The variations in $\ln\theta$ are determined from

$$\frac{\partial \ln\theta}{\partial t} = -\bar{\mathbf{v}}_g \cdot \nabla_p \ln\theta = \frac{1}{f} J_p(\ln\theta, \Phi). \tag{26}$$

The two equations above are sufficient to determine the motion. Their advantage lies in the fact that the problem has been reduced to a purely two-dimensional one: no vertical integrations are required and the initial data need consist only of the height and temperature distribution in the surface $\bar{p}$.

An empirical study of the function $\mu(p)$ gave for $\overline{\sigma^2}$ the approximate mean value 0.14, and the value 600 mb for $\bar{p}$ in winter. In a specific situation $\overline{\sigma^2}$ and $\bar{p}$ seemed to be somewhat influenced by the height of the tropopause, indicating that the position of the tropopause exerts an influence at least on the kinematics of the atmospheric motion.

The Barotropic Model. The barotropic model results from a further simplification of (25). If one assumes that the thermal wind $\sigma\mathbf{v}'$ in (22) is proportional to $\bar{\mathbf{v}}_g$, implying that the isobars are parallel at all heights,

$$\mathbf{v}_g = \left(1 + \sigma\left|\frac{\mathbf{v}'}{\bar{\mathbf{v}}_g}\right|\right)\bar{\mathbf{v}}_g, \tag{27}$$

and equation (25) becomes

$$\nabla_p^2 \frac{\partial \Phi}{\partial t} = J_p\left(\frac{1}{f}\nabla_p^2\Phi + f, \Phi\right) + \left|\frac{\mathbf{v}'}{\bar{\mathbf{v}}_g}\right|^2 \overline{\sigma^2} J_p\left(\frac{1}{f}\nabla_p^2\Phi, \Phi\right) \tag{28}$$

at the level $\bar{p}$. This equation may also be written

$$\frac{\partial \zeta_g}{\partial t} + K\mathbf{v}_g \cdot \nabla_p \zeta_g + v_g \frac{df}{dy} = 0, \tag{29}$$

where

$$K = 1 + \left|\frac{\mathbf{v}'}{\bar{\mathbf{v}}_g}\right|^2 \overline{\sigma^2},$$

in which form it is seen to be practically identical with the so-called "equivalent-barotropic" equation derived by the author [3, p. 384]. The quantity $|\mathbf{v}'/\bar{\mathbf{v}}_g| \approx R\,|\nabla_p\bar{T}/\nabla_p\Phi|$ has been found to have the average value 1.25 and consequently $K = 1 + (1.25)^2(0.14) = 1.22$, which agrees well with the value $K = 1.25$ given in the author's paper just mentioned. Defining p^* by $\mathbf{v}_g(p^*) = K\bar{\mathbf{v}}_g$ and multiplying (29) by K, we obtain

$$\frac{\partial \zeta_g^*}{\partial t} + \mathbf{v}_g^* \cdot \nabla_p \zeta_g^* + v_g^* \frac{df}{dy} \approx \left(\frac{d}{dt}\right)_{p^*}(\zeta_g^* + f) = 0, \tag{30}$$

where the asterisks denote quantities at the level p^*. Hence from (4) we see that p^* is the level of nondivergence and that the vertical component of absolute vorticity $\zeta_g + f$ is conserved at this level. The value of p^* is about 100 mb less than that of $\bar{p}$, or about 500 mb.

The foregoing remarks reveal clearly the exact sense in which the barotropic atmosphere may be considered an approximation to the real atmosphere.

Kibel's Method. In 1940, I. A. Kibel [9] proposed a numerical method for forecasting surface pressure and temperature changes on the basis of the geostrophic hypothesis[1] by assuming that the lapse rate of temperature is everywhere constant, that is,

$$T(x, y, z) = T_0(x, y) - \gamma z, \tag{31}$$

and that the temperature T_H at a fixed upper level is advected with the wind:

$$\frac{\partial T_H}{\partial t} + \mathbf{v} \cdot \nabla T_H = 0. \tag{32}$$

The latter approximation holds at the tropopause where it is considered to be a consequence of the property that the tropopause is a discontinuity surface of the second kind at which the lapse rate of temperature is approximately zero. With $\mathbf{v} = \mathbf{v}_g + \mathbf{v}'$, where $\mathbf{v}'$ is the geostrophic deviation, the adiabatic equation at the ground becomes

$$\frac{\partial T_0}{\partial t} - \frac{RT_0}{fp_0} J(T_0, p_0) + \mathbf{v}_0' \cdot \nabla_h T_0 - \frac{RT_0}{c_p p_0}\left(\frac{\partial p_0}{\partial t} + \mathbf{v}_0' \cdot \nabla_h p_0\right) = 0, \tag{33}$$

where c_p is the specific heat of dry air at constant pressure. Similarly, equation (32) becomes

$$\frac{\partial T_0}{\partial t} - \frac{RT_H}{fp_H} J(T_0, p_H) + \mathbf{v}_H' \cdot \nabla_h T_0 = 0. \tag{34}$$

Noting that

$$p = p_0 \exp\left(-\frac{g}{R}\int_0^z \frac{dz}{T_0 - \gamma z}\right), \tag{35}$$

we obtain the thermal wind equation

$$\frac{\nabla_h p}{\rho} = \frac{RT_H}{p_0}\nabla_h p_0 + \frac{gz}{T_0}\nabla_h T_0 \tag{36}$$

1. A somewhat simplified account of Kibel's method by B. I. Isvekov has been translated into English. See "Professor I. A. Kibel's Theoretical Method of Weather Forecasting." *Bull. Amer. meteor. Soc.*, 27:488–497 (1946).

and the relationship

$$\frac{1}{\rho}\frac{\partial p}{\partial t} = \frac{RT_H}{p_0}\frac{\partial p_0}{\partial t} + \frac{gz}{T_0}\frac{\partial T_0}{\partial t}, \tag{37}$$

from which it follows that (34) may be written

$$\frac{\partial T_0}{\partial t} - \frac{RT_H}{fp_0} J(T_0, p_0) + \mathbf{v}'_H \cdot \nabla_h T_0 = 0. \tag{38}$$

If the winds are assumed geostrophic, (33) and (38) combine to give

$$\left.\begin{aligned} \frac{\partial p_0}{\partial t} &= -\alpha J(T_0, p_0) \\ \frac{\partial T_0}{\partial t} &= \beta J(T_0, p_0) \end{aligned}\right\}, \tag{39}$$

where

$$\alpha = \frac{c_p}{f}\left(1 - \frac{T_H}{T_0}\right); \qquad \beta = \frac{RT_H}{fp_0}.$$

These are Kibel's equations for the first approximation. Their integration becomes especially simple if it is observed that in virtue of (39)

$$\frac{\partial}{\partial t}(\alpha T_0 + \beta p_0) = 0,$$

so that

$$\theta = \alpha T_0 + \beta p_0 \tag{40}$$

is independent of time, and

$$\left.\begin{aligned} \frac{\partial p_0}{\partial t} &= J(p_0, \theta) \\ \frac{\partial T_0}{\partial t} &= J(T_0, \theta) \end{aligned}\right\}. \tag{41}$$

These equations imply that the fields p_0 and T_0 are propagated along lines of constant θ with a velocity numerically equal to $|\nabla_h \theta|$. They may also be interpreted to state that the surface pressure and temperature are steered along the contour lines of the isobaric surface located at approximately the height $Z = \alpha T_0/(\alpha\gamma + g\beta\rho_0)$ and with a velocity given by the wind at this level multiplied by $(\alpha\gamma + g\beta\rho_0)f/g$. In practice, the constants α and β are determined empirically rather than from their formulas.

In 1948 Lettau [10] observed that Kibel's equations of first approximation are very similar to a set derived by Exner in 1906 [8] under similar assumptions. In place of Kibel's tropopause condition, Exner assumed that there exists a level H at which the pressures do not change. If H is identified with Z, Exner's equations become nearly identical with Kibel's.

Equations (39) give zero pressure tendency at a pressure center, where $\partial p_0/\partial x = \partial p_0/\partial y = 0$. For an estimate of the tendency at a center Kibel proceeds to a second approximation. As the derivation is omitted in both Kibel's and Isvekov's papers, it will be supplied here.

The geostrophic deviation

$$u' \approx -\frac{1}{f}\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right)$$

$$v' \approx \frac{1}{f}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right)$$

is further approximated by substituting the geostrophic wind components $u_g = -\rho^{-1}f^{-1}(\partial p/\partial y)$, $v_g = \rho^{-1}f^{-1}(\partial p/\partial x)$ for u and v. Ignoring small terms and using (36) and (37), one obtains, when $\partial p_0/\partial x = \partial p_0/\partial y = 0$,

$$\left.\begin{aligned} u'_H - u'_0 = &-\frac{R(T_H - T_0)}{f^2 p_0}\frac{\partial^2 p_0}{\partial x \partial t} - \frac{gH}{f^2 T_0}\frac{\partial^2 T_0}{\partial x \partial t} \\ &- \frac{g^2H^2}{f^3 T_0^2} J\left(T_0, \frac{\partial T_0}{\partial x}\right) - \frac{gHRT_H}{f^2 p_0 T_0} J\left(T_0, \frac{\partial p_0}{\partial x}\right), \\ v'_H - v'_0 = &-\frac{R(T_H - T_0)}{f^2 p_0}\frac{\partial^2 p_0}{\partial y \partial t} - \frac{gH}{f^2 T_0}\frac{\partial^2 T_0}{\partial y \partial t} \\ &- \frac{g^2H^2}{f^3 T_0^2} J\left(T_0, \frac{\partial T_0}{\partial y}\right) - \frac{gHRT_H}{f^2 p_0 T_0} J\left(T_0, \frac{\partial p_0}{\partial y}\right) \end{aligned}\right\}. \tag{42}$$

Evaluating $\partial^2 p_0/\partial x \partial t$ and $\partial^2 p_0/\partial y \partial t$ from (39) and substituting in (33) and (38) we obtain, by elimination of $\partial T_0/\partial t$,

$$\frac{\partial p_0}{\partial t} = B\left[\frac{\partial T_0}{\partial x} J\left(T_0, \frac{\partial \tilde{\theta}}{\partial x}\right) + \frac{\partial T_0}{\partial y} J\left(T_0, \frac{\partial \tilde{\theta}}{\partial y}\right)\right], \tag{43}$$

where

$$\left.\begin{aligned} \tilde{\theta} &= \tilde{\alpha} T_0 + \beta p_0 \\ \tilde{\alpha} &= \left(\frac{\gamma_a}{\gamma}\right)^2 \alpha \Big/ \left(2\frac{\gamma_a}{\gamma} + \frac{T_0}{T_H} - 1\right) \\ B &= \frac{c_p p_0 \alpha}{RT_0 f}\left(2\frac{\gamma_a}{\gamma} + \frac{T_0}{T_H} - 1\right) \end{aligned}\right\}, \tag{44}$$

and γ_a is the dry-adiabatic lapse rate. Making the approximation $\tilde{\theta} \approx \theta$, it is possible to deduce the consequence that the surface pressure decreases when the θ contours at the steering level Z diverge over the center in the direction of the wind. Conversely, if the contours converge, the surface pressure increases.

In view of the nature of the assumptions made, Kibel's first and second approximations cannot serve as a substitute for the general advective geostrophic equations. There have been references to higher approximations, but the articles unfortunately were not available to the writer.

The Linearized Barotropic Model

The nonlinearity of the quasi-geostrophic equations makes it difficult to study their properties. However, many of the essential properties of the nonlinear motions are preserved in the linearization. Thus, for example, boundary conditions will often be qualitatively the same for both and the numerical technique of integration used for the one can be used for the other. In this manner it is often possible to obtain semi-

quantitative estimates of the convergence criteria for the nonlinear difference equations—estimates that would be difficult if not impossible to obtain by any other means. Also, the study of linearized counterparts may be undertaken for the physical insight they give into the nature of the dynamical processes at work in the atmosphere—without which insight, numerical forecasting is at best a haphazard affair.

One problem to which at least a partial answer is required is the determination of the speeds at which influences propagate in the atmosphere. The elliptic character of (1) demands that the boundary conditions be known as functions of time along vertical surfaces surrounding the forecast region, as well as at the ground and at the "top" of the atmosphere. Since such conditions cannot be known, there appears to be a basic indeterminacy in the problem of forecasting for a limited region. The reason for this indeterminacy is that the introduction of the geostrophic approximation effectively filters out the sound and gravity waves whose finite speeds place an upper limit on the rate of propagation of a disturbance. Thus disturbances propagate instantaneously in quasi-geostrophic motion, and effects from the boundaries are felt immediately at every point in the region they enclose. One feels that since sound and gravity motions interact negligibly with the large-scale systems, the solution to the difficulty must be found in the quasi-geostrophic equations themselves. Let us, therefore, examine these equations in their linearized form.

As we shall be concerned primarily with horizontal propagation it will be permissible to treat the motion barotropically, for it has already been shown that the horizontal motion of a barotropic atmosphere approximates the actual motion at a certain mean level.

For small perturbations on a constant zonal current U, the barotropic equation (30) becomes

$$\nabla_h^2\left(\frac{\partial\Phi}{\partial t}+U\frac{\partial\Phi}{\partial x}\right)+\beta\frac{\partial\Phi}{\partial x}=0. \tag{45}$$

The quantity $\beta = df/dy$ will be given the constant value corresponding to the mean latitude 45°. Equation (45) admits of the plane wave solution

$$\Phi = \exp\left[i(kx+\mu y-\nu t)\right] \tag{46}$$

in which the frequency ν is related to the wave numbers $k = 2\pi/l_x$ and $\mu = 2\pi/l_y$ by

$$\nu = kU-\frac{\beta k}{k^2+\mu^2}. \tag{47}$$

The group velocity components c_{gx} and c_{gy} are

$$\left.\begin{aligned} c_{gx} &= \frac{\partial\nu}{\partial k} = U+\beta\frac{k^2-\mu^2}{(k^2+\mu^2)^2}\\ c_{gy} &= \frac{\partial\nu}{\partial\mu} = \frac{2\beta k\mu}{(k^2+\mu^2)^2}\end{aligned}\right\}. \tag{48}$$

Now it may be shown that the kinetic energy E of a point disturbance obeys approximately the law

$$\frac{\partial E}{\partial t}+\frac{\partial}{\partial x}(c_{gx}E)+\frac{\partial}{\partial y}(c_{gy}E)=0, \tag{49}$$

which states that the energy of the disturbance associated with a given area in the x, y plane will not change with time when each point of the area moves with the local group velocity given by (48). Since an arbitrary initial disturbance may be regarded as a sum of point disturbances, we may state that no influence is propagated faster than the maximum group velocity. An inspection of (48) shows that for increasing l_x and l_y the group velocities as well as the phase velocities become arbitrarily large. However, the magnitudes of l_x and l_y are limited by the finiteness of the earth's surface, and even this limitation is scarcely realistic, for the extent of the motions with which we are concerned is far less than world-wide. To obtain a better estimate of the maximum group velocities, we first provide for the finiteness of the earth by assuming x to have a period equal to the circumference of the 45° latitude circle. If the unit of distance is taken as the radius of this circle and the unit of time as one day, k becomes an integer representing the number of waves encircling the earth and $\beta = 2\pi$.

A typical spectrum determined by Fourier analysis of the v component of the velocity at 45°N for 0300Z on January 11–13, 1946 at 500 mb is shown in Fig. 1,

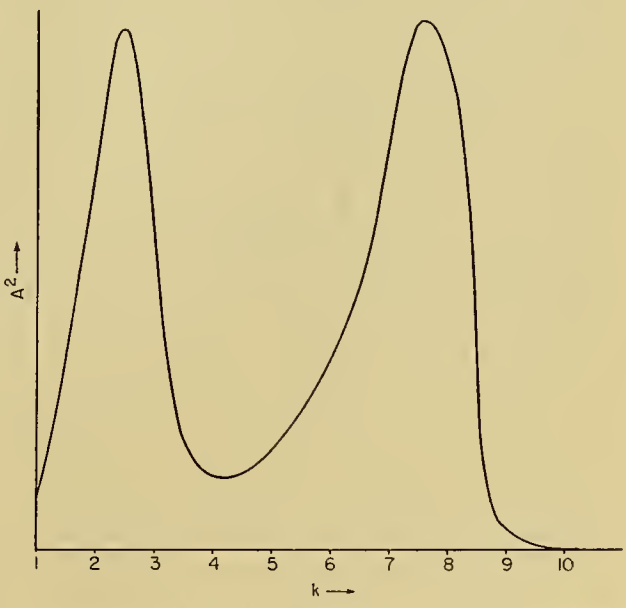


FIG. 1.—Mean spectral distribution of the kinetic energy of the north-south motion at 500 mb and 45° N for 0300Z, January 11, 12, and 13, 1946.

in which the square of the amplitude, A^2, measuring the kinetic energy of the north-south motion, is plotted against k. One energy maximum occurs at $k = 2.5$, corresponding to a wave length of 144° long. at 45° lat., and the other at $k = 7.5$, corresponding to a wave length of 48° long. The first maximum is associated with the very long wave quasi-stationary disturbances which, as Charney and Eliassen [4] have shown, are produced mainly by topographical action. As we are

not here concerned with these disturbances, we may assume a cutoff in the wave number at $k = 4$. The maximum value of c_{gx} then occurs at $\mu = 0$ and is equal to $U + \beta/k^2$ or $(U + 22.5)$ deg long. day^{-1}, and the minimum value occurs at $\mu^2 = 3k^2$ and is equal to $U - \beta/8k^2 = (U - 2.8)$ deg long. day^{-1}. Since it is also unrealistic to take $\mu = 0$, we shall suppose that $\mu \geq 4$ also, in which case max c_{gx} and min c_{gx} are equal to $U \pm \beta/8k^2 = (U \pm 2.8)$ deg long. day^{-1}. The corresponding values of max and min c_{gy} occur at $\mu = \pm k$ and are equal to ± 11.2 deg long. day^{-1} for k, $\mu \geq 4$.

These group velocity estimates can be regarded only as approximations in the asymptotic sense; they are accurate only for large values of time. Moreover, the derivation of the energy propagation equation (49) presupposes a continuous variation of frequency, whereas the frequencies are here restricted to discrete values by the finiteness of the earth. The following alternative approach to the problem avoids these difficulties.

If we assume a sine dependency on the y-coordinate ($\partial^2\Phi/\partial y^2 = -m^2\Phi$), equation (45) becomes

$$\frac{\partial^2}{\partial x^2}\left(\frac{\partial \Phi}{\partial t} + U\frac{\partial \Phi}{\partial x}\right) - m^2\frac{\partial \Phi}{\partial t} + (\beta - Um^2)\frac{\partial \Phi}{\partial x} = 0. \quad (50)$$

Its solution may be written [3]:

$$\Phi(x, t) = \Phi(x - Ut, 0) + \int_{-\pi}^{+\pi} I_{m^2}(x - Ut - \alpha, t)\Phi(\alpha, 0)\, d\alpha, \quad (51)$$

where

$$I_{m^2} = \frac{1}{2\pi}\sum_{-\infty}^{+\infty}(e^{-i\nu t} - 1)e^{ikx}, \quad (52)$$

and

$$\nu = -\frac{\beta k}{k^2 + m^2}. \quad (53)$$

The "influence function" $I_{m^2}(x, t)$ is shown in Fig. 2 for $m^2 = 18$ and $t = 1$. This function is seen to become very small as the absolute value of x is increased. Let us assume that it is negligible for $x > x_1$ and $x < -x_2$. Then the upper and lower limits of integration may be replaced by $x - Ut - x_1$ and $x - Ut + x_2$, respectively. This means that with respect to the mean zonal current, influences propagate eastward with a speed less than x_1 degrees per day and westward with a speed less than x_2 degrees per day. From the figure, one may estimate $x_1 \approx 25°$ and $x_2 \approx 40°$. These estimates for the influence velocities are greater than those obtained for the group velocities because no restrictions have been placed on the energy spectrum of the initial disturbance $\Phi(x, 0)$. Influences from only very long wave disturbances propagate rapidly, and it is likely that a disturbance with most of its energy lying in the middle and short wave length region of the spectrum would again be found to have influence velocities more in conformity with those given by the extreme group velocities.

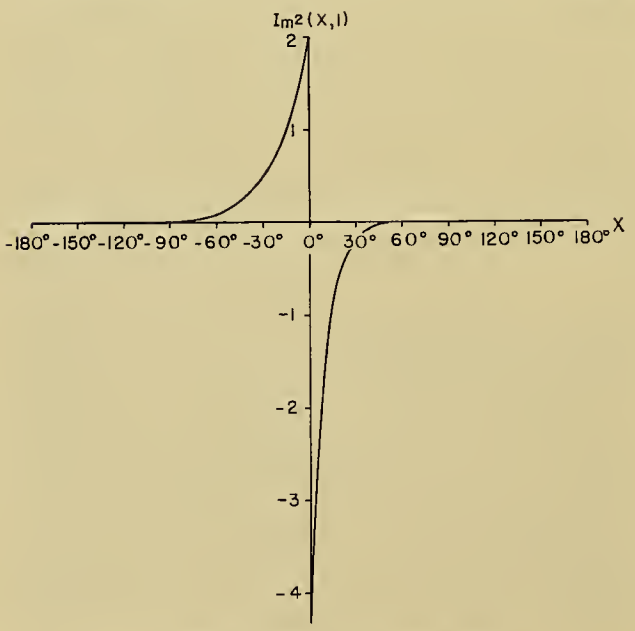


FIG. 2.—Influence function for $m^2 = 18$, $t = 1$.

In summary, one may say that influences have a maximum rate of propagation not much in excess of U (the particle speed) and a minimum rate not much less than U, so that the propagation relative to the ground is eastward. These conclusions have been borne out by several recent integrations of the nonlinear barotropic equations.

The solution for the linearized barotropic equations given by (51) has a certain value in itself. It has been used with some success for forecasting the actual wave-like perturbations of the westerlies [4]. The initial distribution of Φ along a fixed latitude was determined from the 500-mb map, and the integral in (51) was evaluated by means of Simpson's rule. It was shown that the topographical motions become quasi-stationary when $m^2 = 18$, so that the relative variation in the height of an isobaric surface at a fixed latitude could be predicted for a 24-hr period without taking these motions into account. Frictional effects were likewise found to be negligible in forecasts for so short a period. These results have been taken to justify the neglect of topography and friction in first attempts at the integration of the nonlinear barotropic equations.

It has also been shown [3] that influences are propagated at an effectively finite speed in the vertical direction. The maximum vertical group velocity for small plane wave disturbances in a baroclinic atmosphere with fronts perpendicular to the x, z plane has been calculated at approximately ± 4.5 km day^{-1} at 45° lat. Thus in principle it is unnecessary to know the entire vertical structure of the atmosphere initially in order to predict the motion of the lower layers for short periods. This result is perhaps of no great practical importance for numerical computation, since the low energy of the very high level motions renders their influence negligible in any case. It does, however, have a theoretical significance in stating a property of the atmospheric motion that has not generally been recognized. It is this property which may constitute the most serious criticism of the advective model. According to equation (19) the calculation of the geopotential

tendency $S - \bar{A}$ at a point on the ground requires a knowledge of the potential temperature advection throughout the entire vertical column above the point, in contradiction of what has just been stated; and here there is no compensation mechanism that effectively limits the spread of influences as in the case of the geostrophic approximation. It is thus possible that the advective assumption imposes a physically untenable constraint on the atmosphere.

Computational Stability

Once the physical problem of determining the equations of motion and the boundary conditions has been solved, there arises the purely mathematical problem of approximating the solution of the continuous equations by finite-difference methods. Here it may come as a surprising fact that no matter how small one chooses the space and time increments for the finite-difference equations, one has no guarantee that the finite-difference solution will approximate the continuous solution. This phenomenon, first discovered in 1928 by Courant, Friedrichs, and Lewy [6] will be illustrated for the one-dimensional wave equation

$$\frac{\partial^2 \varphi}{\partial t^2} = c^2 \frac{\partial^2 \varphi}{\partial x^2}, \tag{54}$$

whose finite-difference analogue for centered differences may be written

$$\frac{\varphi(x, t + \Delta t) + \varphi(x, t - \Delta t) - 2\varphi(x, t)}{(\Delta t)^2} = c^2 \frac{\varphi(x + \Delta x, t) + \varphi(x - \Delta x, t) - 2\varphi(x, t)}{(\Delta x)^2}. \tag{55}$$

Because of the linearity of (55) we may suppose that the solution is represented by a Fourier series. The long wave length components will be accurately expressed by the solution of (55), but there will inevitably be a distortion of the components whose wave lengths are of the order of Δx. This distortion will be harmless if Δx is small compared with the wave lengths of the physically relevant components and if the small wave length components do not amplify. However, when such an amplification occurs, it does so with exponential rapidity and quickly makes nonsense of the entire solution.

Let us, therefore, consider a Fourier component of the form $\exp[i(kx - \nu t)]$. Substitution in (55) gives

$$\frac{e^{-i\nu\Delta t} + e^{i\nu\Delta t} - 2}{(\Delta t)^2} = c^2 \frac{e^{ik\Delta x} + e^{-ik\Delta x} - 2}{(\Delta x)^2} = -\frac{4c^2}{(\Delta x)^2} \sin^2 \frac{k\Delta x}{2},$$

and, if we set $\omega = e^{-i\nu\Delta t}$,

$$\omega^2 - 2\left(1 - 2c^2 \left(\frac{\Delta t}{\Delta x}\right)^2 \sin^2 \frac{k\Delta x}{2}\right)\omega + 1 = 0. \tag{56}$$

The small error will cause no difficulties if it does not amplify, that is, if $|\omega| \leq 1$. If the two roots of (56) are complex, the square of their common absolute value is given by their product, which by (56) is equal to unity. If the roots are real, one of them must exceed unity in absolute value. Hence the condition for computational stability is that the roots be complex, that is, that the discriminant of (56) be negative. This leads to the condition

$$\frac{\Delta x}{\Delta t} \geq c \sin \frac{k\Delta x}{2},$$

or, since k must be presumed to have any value,

$$\frac{\Delta x}{\Delta t} \geq c. \tag{57}$$

Thus, no matter how small Δx and Δt are chosen independently, the solution of the finite-difference equation will not approximate the continuous solution unless (57) is satisfied. A geometrical interpretation of this criterion has been given by the author in a previous article [3].

Now it may be shown that whenever equations of motion permitting physical propagations are used, the condition (57) must be satisfied for c, the greatest propagation speed. Although Richardson [12] effectively excluded sound waves with the hydrostatic approximation, he did not exclude external gravity waves whose speeds are nearly as great as those of sound, about 300 m sec^{-1}. He chose Δx to be about 200 km; consequently his Δt should have been smaller than 200,000/300 sec or 20 min. In point of fact he chose 6 hr. Hence, a direct application of his method would inevitably have led to computational instability.

By employing the quasi-geostrophic equations one avoids the difficulty of having to choose time increments so small as to be meaningless in relation to the meteorologically significant motions. Nevertheless, the space and time increments cannot be chosen arbitrarily. To show this, let us again consider equation (45).

If we use centered differences, its finite-difference analogue may be written

$$D^2\left[\frac{\Phi(x, y, t + \Delta t) - \Phi(x, y, t - \Delta t)}{2\Delta t} + U\frac{\Phi(x + \Delta s, y, t) - \Phi(x - \Delta s, y, t)}{2\Delta s}\right] = \beta(\Delta s)^2 \frac{\Phi(x + \Delta s, y, t) - \Phi(x - \Delta s, y, t)}{2\Delta s}, \tag{58}$$

where Δs is the space increment and the finite-difference operator D^2 is defined by

$$D^2\psi \equiv \psi(x + \Delta s, y, t) + \psi(x, y + \Delta s, t) + \psi(x - \Delta s, y, t) + \psi(x, y - \Delta s, t) - 4\psi(x, y, t). \tag{59}$$

Consider a Fourier component of the form $\exp[i(kx + \mu y - \nu t] \equiv \omega \exp[i(kx + \mu y)]$. Substitution in (58) gives

$$\omega^2 + 2ai\omega - 1 = 0,$$

where

$$a = \left[U + \frac{\beta(\Delta s)^2/4}{\sin^2 \frac{k\Delta s}{2} + \sin^2 \frac{\mu \Delta s}{2}} \right] \frac{\Delta t}{\Delta s} \sin k\Delta s.$$

The condition for stability, $|\omega| \leq 1$, is then found to be $a \leq 1$, and we obtain

$$\frac{\Delta s}{\Delta t} \geq U \sin k\Delta s + \frac{(1/4)\beta(\Delta s)^2 \sin(k\Delta s)}{\sin^2 \frac{k\Delta s}{2} + \sin^2 \frac{\mu \Delta s}{2}}.$$

An upper bound for the first term on the right-hand side is U. The second term increases with decreasing k and μ, and we may evaluate it by assuming $k\Delta s$ and $\mu\Delta s$ to be small, thus obtaining the upper bound $\beta\Delta s k(k^2 + \mu^2)^{-1}$. Let us now assume that (58) is to be solved for a rectangular region with sides L_x and L_y parallel to the x- and y-axes, respectively. Then k and μ must be greater than $2\pi/L_x$ and $2\pi/L_y$, respectively, and the stability condition becomes

$$\frac{\Delta s}{\Delta t} \geq U + \frac{\beta \Delta s}{4\pi} \frac{L^2}{L_x}, \tag{60}$$

where L is the square harmonic mean,

$$[1/2(L_x^{-2} + L_y^{-2})]^{-\frac{1}{2}},$$

of L_x and L_y.

If Δs is small compared to the dimensions of the rectangular area, the second right-hand term will be small, and we obtain the result that the ratio of the space to the time increment must be greater than the particle velocity. It is shown in reference [5] that a stability condition closely analogous to (60) holds for the general nonlinear barotropic equation.

Integration of the Nonlinear Barotropic Equation

In the following section a description will be given of a method that has been used to integrate the barotropic equation (30).

Equation of Motion. The nondivergence level p^* is taken to be 500 mb. If the spherical surface of the earth is mapped conformally onto a plane and the geopotential Φ of the 500-mb surface is used as dependent variable, the correct form of (30) for a spherical earth becomes

$$m^2 \nabla_p^2 \frac{\partial \Phi}{\partial t} = m^2 J_p \left(\frac{m^2}{f} \nabla_p^2 \Phi + f, \Phi \right),$$

or

$$\nabla_p^2 \frac{\partial \Phi}{\partial t} = J_p \left(\frac{m^2}{f} \nabla_p^2 \Phi + f, \Phi \right), \tag{61}$$

where ∇_p^2 and J_p are to be interpreted as operators in the plane, and m is the scale factor of the mapping.

Except for the factor m^2/f, the left-hand side of (61) is the change in absolute vorticity and the right-hand side is the negative of the absolute vorticity advection. The equation therefore asserts that the vertical component of absolute vorticity is individually conserved.

The Conformal Map. The stereographic projection has the advantage that it is the map on which hemispheric data are usually plotted and analyzed and that, in comparison with the Mercator projection, it has much less distortion at high latitudes so that a square lattice represents more nearly equal areas on the earth. If ϕ is the latitude, a the radius of the earth, and the radius of the equator on the map is chosen as the unit of distance, the scale factor for the stereographic projection is $m = 2a^{-1}(1 + \sin\phi)^{-1}$.

Boundary Conditions. Since data are lacking for the entire earth and as the geostrophic approximation does not hold in the vicinity of the equator, the integration must be performed for a restricted area. The problem then arises of determining the boundary conditions. It is not a serious difficulty that these conditions are not actually known as functions of time, since we already know that influences from the boundary will not propagate inwards at a rate much greater than the particle velocity. One has only to choose the boundary sufficiently far from the area for which the forecast is desired. However, it is still important to assign artificial boundary conditions that are dynamically correct, for experience has shown that if the conditions assigned are not physically possible, computational instabilities arise which propagate faster than the physical disturbances.

If one wishes to solve (61) for the initial tendencies only, it is sufficient to prescribe $\partial\Phi/\partial t$ on the boundaries. This may be done by first making a crude forecast of $\partial\Phi/\partial t$, say by extrapolation from the observed 12-hr height change, or, more roughly, by setting $\partial\Phi/\partial t$ equal to 0. But in forecasting for a finite time interval, it is not sufficient to specify $\partial\Phi/\partial t$, and therefore Φ, for all time. For this case one may show [5] that if fluid is flowing into the forecast region both the flux of vorticity and Φ must be specified, whereas if fluid is flowing out of the region, only Φ need be prescribed. Since the specification of Φ determines the normal velocity at the boundary, one has only to specify the vorticity in addition to Φ where fluid is entering.

Method of Solution. The method to be followed in the solution of (61) will depend on the types of computing instruments available. It may be of some interest to describe briefly the integration procedure used by the writer in collaboration with R. Fjørtoft and J. von Neumann [5].

With the notation $\xi = \nabla_p^2 \Phi$ the basic equation (61) may be replaced by the system

$$\eta = m^2 f^{-1} \xi + f, \tag{62}$$

$$\partial\xi/\partial t = J_p(\eta, \Phi), \tag{63}$$

$$\nabla_p^2(\partial\Phi/\partial t) = \partial\xi/\partial t, \tag{64}$$

with the boundary conditions

$$\partial\Phi/\partial t = 0; \tag{65}$$

$$\partial\xi/\partial t = 0, \quad \text{if} \quad d\Phi/ds < 0, \tag{66}$$

where s is the tangential coordinate along the boundary, directed so as to keep the interior domain always on the right.

For simplicity in computation, a rectangular area with sides μ and ν is chosen, and a rectangular lattice is defined by the coordinates

$$x = (\mu/p)i, \qquad (i = 0, 1, \cdots, p),$$
$$y = (\nu/q)j \qquad (j = 0, 1, \cdots, q),$$

in which the boundary lines are $i = 0, i = p,$ and $j = 0$, $j = q$. If Φ_{ij}, ξ_{ij}, η_{ij}, J_{ij}, $(\partial\Phi/\partial t)_{ij}$, $(\partial\xi/\partial t)_{ij}$ denote the finite-difference approximations to Φ, ξ, η, J_p, $\partial\Phi/\partial t$, $\partial\xi/\partial t$ at the points (i, j), the procedure is as follows: From the initial values of Φ_{ij} one determines first ξ_{ij}, then η_{ij}, and finally J_{ij}. The values of $(\partial\xi/\partial t)_{ij}$ are then immediately found from (63):

$$(\partial\xi/\partial t)_{ij} = J_{ij}(\eta_{ij}, \Phi_{ij}).$$

To obtain $(\partial\Phi/\partial t)_{ij}$ one has to solve the finite-difference analogue of (64), that is,

$$D^2(\partial\Phi/\partial t)_{ij} = (\Delta s)^2(\partial\xi/\partial t)_{ij}, \tag{67}$$

where D^2 is defined by (59) and $\Delta s = \mu/p = \nu/q$ is the grid interval. On the assumption that Φ does not vary on the boundary $(\partial\Phi/\partial t = 0)$ the solution is found to be

$$\left(\frac{\partial\Phi}{\partial t}\right)_{ij} = -\frac{(\Delta s)^2}{pq}\sum_{l=1}^{p-1}\sum_{m=1}^{q-1}\sum_{r=1}^{p-1}\sum_{s=1}^{q-1} \cdot \left(\sin^2\frac{\pi l}{2p} + \sin^2\frac{\pi m}{2q}\right)^{-1} \cdot J_{rs}\sin\frac{\pi lr}{p}\sin\frac{\pi ms}{q}\sin\frac{\pi li}{p}\sin\frac{\pi mj}{q}. \tag{68}$$

In the actual computation, the functions $S_u = \sin(\pi u/p)$ and $T_v = \sin(\pi v/q)$ need only to be tabulated for integral values of u from 0 to $p/2$ and for v from 0 to $q/2$, the functions for other values of the arguments being given by

$$S_{-u} = S_{u+p} = -S_u; \qquad T_{-v} = T_{v+q} = -T_v.$$

This is a decided advantage in a machine with a limited storage capacity for numbers. Had an arbitrary area rather than a rectangular one been selected, the solution would be given by the finite-difference form of (20). Instead of storing the $(p + q)/2$ sines, one would then have to store the $p^2q^2/2$ independent values of a symmetric Green's function.

Having determined the quantities $(\partial\xi/\partial t)_{ij}$, $(\partial\Phi/\partial t)_{ij}$ the time extrapolation is performed by means of the formulas

$$\xi_{ij}^{t+\Delta t} = \xi_{ij}^{t-\Delta t} + 2\Delta t(\partial\xi/\partial t)_{ij}^{t},$$
$$\Phi_{ij}^{t+\Delta t} = \Phi_{ij}^{t-\Delta t} + 2\Delta t(\partial\Phi/\partial t)_{ij}^{t},$$

and the entire process is repeated a required number of times. In the first step, uncentered time differences must, of course, be used.

A number of 24-hr integrations have been carried out by this method on the Eniac[2] [5]. The space interval Δs was chosen to be eight degrees of longitude at 45° on the map and the time intervals 1, 2, and 3 hr were used on different occasions. The results indicated that the space interval was too large—about one-half its value seems to be recommended for future work—whereas even three hours was not too long for the time interval.

Figure 3 shows the results of one such integration. A strip two grid intervals in width at the top and side borders and one grid interval in width at the lower border of each diagram has been excluded to eliminate spurious boundary influences from the forecast. The forecast is seen to be fairly good in regions where adequate data were available. The major discrepancy occurred south of Greenland. In order to ascertain whether the discrepancy was due to the effect of baroclinicity, the 500-mb tendencies were calculated by means of the general advective equations (18) and (19), and to obtain a comparison with observation the computed tendency field was translated in the direction of the mean current for 24 hr with the speed of the trough to give a 24-hr change. The results are shown in Fig. 4.

Objective Analysis

After the initial data were assembled and put on punch cards, the time required for the Eniac to produce a 24-hr forecast using 2-hr time intervals was approximately 24 hr of continuous operating time. However, this time is likely to be considerably reduced by machines with a greater memory capacity. It has been estimated that the time required for the electronic computer at The Institute for Advanced Study, with a memory capacity for 1024 forty-digit binary numbers, will be about ½ hr. It is thus not entirely quixotic to contemplate the preparation of numerical forecasts for practical use in the near future. It then becomes obvious that if the high speed and high capacity of the machines are to be used to greatest advantage, there must be an equally rapid method of preparing the data received by teletype, radio, and telegraph in a form accessible to the machine. Under present conditions the data must first be plotted and subjectively analyzed, both of which operations are now painfully slow. J. von Neumann and H. A. Panofsky have therefore suggested that weather reports be translated into initial data by purely objective methods. As Panofsky [11] has shown, the pressure field may be approximated by an nth order polynomial of the form

$$p(x, y) = \sum_{i,j} a_{ij}x^iy^j \qquad (i + j \leq n)$$

by the method of least squares, provided the number of coefficients a_{ij} is less than the number of points at which p is known. The degree of the polynomial is determined by the amount of smoothing desired. For relatively small areas a third-degree polynomial is held to be adequate.

One can easily envisage a process whereby the machine is instructed in advance to take note of the data at a set of fixed observation points, to perform the

2. The Electronic Numerical Integrator and Computer of the Ballistic Research Laboratories, U. S. Army Proving Ground, Aberdeen, Maryland.

polynomial interpolation, and to calculate the initial data at the points of a predetermined grid. Given the mechanical means for the accomplishment of this program, there would, of course, remain other problems to be solved, not the least of which would be the devising of an objective technique for the location and elimination of errors in the raw data.

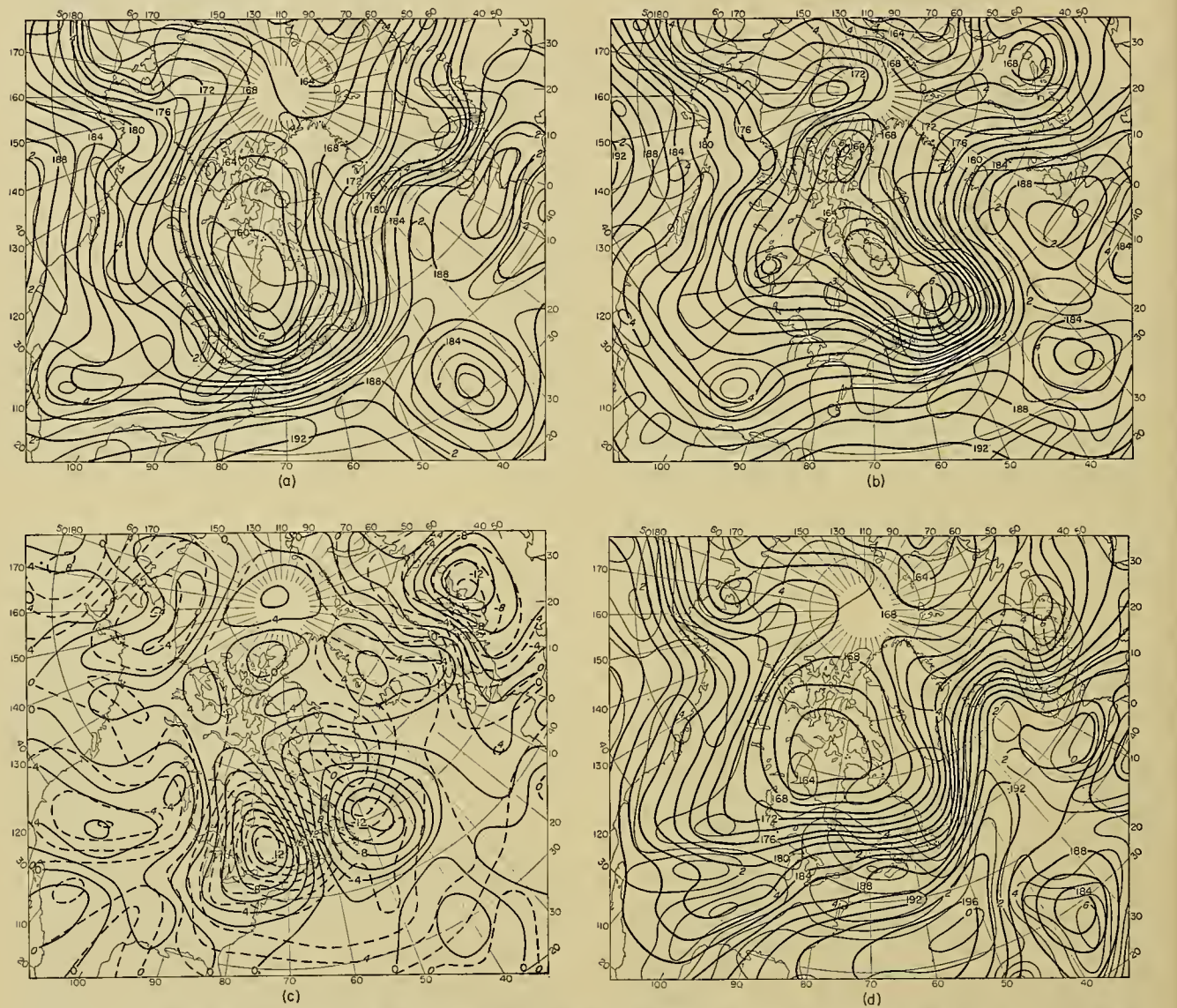


FIG. 3.—Forecast of January 30, 1949, 0300 GMT: (*a*) observed z (heavy lines) and $\zeta + f$ (light lines) at $t = 0$; (*b*) observed z and $\zeta + f$ at $t = 24$ hr; (*c*) observed (continuous lines) and computed (broken lines) 24-hr height change; (*d*) computed z and $\zeta + f$ at $t = 24$ hr. The height unit is 100 ft and the unit of vorticity is $\frac{1}{3} \times 10^{-4}$ sec^{-1}.

Use of the Primitive Equations

The discussion so far has dealt exclusively with the quasi-geostrophic equations as the basis for numerical forecasting. Yet there has been no intention to exclude the possibility that the primitive Eulerian equations can also be used for this purpose. The outlook for numerical forecasting would indeed be dismal if the quasi-geostrophic approximation represented the upper limit of attainable accuracy, for it is known that it applies only indifferently, if at all, to many of the small-scale but meteorologically significant motions. We have merely indicated two obstacles that stand in the way of the application of the primitive equations: First, there is the difficulty raised by Richardson that the horizontal divergence cannot be measured with sufficient accuracy; moreover, the horizontal divergence is only one of a class of meteorological unobservables which also includes the horizontal acceleration. And second, if the primitive Eulerian equations are employed, a stringent and seemingly artificial bound is imposed on the size of the time interval for the finite-difference equations. The first obstacle is the more formidable, for the second only means that the integration must proceed in steps of the order of fifteen minutes rather than two hours. Yet the first does not

seem insurmountable, as the following considerations will indicate.

It is very probable that the state of the motion at any given time will depend continuously upon the

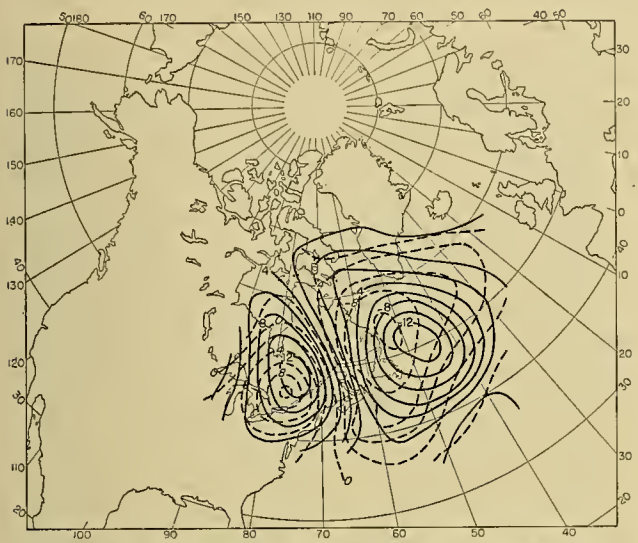


FIG. 4.—The broken lines represent the 24-hr height change computed by translating the baroclinically computed tendency field for 0300 GMT, January 30, 1949 in the direction of the mean current and with the speed of the trough. The solid lines represent the observed 24-hr height change.

initial motion; if the initial conditions are varied slightly, the motion at a subsequent time will vary slightly. Thus if the initial distribution of pressure, temperature, and horizontal wind is known to a relatively high degree of accuracy, the final distribution of these quantities will also be accurately determined and this will be true for whatever initial values are assigned to the unobservables.

To elaborate, let us consider two motions whose velocity components u, v, w and u', v', w' result from the slightly differing initial velocities u_0, v_0, w_0 and u_0', v_0', w_0'. If dm is the element of mass, the integral

$$I = \iiint \tfrac{1}{2}[(u - u')^2 + (v - v')^2 + (w - w')^2]\, dm$$

over the forecast region is a measure of error in the motion when u_0, v_0, w_0 are regarded as correct. We shall assume that u_0', v_0', w_0' are measured with such accuracy that I_0, the initial value of I, is small in comparison with the initial kinetic energy

$$\iiint \tfrac{1}{2}(u_0^2 + v_0^2 + w_0^2)\, dm.$$

Now let us assume that the differences $u - u'$, $v - v'$, $w - w'$ remain fairly small in comparison to u, v, w but allow for the possibility of a relative increase. The differences will then satisfy the linearized perturbation equations for the mean flow u, v, w and reduce to $u_0 - u_0'$, $v_0 - v_0'$, $w_0 - w_0'$ for $t = 0$. If this perturbation motion is stable so that the kinetic energy does not increase with time, the error I will also not increase and the motion u', v', w' will continue to be a good approximation. In practice, when forecasting for a *limited* region, the kinetic energy of the difference motion might conceivably increase as a result of work done at the boundaries or by advection of kinetic energy at the boundary. But in view of what is known about influence velocities one can say that it cannot possibly increase by more than the kinetic energy contained in that region, lying outside the forecast area, from which influences can reach the boundary in the forecast time. Hence, if the kinetic energy within this region is not many times larger than the kinetic energy within the forecast region, the above conclusions still apply. If the difference motion is unstable, the question of the utility of the primitive equations is not easily decided. It is certainly not obvious that dynamic instability necessarily implies their inutility. If, for example, the original motion is a small perturbation on an unstable steady flow with u, v, w, and u', v', $w' \sim e^{\nu t}$ (ν a positive number), the error I will increase exponentially; but the kinetic energy of the actual motion will also increase exponentially in the same proportion so that the relative error will remain constant.

J. C. Freeman and the writer[3] have carried out a numerical integration of the Eulerian equations for small perturbations in a barotropic atmosphere, taking care to satisfy the Courant-Friedrichs-Lewy stability criterion. Had the quasi-geostrophic equation been used, the motion would have consisted only of generalized Rossby waves; instead, the computed motion was found to consist of two superimposed parts, the first nearly identical to the Rossby motion and the second a gravitational wave motion of much smaller amplitude. Because of the necessarily erroneous values used for the acceleration and divergence (these were assumed to be identically zero initially), the changes in the velocity occurring after the first few time steps were quite incorrect. But there soon took place an adjustment of a kind that the motion at no time was found to deviate appreciably from the quasi-geostrophic. The function of the acceleration and divergence fields lay apparently in the mechanism by which the quasi-geostrophic balance was maintained. In a manner of speaking, the gravity waves created by the slight unbalance served the telegraphic function of informing one part of the atmosphere what the other part was doing, without themselves influencing the motion to any appreciable extent. Thus it mattered little that the initial error in the unobservables changed the character of the gravitational waves; only their existence was important.

The question arises whether the unobservables in the baroclinic atmosphere play an analogous role. This question is again not easily answered: in the case of barotropic motion, gravity waves are of the external type which move with the Newtonian velocity of sound; whereas in the baroclinic atmosphere they may be of the internal type—for which the speeds and frequencies are smaller—and might conceivably interfere with the large-scale motions.

In the last analysis, the feasibility of using the primitive equations will be decided only when numerical integrations have been carried out. For this purpose

3. Unpublished report, Meteorology Project, The Institute for Advanced Study, Princeton, New Jersey.

it is more convenient to use pressure as an independent variable in place of height. The complete equations of motion in this variable are given by Eliassen [7]. Under the hydrostatic and adiabatic assumptions they become

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla_p\mathbf{v} + \omega\frac{\partial \mathbf{v}}{\partial p} = -\nabla_p\Phi - 2\Omega\sin\phi\mathbf{k}\times\mathbf{v}, \quad (67)$$

$$\frac{\partial\Phi}{\partial p} = -\frac{1}{\rho}, \quad (68)$$

$$\nabla_p\cdot\mathbf{v} + \frac{\partial\omega}{\partial p} = 0, \quad (69)$$

$$\frac{\partial \ln\theta}{\partial t} + \mathbf{v}\cdot\nabla_p \ln\theta + \omega\frac{\partial \ln\theta}{\partial p} = 0, \quad (70)$$

where $\mathbf{v}$ is the horizontal velocity and $\omega = dp/dt$. The boundary conditions at the ground and at the upper limit of the atmosphere are, respectively,

$$\frac{d\Phi}{dt} = \frac{\partial\Phi}{\partial t} + \mathbf{v}\cdot\nabla_p\Phi - \frac{\omega}{\rho} = 0, \quad (71)$$

and

$$\frac{dp}{dt} = \omega = 0. \quad (72)$$

In the process of numerical computation, the time increment in $\mathbf{v}$ is determined from (67) and ω is found by integrating (69) and taking account of (72):

$$\omega = -\int_0^p \nabla_p\cdot\mathbf{v}\,dp. \quad (73)$$

The time increment in $\ln\theta$ is obtained from (70) and that of Φ is obtained as follows: Combining (73) with (71), we find for the surface height tendency

$$\left(\frac{\partial\Phi}{\partial t}\right)_0 = -\mathbf{v}_0\cdot(\nabla_p\Phi)_0 - \frac{1}{\rho_0}\int_0^{p_0}\nabla_p\cdot\mathbf{v}\,dp, \quad (74)$$

and from the relation

$$\frac{\partial \ln\theta}{\partial t} = -\rho\frac{\partial}{\partial p}\frac{\partial\Phi}{\partial t},$$

we derive by integration

$$\frac{\partial\Phi}{\partial t} = \left(\frac{\partial\Phi}{\partial t}\right)_0 + \int_{p_0}^p\left(\mathbf{v}\cdot\nabla_p\ln\theta + \omega\frac{\partial\ln\theta}{\partial p}\right)\frac{dp}{\rho}. \quad (75)$$

As $(\partial\Phi/\partial t)_0$ is known from (74), $\partial\Phi/\partial t$ is then immediately determined.

Since ω is very nearly equal to w, one would begin by assuming $d\mathbf{v}/dt$, $\nabla_p\cdot\mathbf{v}$, and ω to be 0 and then trust the compensation mechanism to yield a correct forecast. In six hours, the time interval used by Richardson, twenty-four time extrapolations of fifteen minutes each will have been performed, during which time there would presumably be ample opportunity for this mechanism to operate.

Nonadiabatic and Frictional Effects

Nonadiabatic and frictional effects have been ignored in the body of the discussion only because it was thought that one should first seek to determine how much of the motion could be explained without them. Ultimately they will have to be taken into account, particularly if the forecast period is to be extended to three or more days.

Condensation phenomena appear to be the simplest to introduce: one has only to add the equation of continuity for water vapor and to replace the dry by the moist-adiabatic equation. Long-wave radiational effects can also be provided for, since our knowledge of the absorptive properties of water vapor and carbon dioxide has progressed to a point where quantitative estimates of radiational cooling can be made, although the presence of clouds will complicate the problem considerably.

The most difficult phenomena to incorporate have to do with the turbulent transfer of momentum and heat. A great deal of research remains to be done before enough is known about these effects to permit the assigning of even rough values to the eddy coefficients of viscosity and heat conduction. Owing to their statistically indeterminate nature, the turbulent properties of the atmosphere place an upper limit on the accuracy obtainable by dynamical methods of forecasting, beyond which we shall have to rely upon statistical methods. But it seems certain that much progress can be made before these limits are reached.

REFERENCES

1. Bjerknes, V., "Das Problem von der Wettervorhersage, betrachtet vom Standpunkt der Mechanik und der Physik." *Meteor. Z.*, 21:1–7 (1904).
2. Charney, J. G., "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2 (1948).
3. —— "On a Physical Basis for Numerical Prediction of Large-Scale Motions in the Atmosphere." *J. Meteor.*, 6:371–385 (1949).
4. —— and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
5. Charney, J. G., Fjørtoft, R., and Neumann, J. v., "Numerical Integration of the Barotropic Vorticity Equation." *Tellus*, Vol. 2, No. 4, pp. 237–254 (1950).
6. Courant, R., Friedrichs, K., und Lewy, H., "Über die partiellen Differenzengleichungen der mathematischen Physik." *Math. Ann.*, 100:32–74 (1928).
7. Eliassen, A., "The Quasi-static Equations of Motion with Pressure as Independent Variable." *Geofys. Publ.*, Vol. 17, No. 3 (1949).
8. Exner, F. M., "Grundzüge einer Theorie der synoptischen Luftdruckveränderungen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 115:1171–1246 (1906).
9. Kibel, I. A., "Prilozhenie k meteorologii uravnenii mekhaniki baroklinnoi zhidkosti." *Izvestiia Akad. Nauk (SSSR), Ser. geogr. i geofiz.*, No. 5 (1940).
10. Lettau, H., "On Kibel's Method of Forecasting in Relation to Exner's Theory." *Bull. Amer. meteor. Soc.*, 29:201–202 (1948).
11. Panofsky, H. A., "Objective Weather-Map Analysis." *J. Meteor.*, 6:386–392 (1949).
12. Richardson, L. F., *Weather Prediction by Numerical Process*. Cambridge, University Press, 1922.
13. Rossby, C.-G., "Planetary Flow Patterns in the Atmosphere." *Quart. J. R. meteor. Soc.*, 66 (Supp.):68–87 (1940).
14. Southwell, R. V., *Relaxation Methods in Theoretical Physics*. Oxford, Clarendon Press, 1946.
15. Sutcliffe, R. C., "A Contribution to the Problem of Development." *Quart. J. R. meteor. Soc.*, 73:370–383 (1947).

ENERGY EQUATIONS

By JAMES E. MILLER

New York University

The concept of energy has two applications in meteorological problems: it facilitates the analysis of many atmospheric processes, and it makes possible a treatment of temperature changes associated with air motion. While many analyses that are facilitated by the energy concept can be accomplished without it, the relation between characteristics of the air motion and temperature changes following the motion is essentially an energy transformation and should be so treated.

The following discussion of atmospheric-energy equations is based primarily upon the work of Reynolds [11], with extensions and interpretations suggested by the work of many others, especially Margules [6, 7, 8], Richardson [12], and Ertel [4]. The mathematical analysis, which is presented in outline form here, can be found more fully developed, though with a few slight differences in approach, in a recent paper [9].

The Concept of Energy

When discussing energy one must define the system with which the energy is identified. A system may be the gas in a cylinder; it may be a certain interconnected series of rods, wheels, and gears in a mechanical device; it may be an electronic circuit; and it may be a part of a fluid. In the most general sense, a system is a portion of space with prescribed boundaries.

When a body moves through a distance d, opposite to the direction of a force F that is acting on the body, it is said to have done work equal in amount to the product $F \times d$. An insulated system is one that can do work or have work done on it, but is restrained from any other interaction with its environment. Energy, then, is that property of an insulated system which decreases when the system does work and increases when work is done on the system, the amount of increase or decrease being equivalent to the work done.

Different kinds of work are associated with changes of different forms of energy: When a gas system expands, its thermal energy decreases; when the gas is lifted in a gravitational field, its potential energy increases. The change of any form of energy in an insulated system can be identified from the work that accompanies the change. The change can be correlated with measurable properties of the system in a series of suitably controlled experiments and expressed as a mathematical function of those properties. It is assumed that the functions for various forms of energy, though established by a necessarily limited number of experiments, can be used to calculate energy changes in all circumstances—in noninsulated systems of any configuration and in complex processes where many different energy forms are changing simultaneously.

If the concept of energy went no further than the preceding discussion, it would have no value whatsoever; "energy change" would simply be a synonym for "work." There is one additional characteristic of energy that gives the concept its real importance. No exception has ever been found to the rule that energy changes can be expressed in terms of changes of the state parameters of a system, such as temperature, velocity, and position, regardless of how those changes take place. Thus, a change of thermal energy can be expressed in terms of the temperature change, and has the same value regardless of the intermediate states of the system. A similar statement cannot, in general, be made about work.

How, then, can the change of energy, as defined, be independent of intermediate states whereas work is not? It is true that, for an insulated system, both energy change and total work are completely determined by the initial and final states of the system. But for a system that can interact with its environment in other ways besides the performance of work, part of the energy change may be associated with a flow of energy into or out of the system. A gas that is being heated while it is expanding against external pressure will experience a change of thermal energy determined by its initial and final temperatures, while the work and the energy added through the boundary may have many different values, so long as their sum equals the energy change in the system.

Now if a system is restrained from the performance of work, but is not restrained from other external actions, a certain form of energy may flow through the boundary into the system. The increase in the amount of that energy form present in the system can be detected from changes of the parameters that enter into the corresponding energy function. Such changes, associated with energy flux alone, can be correlated with measurable properties on the boundary surface in a series of controlled experiments and expressed as mathematical functions of the boundary properties. It is assumed that these functions can be used to compute energy flux in all circumstances.

The preceding definitions of work, energy, and energy flux, and the empirical evidence that they can always be applied consistently to natural processes, lead to a general law of energy, which is expressed here in terms of changes with time:

$$\begin{bmatrix}\text{Rate of increase of energy in the system}\end{bmatrix} = \begin{bmatrix}\text{Rate at which energy is added through the boundary}\end{bmatrix} - \begin{bmatrix}\text{Rate at which work is done by the system}\end{bmatrix}. \quad (1)$$

The First Law of Thermodynamics

The special form of the energy law that is most commonly used in meteorological analysis is the "first law of thermodynamics":

$$\frac{d}{dt}(c_v T + L) = Q^I + Q^L + \alpha p_{ki} \frac{\partial v_i}{\partial x_k}. \quad (2)$$

Written in this way, the law can be applied to a very small system containing moist air and moving at the velocity of the air. The air is subject to viscous forces, but its thermal-energy function is assumed to be the same as that for a perfect gas. The operator d/dt is the rate of change following the motion of the system or the air; c_v is the specific heat of the mixture at constant volume; T is the absolute temperature; and $d(c_vT)/dt$ is the rate of change of thermal energy per unit mass. The rate of change of latent-heat energy, per unit mass, of water within the system is dL/dt. It is not necessary to establish a zero point for either thermal energy or latent-heat energy, because only their rates of change appear in (2).

The term Q^I on the right-hand side of equation (2) represents the rate at which thermal energy, per unit mass, is added by radiation or molecular conduction through the boundary of the system; and Q^L represents the rate at which latent-heat energy, per unit mass, is added by the net flux of water through the boundary.

The last term represents the rate of work, per unit mass, done on the system by relative motion against the molecular stresses. It is written with the aid of tensor notation which, despite its convenience in a term of this nature, is not widely used in meteorology. The indices i and k may have the values 1, 2, or 3, corresponding to the three directions of a Cartesian coordinate system; thus, x_i or x_k may be x_1, x_2, or x_3. The velocity components v_i or v_k may be v_1, v_2, or v_3, corresponding to the directions of the x_1, x_2, or x_3 axes, respectively. The general derivative $\partial v_i/\partial x_k$ stands for any one of nine different derivatives:

$$\frac{\partial v_i}{\partial x_k} = \frac{\partial v_1}{\partial x_1}, \frac{\partial v_1}{\partial x_2}, \frac{\partial v_1}{\partial x_3}, \frac{\partial v_2}{\partial x_1}, \frac{\partial v_2}{\partial x_2}, \frac{\partial v_2}{\partial x_3}, \frac{\partial v_3}{\partial x_1}, \frac{\partial v_3}{\partial x_2}, \frac{\partial v_3}{\partial x_3}.$$

The specific volume of the mixture in the system is α, and p_{ki} is the molecular stress tensor.

The components of p_{ki} are given by the expression

$$p_{ki} = p_{ik} = -\delta_{ik}\left(p + \frac{2}{3}\mu\frac{\partial v_j}{\partial x_j}\right) + \mu\left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i}\right), \quad (3)$$

where p is the pressure; δ_{ik} is a unit tensor of value 1 when $i = k$ and of value 0 when $i \neq k$; μ is the coefficient of molecular viscosity; and j is an index with the value 1, 2, or 3. In the tensor notation, repetition of an index within a term means that the term is summed over all values of the index. Thus,

$$\frac{\partial v_j}{\partial x_j} = \sum_{j=1}^{3}\frac{\partial v_j}{\partial x_j} = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3},$$

which is the three-dimensional velocity divergence. Any component of p_{ki} is a force per unit area, acting in the plane perpendicular to the x_k-axis. If it is regarded as acting on the fluid lying to the negative side of that plane, then the force is positive in the x_i-direction. If the velocity gradients are vanishingly small, or if the fluid is nonviscous ($\mu = 0$), p_{ki} reduces to the negative of the pressure, p:

$$p_{12} = p_{21} = p_{13} = p_{31} = p_{23} = p_{32} = 0,$$
$$p_{11} = p_{22} = p_{33} = -p.$$

Both i and k are repeated in the last term of (2), and hence

$$\alpha p_{ki}\frac{\partial v_i}{\partial x_k} = \sum_{\substack{i=1\\k=1}}^{3}\alpha p_{ki}\frac{\partial v_i}{\partial x_k}.$$

When this term is summed, expanded, and rearranged, there is obtained:

$$\alpha p_{ki}\frac{\partial v_i}{\partial x_k} = -\alpha p\frac{\partial v_j}{\partial x_j} + \mu\left[\begin{array}{c}\text{Sum of squares of terms}\\ \text{involving derivatives}\\ \text{of velocity components}\end{array}\right],$$

where $\partial v_j/\partial x_j$, the velocity divergence, is equivalent to $\delta_{ik}\,\partial v_i/\partial x_k$. According to the equation of continuity,

$$\frac{\partial v_j}{\partial x_j} = \frac{1}{\alpha}\frac{d\alpha}{dt}. \quad (4)$$

Therefore,

$$\alpha p_{ki}\frac{\partial v_i}{\partial x_k} = -p\frac{d\alpha}{dt} + \mu[\text{Sum of squares}]. \quad (5)$$

The first term on the right, $-p\,d\alpha/dt$, is the rate of work by expansion, and the second term is the rate of work by relative motion against viscous stresses. The latter term, Stokes' "dissipation function," is always positive because both μ and the sum of squares are positive.

The first law of thermodynamics has been established by many experiments, controlled and interpreted with a philosophy of the energy concept similar to that presented in the first part of this paper. The earliest statement of an energy law, by Leibnitz in 1693, dealt not with thermal energy but with kinetic and potential energies (the "live" and "dead forces"). This energy law can be established by a mathematical transformation of Newton's second law of motion, and its validity rests upon the validity of the law of motion.

The Law of Kinetic Energy

Newton's second law, for a compressible viscous fluid on the rotating earth, can be expressed as follows:

$$\left.\begin{aligned}\frac{dv_1}{dt} + (2\omega\cos\phi)v_3 - (2\omega\sin\phi)v_2 &= \alpha\frac{\partial p_{k1}}{\partial x_k},\\ \frac{dv_2}{dt} + (2\omega\sin\phi)v_1 &= \alpha\frac{\partial p_{k2}}{\partial x_k},\\ \frac{dv_3}{dt} - (2\omega\cos\phi)v_1 &= \alpha\frac{\partial p_{k3}}{\partial x_k} - g,\end{aligned}\right\} \quad (6)$$

where ω is the angular velocity of the earth's rotation; ϕ, the latitude; and g, gravity. The x_1-axis is directed toward east; the x_2-axis toward north; and the x_3-axis toward the zenith. When the first equation is multiplied by v_1, the second by v_2, the third by v_3, and the resulting equations are added, the Coriolis terms disappear, and the result is

$$\frac{d}{dt}\left(\frac{v_i^2}{2}\right) = \alpha v_i\frac{\partial p_{ki}}{\partial x_k} - v_3 g,$$

in which $v_i^2 = v_1^2 + v_2^2 + v_3^2$. The first term on the

right is recognized as the rate of work, per unit mass, done on a small volume of the fluid as a result of its motion in the direction of the net molecular stress acting on the volume. It can be transformed to

$$-\alpha p_{ki}\frac{\partial v_i}{\partial x_k}+\alpha\frac{\partial}{\partial x_k}(v_i p_{ki}),$$

and the last term, $-v_3 g$, can be written $-d(gx_3)/dt$, neglecting the variations of gravity. The equation then becomes

$$\frac{d}{dt}\left(\frac{v_i^2}{2}+gx_3\right)=-\alpha p_{ki}\frac{\partial v_i}{\partial x_k}+\alpha\frac{\partial}{\partial x_k}(v_i p_{ki}). \quad (7)$$

Note that, because of the summation convention, each term on the right is summed over the three values of i and k.

This equation, though established by a mathematical transformation of the equations of motion, is interpreted as a special form of the general energy law, equation (1). It will be applied to the same system to which the first law of thermodynamics is applied: a very small volume containing a mixture of air and water vapor. The term $v_i^2/2$ is known as the kinetic energy and gx_3 as the potential energy, both per unit mass. Such interpretation is consistent with the energy concept, for these functions are determined by the system's state and not by its history, and their total increase is equivalent to work done on the system.

For simplicity the following symbols will be used in writing the energy equations:

$$\left.\begin{aligned} I &= c_v T, \quad \text{the thermal energy} \\ K &= v_i^2/2, \quad \text{the kinetic energy} \\ P &= gx_3, \quad \text{the potential energy} \end{aligned}\right\}\ \text{per unit mass}$$
$$M=\alpha p_{ki}\frac{\partial v_i}{\partial x_k}. \quad (8)$$

Equations (2) and (7) then become

$$\frac{d}{dt}(I+L)=[Q^I+Q^L]_A-[-M]_W, \quad (9)$$

$$\frac{d}{dt}(K+P)=-\left[M-\alpha\frac{\partial}{\partial x_k}(v_i p_{ki})\right]_W. \quad (10)$$

The subscripts A and W indicate that the preceding terms in brackets represent energy added to the system or work done by the system, respectively. Since P is understood to mean g times a distance along a line directed opposite to gravity, there is no required orientation of the axes in equations (9) and (10).

The rate of work represented by M contributes to an increase of the sum of thermal and latent-heat energies, and simultaneously takes the same amount from the sum of kinetic and potential energies. It represents an energy transformation through the mechanism of molecular stress, and hence it will be called the molecular transformation function.

Energy equations that may be useful in meteorological problems will be derived from the basic equations (9) and (10) by purely mathematical methods, with the aid of certain simplifying assumptions about the physical properties of the moist air. Before this analysis is begun, a few remarks about the basic equations are in order.

Remarks on the Basic Energy Laws

Equations (6), upon which (10) is based, express Newton's second law of motion in a coordinate system that rotates with the earth. The rotational terms, such as $(2\omega\cos\phi)v_3$, vanish during the derivation of the energy equation, so that the velocity v_i and the rate of change d/dt in equation (10) may be measured in any Cartesian system, whether fixed, rotating, or moving at constant speed. It is customary to refer both (9) and (10) to a Cartesian coordinate system which is instantaneously fixed in the rotating earth, but which also follows the small air system so that two axes always lie in a plane tangent to the earth's surface. The magnitudes of terms in (10) may be different in other coordinate systems. For example, if a 1-g air particle, moving from east to west at 20 m sec^{-1} on the earth's equator, is brought to rest, it loses 2×10^6 ergs of kinetic energy in the commonly used coordinate system; but it gains 90×10^6 ergs in a nonrotating coordinate system. There is a compensating difference in the work in the two coordinate systems. In the former, the air particle performs work in the amount 2×10^6 ergs; in the latter, work is done on it in the amount 90×10^6 ergs.

The terms $\alpha\,\partial p_{ki}/\partial x_k$ in (6) are supposed to represent all body forces, per unit mass, on a small fluid element, with the exception of gravity. Whether they do or not depends upon the accuracy of the expressions for the stress tensor in (3). These expressions are presumed to be correct, but they have been verified experimentally for only a few special cases.

There is a further uncertainty about the equations that is of particular concern to the meteorologist. It is well known that the observed velocity of air motion varies with the type of instrument used. Fluctuations of the velocity in space and time cause a large, sluggish anemometer to react differently from a small, sensitive one. The same sort of scale effect is certainly present in measurements of temperature, pressure, and other properties. This situation poses a problem: How shall the properties be measured in order that the equations will be satisfied?

It can be shown logically that the equations (6), with the stress represented by (3), if correct at all, are strictly correct only in a certain scale domain with time and space periods greater than the molecular periods but smaller than the periods of the smallest turbulent fluctuations [11]. This conclusion presupposes that there is a lower limit to the periods of turbulent fluctuations of the various properties. If the conclusion is correct for the equations of motion, it is also correct for equation (10) and presumably for equation (9).

The latter equation, the first law of thermodynamics, is generally applied only to a stationary, homogeneous fluid whose changes of state take place very slowly. The

equation, which has been limited here to a system small enough to be considered homogeneous, will later be integrated for large, nonhomogeneous systems. Whether the fluid is stationary does not seem to make any difference. If it moves, so that the system contains kinetic energy, the changes of this energy form are completely accounted for in equation (10); and transformations of kinetic into thermal energy are accounted for by the transformation function M, common to both equations.

Equation (9) contains some terms that are not always found in expressions of the first law of thermodynamics. In the first place, the heat released in phase changes is frequently included in a general Q term, implying that the latent heat is supplied by some source outside the system. In equation (9) the latent-heat energy L and its net flux Q^L are written specifically, in order to make a clear distinction between energy transformations within the system and flux of energy across the boundary.[1]

In the second place, equation (9) includes, within the function M, not only the expansional work but also Stokes' dissipation function, which represents work done in relative motion against viscous forces. Since this function usually has a magnitude much smaller than that of expansional work, its neglect is of no consequence in most meteorological problems. But if it is omitted from a discussion of energy transformations, there can be no explanation of the frictional conversion of kinetic energy into thermal energy [14].

The General System

The basic energy equations can be integrated for a large, nonhomogeneous system, whose boundary surface has any shape and moves and changes shape in any prescribed manner, by the following procedure. Each equation is transformed to apply to a unit volume instead of a unit mass and is integrated over the volume of the large system, yielding an equation for the local rate of change of the energy. A term representing energy flux through the boundary surface, associated with the arbitrary and variable motion of the surface, is added to both sides of each equation. The equations then give the rate of change of energy following the large, irregularly shaped, irregularly moving system, the shape of whose boundary may also be changing. This procedure will be illustrated by the transformation of equation (10).

The equation of continuity (4) in the form

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_k}(\rho v_k) = 0,$$

where ρ is the density of the small volume of moist air, is multiplied through by $K + P$. Equation (10) is multiplied through by ρ and the resulting two equations are added; each term in the sum now refers to a unit volume. When this equation is integrated over the volume of the large system, there results

$$\frac{\partial}{\partial t}(K^* + P^*) = -\left[\int_\sigma v_n \rho(K + P)d\sigma\right]_A - \left[M^* - \int_\sigma v_i p_{ni}\, d\sigma\right]_W. \tag{11}$$

Here an asterisk indicates that a term has been multiplied by ρ and integrated over the volume; thus, $K^* = \int_v \rho K\, dV$. Two volume integrals on the right-hand side of the equation have been transformed to surface integrals by the divergence theorem. An element of the boundary surface is $d\sigma$; the component of air motion normal to the boundary and directed outward from the system is v_n ; and the three components of the stress tensor acting in a plane tangent to the boundary surface are p_{ni} .

Finally, the net flux of the energy $K + P$ into the system through the boundary surface, associated with the normal, outwardly directed velocity V_n of the surface, is written as a surface integral. The rate of change following the system is, then, the local change plus the net flux due to the motion of the system's boundary surface:

$$\frac{D}{Dt}(K^* + P^*) = \frac{\partial}{\partial t}(K^* + P^*) + \int_\sigma V_n \rho(K + P)d\sigma, \tag{12}$$

and (11) becomes

$$\frac{D}{Dt}(K^* + P^*) = \left[\int_\sigma u\rho(K + P)d\sigma\right]_A - \left[M^* - \int_\sigma v_i p_{ni}\, d\sigma\right]_W, \tag{13}$$

where u is equal to $V_n - v_n$, or the velocity of the boundary relative to the velocity of the air.

Similarly, equation (9) becomes

$$\frac{D}{Dt}(I^* + L^*) = \left[\int_\sigma [u\rho(I + L) - q_n^I - q_n^L]d\sigma\right]_A - \left[-M^*\right]_W, \tag{14}$$

where q_n^I and q_n^L are the components of the fluxes q_k^I and q_k^L of thermal energy and latent-heat energy at the boundary surface, and are directed normal to the surface out of the system. The fluxes q_k^I and q_k^L , whose units are energy per unit area and time, are defined in terms of Q^I and Q^L :

$$\rho Q^I = -\frac{\partial}{\partial x_k}(q_k^I) \quad \text{and} \quad \rho Q^L = -\frac{\partial}{\partial x_k}(q_k^L). \tag{15}$$

For example, the part of q_k^I associated with molecular conduction is $-c\,\partial T/\partial x_k$, where c is a constant, and the net inward flux per unit volume is $c\,\partial^2 T/\partial x_k{}^2$, a summation over the three values of k.

1. The term L can be expanded into $h_{vs}w_v + h_{ls}w_l$, where h_{vs} and h_{ls} are the latent heats of sublimation and freezing, and w_v and w_l are the mass proportions of water vapor and liquid water to the total mass of moist air. This method of representation is not completely satisfactory because the system, stipulated to consist of moist air only, may also contain some liquid water and ice.

Addition of (13) and (14) gives a total-energy equation:

$$\frac{D}{Dt}(I^* + L^* + K^* + P^*) = \left[\int_\sigma [u\rho(I + L + K + P) - q_n^I - q_n^L]d\sigma\right]_A - \left[-\int_\sigma v_i p_{ni} d\sigma\right]_W. \quad (16)$$

The important effect of adding the equations is the elimination of M^*, the only term on the right-hand side of the equations that depends upon processes within the system. By means of equation (16), it is possible to compute the change of total energy from measurements on the boundary, without any knowledge of processes within the system. If there is no work or energy flux at the boundary, the right side of (16) is zero, and the equation then expresses a law of conservation of energy.

The energy equations are sometimes derived in an order different from that followed here. The principle of conservation of energy is accepted as the basic law, and a total-energy equation in a form similar to (16) is written immediately. The law of kinetic energy is derived from the equations of motion, as is done here, and then the first law of thermodynamics follows upon subtraction of the kinetic-energy law from the total-energy law [15]. In contrast, the procedure used here is, first, the formulation of the law of kinetic energy and the first law of thermodynamics; and second, a statement of the total-energy law derived by combining the two. This procedure is considered more logical because one cannot write the necessary energy and work functions in the total-energy law without recourse to the experimental evidence embodied in the two special laws.

Effects of Averaging Processes

The scale domain of nonturbulent variations, the domain for which the energy equations are presumably valid, may be defined as follows: If instrumental measurements of a property are not appreciably affected by decrease in size and increase in response of the instrument, the property is being measured in a scale domain below the scale of the smallest turbulent fluctuations of that property.

Meteorological measurements of the various physical properties are made with instruments whose size and period of response are certainly greater than the size and period of the smallest turbulent fluctuations. Suppose several anemometers of varying size and response were mounted near each other in the atmosphere. The largest, most insensitive anemometer would indicate a fluctuation of velocity with time; the next more refined instrument would indicate more rapid fluctuations superimposed on those detected by the first instrument; and so on, down the scale. According to the postulate of a scale domain smaller than the smallest fluctuations, all anemometers smaller and more sensitive than a certain anemometer would give indications that were essentially identical in period and amplitude. The velocities indicated by these anemometers are the velocities that are supposed to satisfy the equations of motion and energy. Experimental studies suggest that the anemometers should be at least smaller than one centimeter in diameter and should react to velocity changes within at least one second [3].

Because the terms in the basic energy equations cannot be correctly evaluated by ordinary meteorological observations, it is no great exaggeration to say that the equations are not very useful for meteorological purposes. The equations can be rendered more useful by an analysis that will now be demonstrated; but, as will be seen, the analysis introduces new terms whose evaluation has long been a subject for speculation, postulation, and experimentation.

A property s will be separated into a mean value $\bar{s}$ and a deviation s' from the mean:

$$s = \bar{s} + s'. \quad (17)$$

The mean value is defined by a space-time integral:

$$\bar{s} = \frac{\int_V \int_t s\, dt\, dV}{Vt}, \quad (18)$$

where V is the volume of the space occupied by the instrument, and t is the period for which the instrument, because of its inertia, automatically averages the reading. If the mean value is defined arbitrarily for a selected volume and a selected period, the averaging can be carried out by numerical methods. Thus, any set of observations of s, at different points and different times, may be converted into a mean value over a selected volume and period by calculation of the integral in (18). Whether the averaging is done automatically by the instrument, or by numerical methods, or by both, $\bar{s}$ refers to a central point and time in V and t, and it is a continuous function of space and time. The values of $\bar{s}$ at nearby points and times are determined in overlapping volumes and periods.

The various properties s are replaced in the equations by $\bar{s} + s'$ and the individual terms of the equations are averaged by an integral of the form (18). It is assumed that $\overline{s'}$, the average value of s' for the same space-time as $\bar{s}$, is negligible; and that, therefore, $\bar{\bar{s}}$, which represents the average of the continuously varying $\bar{s}$ in the same space-time, is equivalent to the value of $\bar{s}$ at the center of the space-time.[2] Further, for the sake of simplification, it is assumed that $\rho = \bar{\rho}$, or that the density fluctuations are negligible; but this does not necessarily mean that the fluid is incompressible. The consequences of the assumptions are, for example,

$$\overline{v_i'} = 0, \quad \overline{\rho v_i'} = 0, \quad \overline{p_{ki}'} = 0, \quad \bar{\bar{v}}_i = \bar{v}_i,$$

$$\overline{\rho v_i} = \rho \bar{v}_i, \quad \overline{\frac{\partial}{\partial x_k}(\rho \bar{v}_i v_k')} = 0, \quad \text{etc.}$$

2. Note that $\overline{s'}$ is not the average value of deviations from a fixed mean, because $\bar{s}$ also varies from point to point; thus, $\overline{s'} = 0$ does not follow from the definitions but must be stated as an assumption.

Hesselberg [5] has defined the mean velocity by the formula:

$$\bar{u}_i = \frac{\int_V \int_t \rho v_i \, dt \, dV}{\int_V \int_t \rho \, dt \, dV},$$

as a consequence of which $\overline{\rho u_i'}$ is automatically zero whether ρ' is negligible or not. Energy equations written for this density-weighted mean velocity are similar in form to the ones that follow, but many of the terms have different meanings and values. The choice of the averaging formula for the velocity should be determined finally by the instrumental technique that is used in measuring v_i. If the instrument responds directly to the momentum ρv_i, then it measures the $\bar{u}_i$ of Hesselberg's formula. If it responds directly to the velocity v_i, then it measures $\bar{v}_i$ and formula (18) should be used. If ρ' is truly negligible or is not correlated with velocity components, the two formulas are identical.

Energy Equations for Averaged Properties

The process of analysis for averaged properties is applied to the first law of thermodynamics as follows. By combination with the equation of continuity, equation (9) is transformed to apply to a unit volume:

$$\frac{\partial}{\partial t}(\rho I + \rho L) + \frac{\partial}{\partial x_k}(v_k \rho I + v_k \rho L) = \rho Q^I + \rho Q^L + \rho M. \tag{19}$$

Then v_k is replaced by $\bar{v}_k + v_k'$, and the terms are averaged by formula (18) over the space and time pertaining to $\bar{v}_k$:

$$\frac{\partial}{\partial t}(\rho \bar{I} + \rho \bar{L}) + \frac{\partial}{\partial x_k}(\bar{v}_k \rho \bar{I} + \bar{v}_k \rho \bar{L} + \overline{\rho v_k' I} + \overline{\rho v_k' L}) = \rho \bar{Q}^I + \rho \bar{Q}^L + \rho \bar{M}, \tag{20}$$

Now,

$$\bar{M} = \alpha \bar{p}_{ki} \frac{\partial \bar{v}_i}{\partial x_k} + \overline{\alpha p_{ki}' \frac{\partial v_i'}{\partial x_k}} + \overline{\alpha p_{ki}' \frac{\partial \bar{v}_i}{\partial x_k}} + \overline{\alpha \bar{p}_{ki} \frac{\partial v_i'}{\partial x_k}}, \tag{21}$$

in which the last two terms are assumed to be negligible. The first two terms on the right will be indicated by M_m and M_e, respectively. The energy equation is integrated over the total volume of the arbitrary system and becomes finally

$$\frac{D}{Dt}(\bar{I}^* + \bar{L}^*) = \left[\int_\sigma [\bar{u}\rho(\bar{I} + \bar{L}) - \overline{\rho v_n'(I + L)} - \bar{q}_n^I - \bar{q}_n^L] \, d\sigma\right]_A - [-M_m^* - M_e^*]_W, \tag{22}$$

where $\bar{u}$ is the normal velocity of the boundary surface relative to $\bar{v}_n$, or $V_n - \bar{v}_n$.

The term $\bar{u}\rho(\bar{I} + \bar{L})$ is the flux of the averaged thermal and latent-heat energies at the velocity of the boundary relative to the averaged air velocity. The flux $\overline{\rho v_n'(I + L)}$, averaged locally before integration over the surface, is associated with the "eddy" velocity of the air, v_n', normal to the surface; it can be called the eddy flux of thermal and latent-heat energies. The terms $\bar{q}_n^I$ and $\bar{q}_n^L$ are averaged values of the fluxes that appear in the thermal-energy equation (14); $\bar{q}_n^I$, for example, consists in part of molecular conduction proportional to the gradient of the mean temperature:

$$\bar{q}_n^I = -c \frac{\partial \bar{T}}{\partial x_n} + \text{flux of radiant energy},$$

where x_n is a coordinate normal to the surface and directed outward. The last terms, M_m^* and M_e^*, are the volume integral of the molecular transformation function: M_m^* for averaged stresses and velocity, M_e^* for deviations of stresses and velocity from their averaged values.

There are two alternate ways of deriving an energy equation for averaged properties from the equations of motion: The equations may be averaged first and then converted into an equation for the kinetic energy of averaged motion alone; or they may be converted into an energy equation and then averaged, yielding an equation for kinetic energies of both averaged and eddy motions. The difference between these two equations is the equation for the kinetic energy of eddy motions alone.

In the first method, the equations (6) are converted to the momentum form by combining each of them with the equation of continuity. The velocities and stresses are replaced with their averaged values and deviations, and the terms of the equations are averaged by formula (18). These three equations are multiplied by $\bar{v}_1$, $\bar{v}_2$, and $\bar{v}_3$, respectively; the averaged form of the equation of continuity is multiplied by $-\bar{v}_i^2/2$; and all four equations are added together. The final result is

$$\frac{\partial}{\partial t}\left(\rho \frac{\bar{v}_i^2}{2} + \overline{\rho g x_3}\right) + \frac{\partial}{\partial x_k}\left(\bar{v}_k \rho \frac{\bar{v}_i^2}{2} + \overline{\bar{v}_k \rho g x_3}\right) = \bar{v}_i \frac{\partial T_{ki}}{\partial x_k} + \bar{v}_i \frac{\partial \bar{p}_{ki}}{\partial x_k}. \tag{23}$$

The symbol T_{ki} stands for the "eddy stress" tensor:

$$T_{ki} = T_{ik} = -\overline{\rho v_i' v_k'}. \tag{24}$$

This term, though it arises from the averaging of the accelerational term, is customarily regarded as a stress against the mean motion, acting in the x_i-direction in a plane normal to the x_k-axis, on the fluid lying to the negative side of the plane.

Equation (23) is transformed into an equation valid for a general system by the method already described, and it becomes

$$\frac{D}{Dt}(K_m^* + \bar{P}^*) = \left[\int_\sigma \bar{u}\rho(K_m + \bar{P}) \, d\sigma\right]_A - \left[M_m^* + E^* - \int_\sigma \bar{v}_i(\bar{p}_{ni} + T_{ni}) \, d\sigma\right]_W, \tag{25}$$

where $K_m = \bar{v}_i^2/2$, the kinetic energy, per unit mass, of the averaged motion; $\bar{P} = \overline{gx_3}$, which, incidentally, is equivalent to the unaveraged gx_3 or P; and E is the "eddy-transformation function":

$$E = \alpha T_{ki} \frac{\partial \bar{v}_i}{\partial x_k}. \tag{26}$$

The other terms and their superimposed symbols have the same meanings as in the preceding analyses.

In the second method of deriving a kinetic-energy equation for averaged properties, the starting point is the basic equation of kinetic energy (10), transformed for a unit volume. Again, v_i and p_{ki} are replaced with $\bar{v}_i + v_i'$ and $\bar{p}_{ki} + p_{ki}'$, and the terms are averaged. The resulting equation contains the kinetic energy of eddy motions as well as averaged motion. When equation (23) is subtracted from this equation, there is obtained, after transformation to the general system,

$$\frac{D\bar{K}_e^*}{Dt} = \left[\int_\sigma (\bar{u}\rho\bar{K}_e - \rho\overline{v_n' K_e})\, d\sigma\right]_A - \left[M_e^* - E^* + \int_\sigma \overline{v_i' p_{ni}'}\, d\sigma\right]_W, \tag{27}$$

where $K_e = v_i'^2/2$ and $\bar{K}_e = \overline{v_i'^2/2}$. The function E^* appears with opposite signs in (25) and (27). This explains why E has been called the eddy-transformation function; when positive, it decreases the energy of the mean motion and increases the energy of the eddy motion. It is analogous to the molecular transformation function M.

Applications of the General Energy Equations

Equations (22), (25), and (27) constitute an array of general energy equations for atmospheric processes. Many different special equations can be obtained from them. For example, if turbulent fluctuations are nonexistent ($v_i' = v_k' = p' = p_{ki}' = 0$), if the system's boundary everywhere moves with the air, and if the system is homogeneous, equation (27) vanishes, and (22) and (25) can be reduced to the basic equations (9) and (10).

If one wishes to establish a criterion for the increase of eddying motion, one starts with equation (27) and modifies it according to whatever conditions may be assumed. Richardson's criterion [12], however, cannot be derived from that equation because of the omission of terms involving density fluctuations. One such term omitted from (27) is $-\int_V g\overline{\rho' v_3'}\, dV$. If this term[3] is included and if, following Richardson, one assumes that the boundary work and flux terms and the term M_e^* are negligible, then (27) becomes

$$\frac{D\bar{K}_e^*}{Dt} = E^* - \int_V g\overline{\rho' v_3'}\, dV. \tag{28}$$

Richardson's number is the ratio

$$Ri = \frac{\int_V g\overline{\rho' v_3'}\, dV}{E^*}.$$

According to the assumptions that are usually made in applications of Richardson's criterion, the ratio reduces to the stability divided by the square of the vertical shear; when this ratio is less than a critical value Ri_{crit}, turbulence is supposed to increase. Richardson assumed that $Ri_{\text{crit}} = 1$.

It will not be argued that the terms involving ρ' are negligibly small, for they very well may not be. They have been omitted from the equations developed here solely to permit a discussion of the general aspects of the equations without too much involvement with details. But the fact that this omission eliminates the possibility of deriving Richardson's criterion from the equations may not be a serious fault. The criterion depends upon the term $g\overline{\rho' v_3'}$ being positive and different from zero; in other words, upon a positive correlation between density and vertical velocity. There is no experimental evidence that such correlation is to be expected. The occasional examples cited in support of the criterion are more qualitative than quantitative; according to Sutton [16], "some support can be found for almost any value of Ri_{crit} between 0.04 and 1."

Richardson's criterion appears to be qualitatively correct, because observations show that turbulence tends to be suppressed in a stable layer and to be increased when the vertical wind shear is great. The qualitative success may be explained without considering the term $g\overline{\rho' v_3'}$. Suppose (27) is applied to a homogeneous system, and all terms are dropped except the following:

$$\frac{D\bar{K}_e^*}{Dt} = -M_e^* + E^*. \tag{29}$$

The first term on the right represents the dissipation of eddy energy into thermal energy; it is probably negative most of the time so that it consistently acts to decrease the eddy energy by converting it into thermal energy. The second term E^*, for the frictional layer, is essentially the volume integral of

$$\rho E = -\rho\overline{v_1' v_3'}\frac{\partial \bar{v}_1}{\partial x_3} - \rho\overline{v_2' v_3'}\frac{\partial \bar{v}_2}{\partial x_3}, \tag{30}$$

where the x_3-axis is vertical. According to empirical evidence, the eddy-transformation function E is predominantly positive, so it consistently acts to increase the eddy energy at the expense of the energy of mean motion. The magnitude of E increases with increasing vertical wind shear, $\partial\bar{v}_1/\partial x_3$ and $\partial\bar{v}_2/\partial x_3$; thus, the eddy energy tends to increase with increasing shear. Its magnitude increases also with increasing v_3'; but this fluctuation of the vertical velocity tends to be damped

3. Calder [2] derives the kinetic-energy equations for mean and eddy motion by a procednre similar to that followed here, bnt he retains this one term involving ρ'. His arguments for retaining it appear logical, but it would seem that the arguments apply equally to some of the terms that he does not retain.

out when the stability is great. Hence, with smaller stability v_3' is greater on the average, E is greater, and the eddy energy tends to increase with decreasing stability.

The array of general energy equations may be combined in four ways: any two or all three may be added together. When all three are combined, a total-energy equation is obtained:

$$\frac{D}{Dt}(\bar{I}^* + \bar{L}^* + \bar{K}_m^* + \bar{P}^* + \bar{K}_e^*) = \left[\int_\sigma [\bar{u}\rho(\bar{I} + \bar{L} + \bar{K}_m + \bar{P} + \bar{K}_e) - \rho\overline{v_n'(I + L + K_e)}]\, d\sigma\right]_A - \left[-\int_\sigma [\bar{v}_i(\bar{p}_{ni} + T_{ni}) + \overline{v_i' p_{ni}'}]\, d\sigma\right]_W. \tag{31}$$

If no energy flux or work occurs at the boundary surface, the total energy within the system remains unchanged. This is the law of conservation of energy.

Margules' models of energy transformations in the atmosphere were based upon an equation which can be obtained from (31) by neglecting turbulent fluctuations and boundary activity:

$$\frac{D}{Dt}(I^* + L^* + K^* + P^*) = 0. \tag{32}$$

Margules computed the changes of latent-heat, potential, and thermal energies for closed, stationary systems which progressed, in certain special ways, from a state of instability to a state of stability. He was then able to compute the associated change of kinetic energy by means of (32). His masterful analysis of the models has deeply influenced meteorological thought on atmospheric-energy transformations. Margules, however, was apparently well aware that his models were far from realistic and that, at best, they provided nothing more than general suggestions of the source of atmospheric motions. Furthermore, Spar [13] has investigated the energy changes in two deepening cyclones and has stated that the results "lend no support to Margules' energy theory of cyclones."

To reduce the total-energy equation to Bjerknes' generalization of Bernoulli's equation [1], one must omit a number of terms and undo most of the analysis that produced the total-energy equation. Turbulent fluctuations, viscous stresses, and latent-heat energy are omitted; and the equation is applied to a system consisting of a small volume of air moving at the air velocity, instead of a large volume moving arbitrarily. With these conditions, equation (30) reduces to

$$\frac{\partial}{\partial t}[\rho(I + K + P + p\alpha)] - \frac{\partial p}{\partial t} = -\frac{\partial}{\partial x_k}[\rho v_k(I + K + P + p\alpha)]. \tag{33}$$

After application of the equation of continuity this becomes, for a steady pressure distribution,

$$c_v T + \frac{v_i^2}{2} + g x_3 + p\alpha = \text{const}, \tag{34}$$

along a trajectory.

The eddy flux of thermal and latent-heat energies is the first of the following three eddy terms from the right side of (31):

$$-\rho\overline{v_n'(I + L)} - \rho\overline{v_n' K_e} + \overline{v_i' p_{ni}'}.$$

It differs from the expression for the vertical eddy flux of heat derived by Montgomery [10], partly because he defined the mean velocity by Hesselberg's formula, and partly because he combined the pressure part of the work term $\overline{v_i' p_{ni}'}$ with the flux term. The first difference is unimportant if $\rho = \bar{\rho}$. As for the second difference, the term $\overline{v_i' p_{ni}'}$ becomes either $-\overline{v_n' p'}$ or $-\overline{v_n'}\,\overline{p}$ (since $\overline{v_n'}\,\overline{p} = 0$) if the viscous stresses are ignored; and this, combined with the flux, gives Montgomery's expression for the x_3-direction:

$$-\rho\overline{v_3'(I + L + p\alpha)} = -\rho\overline{v_3' h_t},$$

where $h_t = I + L + p\alpha$. For dry air, $h_t = h = c_v T + p\alpha$, the enthalpy.

The preceding examples are only a small sample of the numerous studies of atmospheric energy. Many others can be found without difficulty in the literature. Since no general approach to the philosophy of atmospheric energy has been adopted by meteorologists, the problems discussed in the literature may sometimes seem to have very little to do with the energy laws discussed here. It is unfortunate that the experimentally determined facts are not always kept distinct from the definitions, assumptions, and speculations. If this distinction were preserved, the work of different investigators could be more readily fitted into a single philosophy.

The general energy equations permit the use of averaged values of velocity and stresses, but they require also a knowledge of eddy terms that are not measured directly. These terms have been the subject of many investigations during the past forty years. The most common approach to their evaluation seems to be guided by their analogy to terms associated with molecular motions. The molecular stress p_{ki} is a function of the pressure p, viscosity μ, and velocity gradients. By analogy, one might wish to express the eddy stress T_{ki} by a similar formula in terms of an "eddy pressure" [12], "eddy viscosity," and gradients of the averaged velocity. The molecular flux of heat is a function of a coefficient of heat conduction and the temperature gradient; and by analogy, the eddy flux of thermal energy may be expressed in terms of a coefficient of eddy conduction and the gradient of mean temperature:

$$\overline{v_k' I} = -C\frac{\partial \bar{I}}{\partial x_k}.$$

The trouble with the analogy is that the coefficients of eddy viscosity and thermal conductivity are too variable and unpredictable, and are not positive at all times;

while the corresponding coefficients on the molecular scale are established physical properties, varying in a predictable manner when other properties vary.

It is apparent, then, that the general energy equations cannot be applied successfully to the real atmosphere until more is known about the eddy terms. The problem of these terms can be formulated more effectively if their mutual relationship, as shown by general equations, is taken into consideration. For example, whether eddy flux of thermal energy, $-\overline{\rho v_k' c_v T}$, or eddy flux of enthalpy, $-\overline{\rho v_k'(c_v T + p\alpha)}$, should be called the eddy flux of heat does not appear to be an important question. The latter, however, combines the flux of a property recognized as energy with an eddy term recognized as work, and there may be some point in maintaining the distinction between energy and work.

It is well known from observations that changes of potential and thermal energies, and sometimes latent-heat energies, are ten to a hundred times greater than changes of kinetic energy in the atmosphere. This difference in magnitudes evidently is caused by some property of the atmosphere, but the energy equations give no clue to the nature of the property. Neither do the equations give any clue as to which factors are most important in determining changes of kinetic energy. The equations, being in differential form, represent relationships between instantaneous rates, but they tell nothing about the relative magnitudes of the various rates of work, flux, and changes of energy. The solution to the problem lies in conditions that are not reflected in the equations; specifically, it lies in the initial and boundary conditions.

For example, increasing kinetic energy of the mean motion in the air near the surface of the earth may lead, because of perturbations set up by the roughness of the surface, to an increasing value of the eddy transformation function E. This effect, as can be seen from the energy equations, results in a more rapid transformation of the mean kinetic energy into eddy kinetic energy, so that the mean energy may approach a limiting value. But the increasing eddy energy implies greater turbulent fluctuations, hence stronger gradients of the unaveraged motion, and hence an increasing value of the molecular transformation function M_e. As this function increases, the eddy energy is more rapidly transformed into thermal energy. The net results are that the mean and eddy kinetic energies may approach limiting values where energy is supplied by some source to the mean motion and degraded finally to thermal energy.

The energy equations, however, quite clearly do not tell the whole story. Some additional guiding principles are required to permit a complete solution to problems of this kind. Classical thermodynamics has such guiding principles—for example, the second law. There should be more attempts to formulate the necessary principles in studies of atmospheric energy.

REFERENCES

1. Bjerknes, V., "Theoretisch-meteorologische Mitteilungen. 4. Die hydrodynamisch-thermodynamische Energiegleichung." *Meteor. Z.*, 34:166–176 (1917).
2. Calder, K. L., "The Criterion of Turbulence in a Fluid of Variable Density with Particular Reference to Conditions in the Atmosphere." *Quart. J. R. meteor. Soc.*, 75:71–88 (1949).
3. Dryden, H. L., "A Review of the Statistical Theory of Turbulence." *Quart. appl. Math.*, 1:7–42 (1943).
4. Ertel, H., "Die hydrothermodynamischen Grundgleichungen turbulenter Luftströmungen." *Meteor. Z.*, 60: 289–295 (1943).
5. Hesselberg, T., "Die Gesetze der ausgeglichenen atmosphärischen Bewegungen." *Beitr. Phys. frei. Atmos.*, 12:141–160 (1926).
6. Margules, M., "Über den Arbeitswert einer Luftdruckvertheilung und über die Erhaltung der Druckunterschiede." *Denkschr. Akad. Wiss. Wien*, 73:329–345 (1901). (English translation by Cleveland Abbé in "The Mechanics of the Earth's Atmosphere," Third Collection. *Smithson. misc. Coll.*, 51:501–532 (1910).)
7. —— "Über die Energie der Stürme." *Jb. ZentAnst. Meteor. Wien* (1903). (Translation by Cleveland Abbé in "The Mechanics of the Earth's Atmosphere," Third Collection. *Smithson. misc. Coll.*, 51:533–595 (1910).)
8. —— "Zur Sturmtheorie." *Meteor. Z.*, 23:481–497 (1906).
9. Miller, J. E., "Energy Transformation Functions." *J. Meteor.*, 7:152–159 (1950).
10. Montgomery, R. B., "Vertical Eddy Flux of Heat in the Atmosphere." *J. Meteor.*, 5:265–274 (1948).
11. Reynolds, O., "On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion." *Phil. Trans. roy. Soc. London*, (A) 186:123–164 (1895).
12. Richardson, L. F., "The Supply of Energy from and to Atmospheric Eddies." *Proc. roy. Soc.*, (A) 97:354–373 (1920).
13. Spar, J., "Synoptic Studies of the Potential Energy in Cyclones." *J. Meteor.*, 7:48–53 (1950).
14. Stewart, H. J., "The Energy Equation for a Viscous Compressible Fluid." *Proc. nat. Acad. Sci., Wash.*, 28:161–164 (1942).
15. —— "Kinematics and Dynamics of Fluid Flow" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See p. 429)
16. Sutton, O. G., *Atmospheric Turbulence*. London, Methuen, 1949. (See p. 92)

ATMOSPHERIC TURBULENCE AND DIFFUSION

By O. G. SUTTON

Military College of Science, Shrivenham, England

THE AERODYNAMICAL BACKGROUND

The Nature of Turbulent Flow. A particle in a stream of fluid can never follow a perfectly smooth path because of minute random disturbances arising from the molecular structure of the fluid (Brownian motion), but observation shows that, in certain circumstances, oscillations appear in the path which are much too large to be ascribed to molecular agitation. Such irregularities must imply the existence of rapid and apparently random fluctuations in the velocity of the stream, constituting a permanent and characteristic feature of this type of flow. Turbulence can hardly be defined in a strict mathematical sense, but is generally understood to imply a motion characterized by a continuous succession of such finite disturbances and in this sense nearly all natural motion, whether of water or of air, is turbulent. Only exceptionally is there found in nature a truly nonturbulent or laminar flow, in which the only random disturbances are the infinitesimal fluctuations due to molecular agitation.

Anemometer records show that in general, and especially in the lower layers of the atmosphere, the wind is highly turbulent, the velocity being a complex of oscillations of duration varying from a fraction of a second to many minutes and of an amplitude which is often a substantial fraction of the average speed. Similar irregular oscillations are shown by direction indicators, so that the speed of the wind changes not only from instant to instant but also from point to point of space. A complete specification of the velocity field over even a limited portion of the atmosphere is in practice unattainable and, to make progress, attention must be concentrated upon mean values and other statistical functions of the velocity. The study of atmospheric turbulence is chiefly concerned with the analysis of the mean distribution of momentum, heat and suspended matter in, and as a result of, this highly complex and rapidly changing field.

Air flow near the surface of the earth, the region in which atmospheric turbulence is of greatest importance, resembles in many respects turbulent motion in long straight pipes or near solid boundaries, as in wind tunnels, and aerodynamics thus affords a natural and convenient starting point for the meteorological problem. The earliest recognition of two distinct types of flow—laminar and turbulent—seems to have been made by Hagen about 1839, but the detailed and systematic study of turbulence undoubtedly opens in 1883 with the famous experiments of Osborne Reynolds [49] on the flow of water through long straight glass tubes. Reynolds, by the simple device of making the flow visible by a thin stream of dye, was able to demonstrate that the transition from an orderly rectilinear motion (laminar flow) in which the thread of dye remains intact from inlet to outlet, to a disorderly or turbulent flow, evinced by the rapid disintegration of the filament of dye, takes place when the Reynolds number $\bar{u}d/\nu$ ($\bar{u}$ = mean speed, d = tube diameter, ν = kinematic viscosity) exceeds a certain value. At the same time other properties of the flow are changed, in particular the distribution of mean velocity across the pipe. In the laminar state the velocity profile is parabolic from the wall to the centre, but in turbulent flow the velocity is almost uniform over a central core, declining sharply to zero in a very thin layer adjacent to the wall itself. Such a change is easily accounted for in general terms. In laminar flow the only intermingling of adjacent layers of fluid is that due to molecular agitation (viscosity, conduction, diffusion), but in turbulent motion the fluid elements, because they follow extremely tortuous paths, transfer momentum, heat and matter freely from one layer to another and thus tend to smooth out local differences in velocity, temperature and concentration of suspended matter. In other words, the principal effect of the turbulence is to cause enhanced *mixing*, and it is this aspect which is of paramount importance in meteorology. Very similar features are found in flow near solid surfaces. When a uniform stream of air meets a solid body, the influence of viscosity is felt only in the immediate vicinity of the surface, in the so-called *boundary layer*, a very shallow region characterized by large velocity gradients. Flow in such layers may be laminar, partly laminar and partly turbulent, or wholly turbulent. In a laminar boundary layer the velocity increases fairly regularly from the wall to the free stream, but in a turbulent boundary layer the velocity profile is much more uniform over the greater part of the layer, decreasing sharply to zero on approaching the wall itself. The effect of the turbulence is to bring down elements of fast-moving air from the free stream to the wall and, conversely, to remove the retarded air at the wall into the free stream so that, except in the immediate vicinity of the surface, the velocity gradient is much reduced compared with that found in laminar flow.

These observations have their counterparts in natural flow near the ground, but here the problem is much complicated by two extremely important differences. Laboratory investigations at low speeds deal almost exclusively with fluids in which marked density differences do not exist and for which the effects of gravity on the flow are negligible. The maintenance of the turbulent state calls for a continuous supply of energy to be used in moving elements of the fluid from one level to another. If the lower layers of the fluid considerably exceed in density those above, the work which has to be done against the gravitational field in lifting masses of

fluid from lower to higher levels must reduce the intensity of the turbulence and may ultimately cause a transition to laminar flow. On the other hand, a fluid in which density increases with height is favourable to the formation of upward currents and therefore to the maintenance and the growth of turbulence. The atmosphere is a fluid whose lower boundary is subjected to strong heating and cooling, especially in clear weather, so that near the ground the turbulence of the wind tends to rise to a maximum in the mid-hours of a sunny day, when there is usually a pronounced superadiabatic lapse rate in the lowest layers, and to diminish or even die away completely during a clear night, when the radiative cooling of the ground creates a marked inversion of temperature gradient in the lowest layers. The gradient of temperature in the vertical is thus to be regarded as a factor exerting a powerful control on the turbulence of the wind. A second difficulty peculiar to meteorology arises from the variable nature of the surface over which the air flows. Provided that the obstacles which cover a surface are not too large and are evenly distributed, aerodynamic investigations have indicated a rational method of allowing for their effects, and the same concepts have been applied with considerable success in many problems of atmospheric motion, but in meteorology cases frequently arise in which the surface irregularities are too large or too unevenly distributed to be treated in this way. This is especially the case, for example, in considering the diffusion of atmospheric pollution in built-up areas.

Mathematical Treatment. It is generally accepted that in any fluid motion the component velocities u, v and w, along axes of x, y and z respectively, satisfy the Navier-Stokes equations, that is, three relations of the type

$$\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}\right) = -\frac{\partial p}{\partial x}+\mu\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+\rho X,$$

where ρ is the density, p is the pressure, μ is the dynamic viscosity and X the x-component of any external force. To these must be added the equation of continuity, which expresses the conservation of mass. Exact solutions of this set of nonlinear equations are known only for certain special cases which have little or no interest for meteorology.

The mean velocities which appear so prominently in studies of turbulence are usually averages over an interval of time (T), that is, they are defined by the relations

$$\bar{u}=\frac{1}{T}\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T} u\,dt, \qquad \bar{v}=\frac{1}{T}\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T} v\,dt,$$

$$\bar{w}=\frac{1}{T}\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T} w\,dt,$$

in which case the fluctuations or eddy velocities u', v' and w' are defined by

$$u=\bar{u}+u', \quad v=\bar{v}+v', \quad w=\bar{w}+w'.$$

If the interval of time T is sufficiently long, it may be asserted that

$$\bar{u}'=\bar{v}'=\bar{w}'=0.$$

The effect of viscosity is to set up in the fluid a system of stresses, namely three normal stresses p_{xx}, p_{yy} and p_{zz} defined by

$$p_{xx}=-p+2\mu\frac{\partial u}{\partial x}, \qquad p_{yy}=-p+2\mu\frac{\partial v}{\partial y},$$

$$p_{zz}=-p+2\mu\frac{\partial w}{\partial z},$$

and three tangential stresses, p_{xy}, p_{yz}, p_{zx} defined by

$$p_{xy}=\mu\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right), \qquad p_{yz}=\mu\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right),$$

$$p_{zx}=\mu\left(\frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\right).$$

Introducing these stresses into the equations of motion, we have three equations of the type

$$\rho\frac{\partial u}{\partial t}=\frac{\partial}{\partial x}(p_{xx}-\rho u^2)+\frac{\partial}{\partial y}(p_{xy}-\rho uv)+\frac{\partial}{\partial z}(p_{zx}-\rho uw)+\rho X. \tag{1}$$

Putting $u=\bar{u}+u'$, etc., and taking means, we have

$$\rho\frac{\partial \bar{u}}{\partial t}=\frac{\partial}{\partial x}(\bar{p}_{xx}-\rho\bar{u}^2-\rho\overline{u'^2})+\frac{\partial}{\partial y}(\bar{p}_{xy}-\rho\bar{u}\bar{v}-\rho\overline{u'v'})+\frac{\partial}{\partial z}(\bar{p}_{zx}-\rho\bar{u}\bar{w}-\rho\overline{u'w'})+\rho X, \tag{2}$$

and analogous equations for the y- and z-components. The only formal change which has occurred is that in equations (2) certain additional terms, depending upon the eddy velocities, have been added to the original viscous stresses. These additional terms, called the *Reynolds stresses*, are the mathematical expression of the effect of the velocity fluctuations in transporting momentum across a surface in the fluid, just as the original stresses represent the effect of the molecular agitation in transporting momentum by viscosity. In general, the Reynolds stresses are considerably greater than the corresponding viscous stresses, and it is usually possible to ignore the latter in problems of atmospheric turbulence.

No general method has yet been evolved for expressing the Reynolds stresses in terms of the velocity components and their spatial derivatives by exact analysis, and this constitutes the principal difficulty in proceeding further along these lines. The study of a special case, however, suggests semi-empirical methods by which progress can be made despite the formidable difficulties of the complete problem.

Flow Near a Boundary. Consider a steady flow in

which the mean motion has the same direction at all points and in which the turbulence, specified by the mean squares of the oscillations, is uniform in all directions. Such conditions are approached fairly closely in motion very near a plane solid boundary, such as the surface of the earth. In the notation of the previous section, if x be in the direction of mean flow and z distance normal to the surface ($z = 0$),

$$\bar{u} = \bar{u}(z), \quad \bar{v} = \bar{w} = 0.$$

If we introduce the eddy shearing-stress $\tau = -\rho\overline{u'w'}$, the equations of motion reduce to the single equation

$$\frac{\partial \tau}{\partial z} = \frac{\partial \bar{p}}{\partial x},$$

in which molecular terms have been neglected. If the pressure gradient $\partial\bar{p}/\partial x$ is invariable throughout the shallow layer concerned,

$$\tau = \tau_0 + z\,\frac{\partial \bar{p}}{\partial x},$$

where τ_0 is the value of τ as $z \to 0$.

In many aerodynamical problems the pressure gradient is effectively zero, and in micrometeorological applications $\partial\bar{p}/\partial x$ is usually small compared with τ_0/z. For moderate values of z the shearing stress τ may thus be considered invariable with height and equal to the value at the surface, τ_0. This assumption is almost always made in problems of turbulence near the ground. For a more detailed examination of the meteorological problem the reader should consult papers by Ertel [17] and Calder [6], who conclude that in the atmosphere the constancy of τ with height may usually be safely assumed for values of z not exceeding about 25 m. In this special case the transport of momentum across any horizontal plane is effectively measured by the eddy shearing-stress $-\rho\overline{\bar{u}'\bar{w}'}$, that is, by a quantity involving the correlation between the eddy velocities u' and w'. The existence of this correlation expresses the fact that gusts or positive values of u' are more frequently associated with downward-moving air, and lulls (negative u') with upward-moving air, than vice versa.

The form of the profile of mean flow, however, cannot be deduced unless some further hypothesis is introduced. The earliest attempt to frame such a hypothesis appears to have been that of Boussinesq [3] in 1877. There is an obvious analogy between the action of the velocity fluctuations in a turbulent fluid and the motion of the molecules in a gas, which suggests that the effects produced by the turbulence may be ascribed to the movements of discrete masses of fluid, called *eddies*, from one level to another. In the corresponding problem in purely laminar flow the shearing stress is $\mu\,\partial u/\partial z$, which suggests that, on this analogy, the Reynolds stress may be expressed as the product of the gradient of mean velocity and a virtual viscosity, that is,

$$\tau = -\rho\overline{u'w'} = A\,\partial\bar{u}/\partial z = K\rho\,\partial\bar{u}/\partial z. \quad (3)$$

The quantity A, called an interchange coefficient (*Austauschkoeffizient*), corresponds to the dynamic viscosity μ, and the quantity K, called the eddy viscosity, corresponds to the kinematic viscosity ν. It is obvious that the same concept can be applied equally well to the transport by turbulence of heat or suspended matter, by defining in a similar fashion the eddy conductivity and the eddy diffusivity. This simple but powerful concept has had a much greater influence on dynamical meteorology than on other branches of fluid motion. As an initial assumption it is natural to suppose that A and K behave exactly like their molecular counterparts, that is, are true constants and thus independent of position in the field. If this were strictly true, turbulent motion would be simply an enlarged copy of laminar flow, which is far from being the case. This hypothesis is now abandoned, but it should be recognized that in meteorology it has played an invaluable part in bringing to prominence the enormous difference in magnitude between molecular and turbulent transport. The magnitudes usually quoted in meteorological literature for K are 10^3, 10^4 or 10^5 cm^2 sec^{-1}, which should be contrasted with the kinematic viscosity of air, which is of the order of 10^{-1} cm^2 sec^{-1}.

The Mixing-Length Hypothesis. It was soon recognized, especially by workers in aerodynamics, that the eddy viscosity itself was likely to prove too complicated as a starting point for the analysis of turbulent flow. An attempt to improve the treatment without entirely abandoning the analogy with molecular theory was made by Prandtl [44] in 1925. This has proved extremely fruitful although, like many other theories of turbulence, it is semi-empirical, and intuitive rather than analytical.

Measurements of turbulent flow indicate that the virtual stresses set up by the turbulence are approximately proportional to the square of the mean velocity. It is convenient to define an auxiliary reference velocity u_*, known as the friction velocity (*Schubspannungsgeschwindigkeit*) for which this relation is exact, that is,

$$u_*^2 = \tau/\rho = \overline{u'w'}.$$

Thus u_* is a quantity of the same order of magnitude as the eddy velocities (that is, in most meteorological applications u_* is about $\frac{1}{10}$ of the mean speed, the exact value depending on the nature of the terrain). A list of typical values of u_* has been given by Sutton [63]. Prandtl then introduces a characteristic quantity l, called the *mixing length* (*Mischungsweg*), defined by

$$u_* = l\,\partial\bar{u}/\partial z,$$

so that

$$K = u_* l = l^2\,\partial\bar{u}/\partial z.$$

Thus l is a length which resembles, but is very much greater than, the free path of the kinetic theory of gases.[1] It may be interpreted in a general way as the distance which an eddy moves from its point of departure from the mean motion until it mixes again with the

1. Compare $\nu = \frac{1}{3}c\lambda$, where c = molecular velocity, λ = free path.

main body of the fluid. A mass of fluid leaving a layer z in which it has acquired the mean motion $\bar{u}(z)$, and moving a vertical distance l while conserving its momentum, will give rise to the fluctuation

$$\bar{u}(z + l) - \bar{u}(z) \approx l\, \partial\bar{u}/\partial z$$

at the new level. Thus $l\, \partial\bar{u}/\partial z$ may be regarded as representing either an eddy velocity or the friction velocity.

A somewhat different method [22] of introducing a characteristic length is as follows: Suppose that $E(z)$ is any transferable property, such as momentum, temperature or concentration of suspended matter, whose mean value is constant over any (x, y) plane and which is supposed to be conserved during a transfer from the plane $z = z_1$ to $z = z_2$. The mean rate at which $E(z)$ is transported across a unit area of an (x, y) plane is

$$q = -\overline{w'[E(z_2) - E(z_1)]} \approx -\overline{w'(z_2 - z_1)}\, \partial\bar{E}/\partial z.$$

If a length l be defined such that

$$l\sqrt{\overline{w'^2}} = \overline{w'(z_2 - z_1)},$$

the rate of transfer is

$$q = -l\sqrt{\overline{w'^2}}\, \partial\bar{E}/\partial z.$$

Applying this to the transfer of momentum, in which $E = \rho u$ and $q = -\tau$, we have

$$\tau = \rho l\sqrt{\overline{w'^2}}\, \partial\bar{u}/\partial z,$$

and so

$$K = l\sqrt{\overline{w'^2}}.$$

The advantages to be gained by such dissections are not immediately obvious, since the analysis gives no clue to possible variations of l with any of the variables. In practice, the introduction of the mixing length has proved very useful, since it has been found that only very simple assumptions regarding l are required to obtain a satisfactory mathematical representation of many complex phenomena. It is doubtful if it is possible to assign a definite physical meaning to the mixing length, but broadly it is clear that l is a measure of the average "size" of the eddies responsible for the mixing, or equally, a rough indication of the average depth of the layers over which mixing takes place.

Velocity Profile Near a Boundary. When a turbulent stream flows over a smooth surface, three regions of motion can be distinguished:

1. A shallow zone of laminar flow immediately adjacent to the wall in which the shearing stress is chiefly due to viscosity ("laminar sub-layer").
2. The turbulent boundary layer proper, lying above the laminar sub-layer, in which the Reynolds stress is at least as important as the viscous stress.
3. The main body of the turbulent fluid, lying above the boundary layers, in which viscosity plays a negligible part.

A surface whose irregularities are large enough to prevent the formation of a laminar sub-layer is said to be "aerodynamically rough."

Smooth Surface. We consider again motion in a shallow layer in which τ (and hence u_*) is invariable with height and in which the mean velocity $(\bar{u})$ is a function of z only. If the velocity profile be assumed to depend only upon l, ν, u_* and the independent variable z, it follows that the dimensionless ratio $\bar{u}/u_*$ must be expressible as a function of l/z and u_*z/ν, since these are the only nondimensional ratios which can be formed from these variables. If, in addition, the scale of the mixing is assumed to be proportional to distance from the boundary, that is, $l = kz$, where k is a pure number (Kármán's constant), the definitions of u_* and of τ lead to the first-order differential equation,

$$\frac{d\bar{u}}{dz} = \frac{u_*}{l} = \frac{u_*}{kz},$$

whence

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln\left(\frac{u_*z}{\nu}\right) + \text{const}, \tag{4}$$

where the "constant" is to be determined by a suitable boundary condition. The usual requirement that $\bar{u} = 0$ on $z = 0$ cannot be satisfied, but Nikuradse has shown that equation (4) agrees closely with observations made on flow in smooth pipes in the form,

$$\frac{\bar{u}}{u_*} = \frac{1}{0.4} \ln\left(\frac{u_*z}{\nu}\right) + 5.5 \approx 2.5 \ln\left(\frac{9u_*z}{\nu}\right), \tag{5}$$

that is, $k = 0.4$ and the velocity vanishes on the plane $z = \nu/9u_*$, the equation having no meaning for smaller values of z.

Rough Surface. It is a well-established experimental fact that in the case of an aerodynamically rough surface, the influence of viscosity is negligible, in which case the solution of equation (3) may be written

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln\frac{z}{z_0}, \tag{6}$$

where z_0 is another constant of integration. This form of the equation implies that $\bar{u} = 0$ on $z = z_0$ (the equation having no meaning for smaller values of z), so that z_0, usually termed the *roughness length*, is a quantity which may be regarded as specifying, in some manner, the effect of the irregularity of the surface on the mean flow. This is borne out by the fact that, for pipes whose interior surface is uniformly roughened with fine grains of sand, it has been found that a definite relation exists between the roughness length and the size of the irregularities ($z_0 \approx \frac{1}{30}$ of the average diameter of the grains). Schlichting [56] has gone further and has shown that a surface may be regarded as smooth if $u_*z_0/\nu < 0.13$ and fully rough if $u_*z_0/\nu > 2.5$.

The analysis above appears to be adequate if the layer concerned is very shallow, but since it implies that the effect of the irregularities is equally felt at all distances above the surface it cannot be applied, without care, to the deeper layers met with in most meteorological applications. An alternative treatment of the rough-surface problem, due to Rossby and Mont-

gomery [55], appears to be more appropriate for such cases. Rossby and Montgomery assume that

$$l = k(z + z_0),$$

which expresses the intuitive conception that the effects of the irregularities are most strongly felt in the immediate vicinity of the surface and hardly at all at the greater heights. Equation (3), with this form for l, yields

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln \left(\frac{z + z_0}{z_0} \right), \tag{7}$$

using the boundary condition $\bar{u} = 0$ on $z = 0$. Thus theoretically, the profile of mean velocity over a rough surface is fully determined provided both u_* and z_0 are known (k being regarded as a universal constant). These may be combined to form the quantity termed by Sutton [63, 65] the *macroviscosity*, defined by $N = u_* z_0$. The introduction of this quantity enables a single profile, applicable to both smooth and rough surfaces, to be obtained, thus overcoming the difficulty felt in the free use of equations (6) and (7), that neither of these tends to the smooth-surface form (5) as $z_0 \to 0$. The generalized profile is

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln \left(\frac{u_* z}{N + \nu/9} \right),$$

corresponding to $l = kz$, or

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln \left(\frac{u_* z + N}{N + \nu/9} \right)$$

if $l = k(z + z_0)$. As z_0 (and hence N) tends to zero these equations tend to the smooth-surface form, but in most meteorological applications N is of the order of 10 to 10^3 cm^2 sec^{-1} and is thus much greater than ν. Schlichting's criteria may be written

$$\text{smooth flow: } N < 0.13\nu \approx 0.02 \text{ cm}^2 \text{ sec}^{-1},$$

$$\text{rough flow: } N > 2.5\nu \approx 0.4 \text{ cm}^2 \text{ sec}^{-1}.$$

Power-Law Profiles. A somewhat less satisfactory representation of the flow near a boundary is obtained by the use of power laws, but this type of profile is often much easier to handle mathematically than the logarithmic profile. The usual form of the profile near a smooth surface is

$$\bar{u} = \bar{u}_1 (z/z_1)^p, \tag{8}$$

where $\bar{u}_1$ is the mean wind speed at height z_1 and p is a positive number whose value for moderate Reynolds numbers is about $\frac{1}{7}$. If $\tau/\rho = K \, \partial\bar{u}/\partial z$ is invariable with height, it follows that

$$K = K_1 (z/z_1)^{1-p}, \tag{9}$$

where K_1 is the value of K at $z = z_1$. Equations (8) and (9) constitute the so-called "conjugate power-laws" of Schmidt.

The Vorticity-Transport Hypothesis. The preceding analysis is based essentially upon the hypothesis that momentum is conserved during the motion of an eddy, and for this reason the Prandtl treatment is often referred to as the "momentum-transport theory." From the aspect of exact hydrodynamical theory this hypothesis is questionable, since it involves the assumption that the pressure fluctuations do not affect the transfer of momentum. An alternative hypothesis, in many ways more attractive, is that *vorticity* is conserved and this forms the basis of Taylor's treatment of the problem [72]. Space does not allow an adequate discussion of the matter here, nor, in the present state of development of the theory of atmospheric turbulence, are the differences between the two theories of major importance for meteorology, except perhaps in one respect. According to the momentum transport theory, the rate at which momentum is communicated to unit volume of the fluid by turbulence is

$$\frac{\partial \tau}{\partial z} = \rho \frac{\partial}{\partial z} \left\{ K(z) \frac{\partial \bar{u}}{\partial z} \right\},$$

whereas on the vorticity transport theory,

$$\frac{\partial \tau}{\partial z} = \rho K(z) \frac{\partial^2 \bar{u}}{\partial z^2}.$$

These two forms are equivalent if, and only if, $K(z) =$ constant. For further details the reader is referred to Taylor's original papers, or to the accounts given by Brunt [4] and Goldstein [22]; the latter gives considerable detail as regards tests on the laboratory scale. An illuminating discussion, chiefly from the standpoint of the experimental worker, of the validity of the various hypotheses which have been advanced in the last two or three decades in order to develop further the Reynolds theory of stresses has been given by Dryden [14]. When the predictions of the various theories are compared with the results of accurate measurements made in turbulent boundary layers, the disquieting conclusion is reached that all such theories fail at some point or other. Considerable doubt is now thrown on the validity of the mixing-length idea and even on the fundamental hypothesis that the eddy velocities and the turbulent shear-stresses are directly related to the mean velocity and its derivatives at a point (equations (3) above). These considerations may perhaps indicate that one fruitful period of development in the theory of turbulence is drawing to its close, and that the time is now ripe for advances along quite different lines.

Statistical Theories of Turbulence. One such fundamentally different approach to the general problem is that of the so-called "statistical" theory of turbulence, initiated by Taylor [73] in 1920 and later expanded by him in a remarkable series of papers in 1935 [74]. Subsequent developments were made by Kármán and Howarth [32] and an account of the position reached by 1943 has been given by Dryden [15].

All theories of turbulence are necessarily statistical in some sense or other, but whereas the theories described in the first part of this article start with a measured distribution of mean velocity and mean pressure and relate these to the virtual stresses, the methods about to be described originate in the statistical properties of the fluctuations and seek to establish exact

relations between these properties and the mean motion. The two approaches have their counterparts in molecular theory, one being analogous to the kinetic theory of the elastic sphere models while the other has certain affinities to the statistical mechanics of an assembly of molecules.

In his initial (1920) contribution Taylor considered what statistical properties of a fluctuating field of flow are required to determine the diffusion of a group of particles suspended in the medium. If the turbulence is statistically uniform and steady ($\overline{u'^2}$ independent of locality and time), the spread of a cluster of particles over a plane is given by

$$\overline{X^2} = 2\overline{u'^2} \int_0^T \int_0^t R(\xi)\, d\xi\, dt,$$

where X is the distance travelled by a particle in time T and $R(\xi)$ is the correlation coefficient between the fluctuating velocities which affect a particle at times t and $t + \xi$. Taylor's theorem means, in effect, that a knowledge of the mean eddying energy and of the correlation $R(\xi)$ specifies completely the diffusion in such a field. The application of this result to meteorological problems is considered later. The analysis also indicates a length l_1, analogous to the mixing length but independent of any model, defined by

$$l_1 = \sqrt{\overline{u'^2}} \int_0^\infty R(\xi)\, d\xi,$$

which is finite if $R(\xi)$ tends to zero in a suitable manner as $\xi \to \infty$.

Another way of representing such a field utilises the correlation $R(y)$ between velocities at points separated by a variable distance y. The length L defined by

$$L = \int_0^\infty R(y)\, dy$$

is called the "scale of the turbulence" and may be regarded as a measure of the average size of the eddies which constitute the pattern of flow. Evidence presented later suggests that in meteorological problems L has no effective upper bound or, in other words, that in atmospheric diffusion, account must be taken of eddies ranging from the minute convectional whorl to full-size disturbances which affect the general circulation of the atmosphere.

In later development of this theory attention has been particularly concentrated upon the problem of the decay of disturbances (*e.g.*, downstream of a grid in a pipe), especially in isotropic turbulence, for which Taylor has introduced another important length λ called the "microscale of turbulence" and defined by

$$1/\lambda^2 = \lim_{y \to 0} \left\{ \frac{1 - R(y)}{y^2} \right\}.$$

The microscale of turbulence indicates the average size of the small eddies which are responsible for the greater part of the dissipation of energy, and is also related to the curvature of the $R(y)$ curve at the origin.

Later Developments; Kolmogoroff's Theory. The most striking advances in the statistical theory of turbulence since 1941 are due to Kolmogoroff [34], Onsager [37], Weizsäcker [79], and Heisenberg [25]. These theories have many points of resemblance, and the most promising appears to be that advanced by Kolmogoroff, whose work has been well summarized in English by Batchelor [1].

Kolmogoroff's basic conception is that, at very high Reynolds numbers, all turbulent motion has the same sort of small-scale structure, although the mean flow may differ widely from situation to situation, and that the motion caused by the small eddies is isotropic and statistically uniform. A turbulent flow may be thought of as giving rise to a spectrum of oscillations whose wave lengths vary in scale from those characteristic of the mean flow to a lower limit below which the motion is entirely laminar. Within this spectrum, energy is continually being transferred from one length scale to another by the generation, from any given band of oscillations, of a set of smaller oscillations, and so on, until it is no longer possible to form smaller eddies. Thus the process may be thought of as one in which energy from the mean motion is fed into the long-wave end of the spectrum by the largest eddies and passed down the spectrum in the direction of decreasing wave length, until ultimately it is absorbed into the random heat motion of the molecules by viscosity. Kolmogoroff then puts forward two similarity hypotheses: (1) that the statistical characteristics can depend only on the mean energy dissipation per unit mass of fluid (ϵ) and on viscosity; (2) that the statistical characteristics of the motion due to the larger eddies are independent of viscosity and depend only upon the energy dissipation ϵ. From these plausible hypotheses it is possible, by arguments chiefly of a dimensional character, to make certain definite predictions about the properties of the mean flow, and in particular, to show that as the Reynolds number increases without limit, the coefficient of correlation between fluctuations at points distance y apart tends to the form $1 - Ay^{2/3}$ (A = constant), provided that y is small compared with the scale of the turbulence. Wind-tunnel measurements so far have provided excellent confirmation of deductions made from Kolmogoroff's hypotheses and there is little doubt that the theory constitutes a powerful tool for the resolution of many complex problems, and one which should especially appeal to meteorologists because of the emphasis laid upon the passage of energy from disturbances of one size to another. An account of this and other theories, together with some promising new developments, has recently been given by Kármán [31], who considers that the principal aim of the present period is to find the laws which govern the shapes of either the correlation or the spectrum functions.

We now pass from these theoretical and laboratory studies to detailed consideration of the meteorological problem.

PHYSICAL FEATURES OF ATMOSPHERIC TURBULENCE

Experimental investigations into atmospheric turbulence may be conveniently divided into: (1) those dealing with localized effects, usually in shallow layers near the ground, in which the consequences of the earth's rotation may be disregarded; (2) those relating to larger-scale processes, usually in the so-called "friction layer" or "planetary boundary-layer" (Lettau), extending from the surface to the level at which the geostrophic velocity is attained (*c.* 500 m); and (3) those concerned with the atmosphere as a whole.

The Temperature Field. The most direct manifestation of atmospheric turbulence is the presence of fluctuations in the wind, but in view of the controlling influence of temperature gradient, it is convenient to begin by discussing the temperature field near the ground. Accurate continuous observations of temperature differences between various heights extending from 2.5 cm to 87.7 m have been tabulated and analysed for southern England by Johnson [28], Best [2], Johnson and Heywood [30], and by Flower [19] for Ismailia, Egypt, while Ramdas [48] has given mean values of air temperature at various levels from 2.5 cm to 10.6 m at maximum- and minimum-temperature epochs at Poona, India. From these investigations there emerges the now familiar picture of the temperature field in the lowest 100 m. In clear weather there is a well-marked diurnal variation of temperature gradient, with superadiabatic lapse rates during the hours of daylight and pronounced inversions at night. With overcast skies, and especially with strong winds, the gradient remains close to the adiabatic lapse rate both day and night.

The gradient of temperature near the ground is subject to such large variations with locality and season, time of day and height above the surface that it would be misleading to attempt to give representative values here. Very large gradients, as high as thousands of times the adiabatic lapse rate, are a persistent feature of the temperature field in the first few centimetres above the ground, especially in summer, but the sharp curvature in the temperature-height curve is confined to the first few metres. Sheppard [60] states that on the average the daytime fall of temperature is roughly proportional to the logarithm of the height, but more detailed investigations by Deacon [12] indicate that, in general, the gradient of temperature is inversely proportional to a power of the height over the first 17 m, the index being greater than unity during lapse conditions and less than unity in inversions.

In warm weather the temperature of the air near the ground shows rapid fluctuations of considerable amplitude, as much as several degrees centigrade. Schmidt [58] gives observations by Robitzsch which show that these oscillations have a well-marked diurnal variation in phase with the temperature gradient, the fluctuations tending to die away as the lapse rate approaches the adiabatic value. Sutton [64], from an examination of records obtained in clear warm weather, has found that in these conditions the oscillations decrease as $z^{-0.4}$ over the range 7 m $\leqq z \leqq$ 45 m, but this is an isolated result and a complete picture of the variation with height does not exist at present. During the inversion period the gradient shows large long-period fluctuations unlike those met with in daytime.

There is much less information concerning temperature gradients over the ocean, but data have been given by Wüst [80], Johnson [29], Montgomery [35], Craig [10], Emmons [16], and Sverdrup [71]. Because of the relatively small effects of insolation and long-wave radiation on the sea, the gradients are much less than those found over land and diurnal effects are hardly noticeable, except in shallow landlocked waters.

The exploration of the mean temperature field near the ground is thus virtually complete except in one important respect. It has now become clear that radiation, especially in the longer wave lengths, is quite as important as convection in the problem of heat transfer, and systematic simultaneous records of both temperature gradient and radiative flux are needed to complete the picture of the thermal structure of the lower atmosphere.

The Velocity Field. Following the lead given by the theory of the turbulent boundary layer in aerodynamics, it has now been shown conclusively by many workers that the profile of mean velocity near the ground in conditions of small temperature gradient is adequately represented by a logarithmic law (*e.g.*, of the type proposed by Rossby and Montgomery, equation (7)), provided that the vegetation cover is not too high. For profiles measured above long grass it is necessary to use an empirical modification of the equation, namely

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \ln\left(\frac{z-d}{z_0}\right) \qquad (z \geqq z_0 + d),$$

where d is a zero-plane displacement, of the order of the depth of the layer of air trapped among the plants. Model has used the same equation for the wind profile over the sea.

The most significant feature of the work referred to above is that which emerges from the detailed investigations of Sheppard [61], Paeschke [38], Deacon [12] and others. In conditions of small temperature gradient, air flow near the ground is identical with that observed in the turbulent boundary layer of a fully rough surface in the laboratory, despite the great differences in scale. Sheppard has shown that the Kármán constant k has much the same value as in wind-tunnel work, while Deacon has provided good evidence that the aerodynamical theory of the roughness length is equally appropriate for natural surfaces.

The real difficulties and complexities of the meteorological problem begin to appear when the temperature gradient differs from the adiabatic lapse rate. The general high level of turbulence during the superadiabatic lapse-rate period promotes a free exchange of momentum between higher and lower levels, while the reverse holds during inversions. Thus, in general, the velocity gradient exhibits a diurnal variation, being small in daytime and large at night, but this is not all. Thornthwaite and Kaser [78] and, more recently, Deacon [12] have shown that for nonadiabatic gradients

the $\bar{u}$, ln z plots are no longer linear, but are convex to the $\bar{u}$-axis in the lapse period and concave to the same axis in the inversion period. Deacon concludes that, in all conditions,

$$d\bar{u}/dz = az^{-\beta} \qquad (z \leqq 13 \text{ m})$$

(a independent of z), with $\beta > 1$ for superadiabatic gradients, $\beta = 1$ for the adiabatic lapse rate and $\beta < 1$ for inversions. Thus the logarithmic profile is valid only for adiabatic gradients, and for other conditions Deacon proposes the profile

$$\frac{\bar{u}}{u_*} = \frac{1}{k(1-\beta)}\left[\left(\frac{z}{z_0}\right)^{1-\beta} - 1\right] \qquad (z \geqq z_0),$$

where z_0, as usual, is the roughness length.

For many applications, and particularly those dealing with diffusion, the logarithmic profile makes the relevant differential equations difficult to handle, and in such problems a simple power law, $\bar{u} = \bar{u}_1(z/z_1)^p$, is usually employed as an approximation. This is fairly satisfactory except near a rough surface, where a modification has to be introduced to allow for the effect of the irregularities (*v.* Calder [7] and Sutton [65]). When a law of this type is adopted the effect of temperature gradient is shown by variations in the index p. Values of p ranging from 0.01 (large lapse rate) to 0.77 (large inversion) have been given (Brunt [4], Frost [20]), and it is generally accepted that the value $p = 1/7$ is appropriate for small gradients, except very near the ground or over very high vegetation. For conditions very near the ground ($z < 2$ m), the evidence shows that the logarithmic profile can be assumed without serious error for all except the largest gradients by allowing the roughness length and the Kármán constant to vary with temperature gradient, but if deeper layers are involved, the departure of the velocity profile from the logarithmic form must be taken into account.

The profile of mean velocity near the surface has now been thoroughly explored, but the same cannot be said about the eddy velocities. One of the great difficulties in dealing with atmospheric turbulence compared with that found in wind tunnels is that the natural wind is made up of fluctuations of widely different periods, and the interval of measurement and the response characteristics of the anemometer must always be taken into account, particularly in problems of diffusion. The most reliable data on the velocity fluctuations relate to what Scrase [59] has called "intermediate-scale turbulence," mean values over intervals of the order of a few minutes. It was early demonstrated by Taylor and later confirmed by Scrase [59] and Best [2] that such turbulence is anisotropic near the ground, with the cross-wind component about 50 per cent greater than either the downwind or vertical component at about 2 m over downland in conditions of small lapse rate. Best [2] has given a reasonably complete picture of the behaviour of the three components near the ground by making use of the bi-directional vane, and he concludes that the anisotropy is likely to be negligible at heights greater than about 25 m.

As would be expected, the eddy velocities obey the Maxwell law of frequency distribution, and Best has given the precise formula. Best also examined the variation of the fluctuations with height and found that both the lateral and vertical components increase slowly from 25 cm to 5 m and presumably begin to decrease at higher levels. For winds blowing over the sea there is still less information, but all three components of gustiness are much reduced compared with those over land, probably to about one half.

At the present time, much remains to be learnt concerning the structure of turbulence near the ground. While it is known that the amplitude of the oscillations shows a continuous decrease as the temperature gradient changes from lapse to inversion, it is impossible to quote reliable laws, even empirical, which express this fact quantitatively. There is very little information on the distribution of energy among the fluctuations; Brunt [4] has deduced from the work of Scrase that near the ground and in conditions of small lapse rate most of the eddying energy is associated with oscillations of periods less than five seconds, but a complete picture of the eddy spectrum is not yet available.

The Humidity Field. Investigations into the propagation of high-frequency electromagnetic waves over the surface of the earth have directed attention to the study of water-vapour gradient in the atmosphere, and fairly detailed accounts have been given by Sheppard [60] and Burrows and Attwood [5]. In the daytime, vapour pressure in the lower layers decreases with height even over ground which appears dry, but during the inversion period the gradient may change sign (vapour pressure increasing with height), often with the formation of dew. Sheppard concludes that in the main the vapour-pressure profile conforms fairly closely to a logarithmic law up to 100 m at least, and that there is no marked diurnal variation. Information concerning the distribution of water vapour over the sea is to be found in the papers by Craig, Emmons and Sverdrup cited above.

The Transference of Momentum, Heat and Water-Vapour in the Vertical. In Reynolds theory the transport of momentum by turbulence across a horizontal plane is expressed by the eddy shearing stress $\tau = -\rho\overline{u'w'}$, if the molecular term is disregarded. In the surface layers, it is customary to assume that τ is virtually independent of height and therefore equal to its value at the surface, τ_0. The frictional effect of the ground may also be expressed by means of the skin-friction coefficient C_D by writing

$$\tau_0 = \tfrac{1}{2}C_D\rho\bar{u}^2,$$

where $\bar{u}$ is the mean velocity at some fixed height.

Laboratory investigations show that the skin-friction coefficient of a fully rough surface is virtually independent of the viscosity of the fluid, and in these small-scale experiments C_D usually varies between 10^{-3} and 10^{-2}. Taylor [75], by measuring the approach to the gradient wind in the friction layer from pilot-balloon observations, found C_D for downland to be about 5×10^{-3}, and very similar values were found by Sutcliffe [62], using a somewhat different method.

Hence there is no essential difference between laboratory and open-air results as regards the magnitude of the skin friction.

Direct observations of the eddy shearing stress are few. Sheppard [61] measured τ_0 by observing the deflection of a small plate floating on oil in a streamlined wooden surround placed at the centre of a large concrete surface. For this unusually smooth terrain he found C_D to be about 2×10^{-3} in conditions of small lapse rate. A table of values of C_D, based chiefly on data given by Sheppard and Deacon, has been published by Sutton [63]. The range is from $C_D = 2 \times 10^{-3}$ for a very smooth surface to $C_D = 4 \times 10^{-2}$ for a field of fully grown root crops.

Scrase [59] determined the Reynolds stress at two levels by the direct evaluation of $-\rho\overline{u'w'}$ from ciné records of the motions of light vanes over downland and found, disconcertingly, a fourfold increase of τ with height over the interval 1.5 m–19 m in conditions of small temperature gradient. Thus the only direct observation yet made of the Reynolds stress in the atmosphere does not support the theoretical deduction of invariability with height, and the discrepancy has still to be explained.

Some data relating to superadiabatic lapse rate and inversion conditions have been given by Sheppard [61], who shows that if the Kármán constant k in the logarithmic profile be allowed to vary with temperature gradient, C_D may be regarded as approximately constant for any given terrain. Since it is now clear that the logarithmic law can hold only for gradients near the adiabatic lapse rate, this conclusion loses some of its force. A complete investigation of the drag of the surface or of the transfer of momentum in large temperature gradients has not yet been made.

In the analogous problem of heat flux, the situation is complicated by influences not directly due to turbulence. The balance of heat exchange near the ground involves short-wave radiation (direct, diffuse and reflected), long-wave radiation from the ground and the atmosphere, heat absorbed by the soil and vegetation and by evaporation from the ground or released by the formation of dew, and finally the flux associated with turbulence, that is, by convection, either natural or forced. It is not yet possible to give representative values for all these components, but a recent contribution by Pasquill [39], giving very complete data for a few selected occasions, makes it clear that the flow of heat caused by outgoing long-wave radiation can be as large as, or even greater than, the turbulent flux.

In conditions of calm warm weather Sutton [64], using data for southern England, has shown that in the hours around noon the net upward flux of heat (q) in the lowest 100 m is expressed by

$$q = 3 \times 10^{-3} - 5.4 \times 10^{-8} z \qquad (\text{g cal cm}^{-2} \text{ sec}^{-1}),$$

where the height z is measured in centimetres. This means that near the ground the heat flux, like the eddy shear-stress, is invariable with height during the hottest part of the day, the average value being about 3×10^{-3} g cal cm^{-2} sec^{-1}. A very similar estimate has been made by Priestley and Swinbank [47] from different data, and the figure is also supported by Pasquill's measurements [39], while Sutton [64] has pointed out that the estimate agrees with laboratory data on the loss of heat by natural convection from a small horizontal plate maintained at a constant temperature.

The upward flux of water vapour is the rate of evaporation from the surface, and will be dealt with at greater length in the discussion of small-scale diffusion processes.

To summarize generally what has been achieved in the determination of the physical features of atmospheric turbulence, it may be said that as regards mean properties, the picture is reasonably complete. The outstanding need at present is for a systematic investigation of the fluctuations themselves, and especially the distribution of energy among the various wave lengths, and the correlations between fluctuations in different planes or of different entities. Recent developments in electronic methods should enable many of these problems to be solved.

ANALYSIS OF LARGE-SCALE PROCESSES

The initial successes of the theory of atmospheric turbulence came in the pioneer work of Taylor, Ekman, Schmidt and Richardson on large-scale phenomena such as the approach of the wind in the friction layer to the geostrophic velocity, the propagation of the diurnal wave of temperature from the surface, and the large-scale diffusion of water vapour and pollution. These are the "classical" problems of atmospheric turbulence; they are usually dealt with at length in textbooks of dynamical meteorology and for this reason only a brief account will be given here.

If we consider first the slowing down of the free-stream (geostrophic) velocity by the friction of the earth, the effect of turbulence on the transport of momentum may be expressed quite generally by the introduction of virtual stresses τ_{xz} and τ_{yz}. If $\bar{u}$ and $\bar{v}$ are the components of the mean velocity, assumed horizontal, along axes pointing east and north respectively, equations (2), neglecting any change of pressure gradient with height, become

$$\frac{\partial \bar{u}}{\partial t} - 2\omega\bar{v} \sin\phi = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{1}{\rho}\frac{\partial \tau_{xz}}{\partial z},$$

$$\frac{\partial \bar{v}}{\partial t} + 2\omega\bar{u} \sin\phi = -\frac{1}{\rho}\frac{\partial p}{\partial y} + \frac{1}{\rho}\frac{\partial \tau_{yz}}{\partial z}.$$

The terms involving $\omega \sin\phi$ (ω= angular velocity of earth, ϕ = latitude) are the Coriolis accelerations.

The component eddy stresses τ_{xz} and τ_{yz} must vanish or become small at the top of the friction layer, but apart from this there is no a priori knowledge of their behaviour, and the equations can be solved only when some additional information is deduced or postulated concerning these stresses. An obvious first step is to adopt the analogy with molecular theory by expressing the shear stresses as the product of a constant virtual viscosity and the velocity gradient. The solution obtained in this way is usually displayed as the familiar

"Ekman spiral," a graph of $\bar{u}$ against $\bar{v}$ with the geostrophic velocity as the limit point.

Although the assumption of a constant eddy viscosity (K) or exchange coefficient ($A = K\rho$) can be no more than a crude first approximation, the Ekman spiral undoubtedly gives a picture of the approach to the geostrophic wind which is easily grasped and which, except near the surface, does not depart too far from the truth. The most striking feature of this early work is that the concept of eddy viscosity implies values of K at least ten-thousand or a hundred-thousand times greater than the kinematic viscosity of air.

In the corresponding problem of the transport of heat by eddies, the assumption of a constant eddy conductivity leads to the familiar equation

$$\frac{\partial T}{\partial t} = K \frac{\partial^2 T}{\partial z^2},$$

where T is absolute temperature (Brunt [4]). If the diurnal variation of surface temperature be given by

$$T = T_0 + B \sin qt,$$

the solution which expresses the progress of the diurnal wave upwards is

$$T = T_0 + Be^{-bz} \sin (qt - bz), \qquad (10)$$

where $b = \sqrt{q/2K}$. This solution has been used by Schmidt and Taylor for the analysis of the Eiffel Tower observations and by Johnson [28], Best [2], and Johnson and Heywood [30] in a very detailed examination of the temperature field up to 100 m above the surface. These early researches indicated values of K, the eddy conductivity, of the same order of magnitude as those found for the eddy viscosity.

Finally, if $\bar{\chi}$ represents the mean concentration or mass per unit volume of some easily identified constituent of the atmosphere, such as water vapour, smoke or dust, the assumption of a constant eddy diffusivity leads to Fick's equation

$$\frac{D\bar{\chi}}{Dt} = K\nabla^2\bar{\chi},$$

where D denotes "differentiation following the motion" and ∇^2 is the Laplace operator. For moderate distances of travel, say not exceeding a few hundred metres, the values obtained for the eddy diffusivity are of the same order of magnitude as those found for the turbulent transport of momentum and heat.

The picture of eddy diffusion given by this early work is consistent and convincing—the atmosphere behaves like a gas of greatly enhanced viscosity, conductivity and diffusivity, and the analogy is reinforced by the fact that all three coefficients appear to have much the same order of magnitude. Detailed examination of the results, however, reveal discrepancies which are far too serious to be ignored or ascribed to errors of measurement. Best [2] and Johnson and Heywood [30], by applying the solution (10) to a succession of shallow layers, showed that to explain the observations the "constant" eddy conductivity must be made to increase rapidly with height, especially in clear summer weather. The solution (10) indicates that the phase of the diurnal wave should change linearly with height, whereas Best found that in reality the change of phase with height is relatively slow, being approximately as the one-fifth root of the height. Cowling and White [9] made a searching analysis of the Leafield temperature results and concluded that the eddy conductivity must vary with height and have a diurnal variation, and they found additional support for a conclusion reached by Chapman in 1925 that the temperature field near the surface cannot be explained on the basis of eddy conduction alone, and that some other factor (possibly radiation) must play an important part. Richardson [52] obtained equally striking results in a series of notable researches on the large-scale diffusion of matter. If the eddy diffusivity were truly constant, the scatter of matter originally concentrated at a point would follow the law deduced by Einstein for Brownian motion, namely

$$\sigma^2 = 2Kt,$$

where σ is the standard deviation of the distances travelled by the particles in time t. By applying this equation to processes ranging from the drift of thistledown over a few metres to the scatter of free balloons and volcanic dust over many kilometres, Richardson demonstrated that K in this equation must increase indefinitely with the distance travelled, values as high as 10^8 cm^2 sec^{-1} being found for the longer distances.

The recognition of the fact that the mixing processes of the atmosphere cannot be regarded as a kind of enlarged version of molecular diffusion marks the beginning of the modern phase of the theory of atmospheric turbulence. Obviously, if the concepts of eddy viscosity, conductivity and diffusivity are to be retained, it is no longer possible to regard these coefficients as independent of position in the field, but the next step in the development of the theory is by no means obvious. To fix ideas, consider the relatively simple problem of the diffusion of smoke from a continuous source at ground level, a matter of importance in the military problem of the screening of targets and for which the experimental results are known to a high degree of accuracy for conditions of small temperature gradient and moderate distances of travel (Sutton [67]). The solution of Fick's equation for a steady continuous point source is

$$\bar{\chi}(x, y, z) = \frac{Q}{4\pi Kr} \exp\left\{-\bar{u}\frac{(y^2 + z^2)}{4Kr}\right\},$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the distance from the source Q; x, y and z are measured downwind, across wind and vertically; and $\bar{u}$ is the constant mean wind (Roberts [54]). Thus the concentration of smoke on the axis of the cloud ($y = z = 0$) should decrease inversely as $1/x$, but the measurements show that the actual decrease is as $1/x^{1.76}$ in conditions of small temperature gradient. Hence K must be made to increase with distance from the source if the measured concen-

trations are to be made to agree with the mathematical solution. There are obvious difficulties in ascribing physical reality to a quantity whose value depends on the distance of the sampling point from an arbitrary point in space and any variation in K must be regarded, more rationally, as an expression of the fact that (as Richardson puts it) in a turbulent fluid the average rate of separation of a pair of marked particles is a function of their distance apart, a type of diffusion quite unlike that contemplated in the kinetic theory of gases. In particular, it does not seem possible to assign any upper limit to the "size" of eddy which can take part in atmospheric diffusion, and there is no definite "scale" of atmospheric turbulence.

The most natural and acceptable form of variation of K is to allow the coefficient to increase with the depth of the layer under consideration. The mathematical technique required for the solution of diffusion problems is considerably simplified if the interchange coefficients are assumed to vary as a simple power of the height. The problem of the approach of the wind in the friction layer to the geostrophic velocity has been dealt with very thoroughly by Köhler [33] for the case $K \propto z^m$, $m \geqq 0$. The solution involves Bessel functions and the results show a fair measure of agreement with observation. Rossby and Montgomery have also discussed the same problem by making the reasonable assumption that the mixing length (and hence the eddy viscosity) first increases linearly with height in a relatively shallow surface layer (thickness about 100 m) and then decreases slowly so as to allow a small "residual turbulence" at the top of the friction layer.

In the problem of heat transfer Brunt [4], following the lines of Taylor's early work, has shown that if T be the absolute temperature of the air at height z, the equation of eddy conduction is

$$\rho \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left\{K\rho\left(\frac{\partial T}{\partial z} + \Gamma\right)\right\},$$

where Γ is the adiabatic lapse rate. This implies that the net turbulent flux of heat across a horizontal surface is proportional to the difference between the existing lapse rate and the dry-adiabatic lapse rate, so that in a stable atmosphere, the flow of heat due to eddy motion is downwards, and the ultimate effect of mixing is to produce an atmosphere with constant lapse rate equal to the adiabatic value. This equation and the conclusions drawn from it have been the subject of much discussion in recent years, but this aspect will be dealt with at greater length in the section dealing with convection.

From the purely mathematical aspect the equation of conduction with K variable has been well explored; Haurwitz [24] has given the solution for ρ = constant, $T = T_0 + B \sin qt$ on $z = 0$, $K = a_0 + a_1 z$ (a_0, a_1 constant) in terms of ker and kei functions, and Köhler [33] has published a very thorough treatment of the case in which $K \propto z^m$, $m \geqq 0$, the solution in this case being expressed in terms of Bessel functions of imaginary argument. The implications of this work are considered later.

Turbulence in the General Circulation. In 1921, Defant [13] introduced a picture of the depressions and anticyclones of the synoptic meteorologist as "eddies" in the general circulation, a theory which leads to the concept of *Grossturbulenz* (Lettau), the transport of heat, momentum and water vapour on a planetary scale. A clear account of this work is to be found in Chapter XII of Lettau's *Atmosphärische Turbulenz*. Lettau adopts the ordinary interchange coefficient theory of turbulent transport and by consideration of the deviations from the geostrophic wind finds, for example, that the meridional component of the *Grossaustausch* is of the order of 10^8 g cm^{-1} sec^{-1}, with a "mixing length" of the order of several degrees of longitude.

A very different approach, not involving mixing-length concepts is that given by Priestley [46] in a recent paper in which he develops a method whereby the meridional flux of heat, water vapour and momentum can be evaluated from upper-air soundings. From a preliminary survey Priestley concludes that the atmospheric eddy flux of heat is of the magnitude required to equalize the heat losses and gains in adjacent zones of the earth and atmosphere as a whole, and that the zonal stresses arising from deep meridional currents can maintain wide zonal circulations against the effects of surface friction.

Richardson's investigations into large-scale diffusion have already been mentioned; in 1932 Sutton [66], by means of a semi-empirical theory, showed that in a turbulent medium the Einstein law of molecular diffusion (Brownian motion) should be replaced by the equation

$$\sigma^2 = C^2(\bar{u}t)^m. \tag{11}$$

Here C is a generalized coefficient of diffusion and m is a constant whose value is about 1.75. This equation is certainly satisfied for diffusion over distances of the order of a few hundred metres, and Richardson's data, re-analysed on the basis of this equation, indicate a strong probability that the law is valid also for distances of the order of hundreds of kilometres. In assessing the value of this conclusion it should be borne in mind that the data on which it is based are extremely crude for the greater distances, but the evidence available so far is that a law of the type (11) is capable of representing atmospheric diffusion, to a first approximation at least, over a very wide range.

It will be evident from the discussion above that although the broad features of many large-scale atmospheric processes can be explained reasonably well by the simpler forms of turbulence theory, as yet little has been attempted in the way of detailed analysis. Many opportunities present themselves here. The recent brilliant American work[2] on the structure of thunderstorms has directed the attention of meteorologists to the problems of free jets and of the general mechanism of the entrainment of air by turbulent mixing on the boundary of a thermal current. Much remains to be

2. Described in *The Thunderstorm* by H. R. Byers and others, Supt. of Documents, Washington, D. C., 1949.

explained concerning the details of large cellular convective motion, first studied in the laboratory by Bénard and investigated mathematically by Rayleigh, Jeffreys and others. The whole problem of turbulence in the upper atmosphere has recently arisen in an acute form in relation to the design of large aircraft. Finally, the applicability of the Reynolds process of averaging in dealing with major atmospheric motions is still uncertain and it is here, perhaps more than anywhere else, that the statistical theories are likely to make the most decisive contributions.

SMALL-SCALE DIFFUSION PROCESSES

Because of their importance in military operations and in studies of atmospheric pollution, specialized problems relating to the spread of suspended matter (smoke and gas) over distances of the order of a few kilometres have been the subject of intensive research, both practical and theoretical. In experimental investigations a measure of control is possible, and the data on diffusion thus obtained are among the most reliable and accurate in micrometeorology. A summary of the properties of continuously generated smoke clouds, based on extensive trials conducted at Porton, England, has been given by Sutton [67]; these refer exclusively to conditions of small temperature gradient and, as yet, no corresponding set has been published for inversions or for conditions of a superadiabatic lapse rate.

Concurrently with the experimental work, there has been a great deal of activity on the theoretical side, with the result that a workable but semi-empirical theory of turbulent diffusion has been built up and verified for conditions of neutral equilibrium (adiabatic lapse rate). This work will now be summarized briefly.

Properties of Continuously Generated Clouds in Conditions of Small Temperature Gradient. The smoke cloud from a continuous point-source stretches downwind in a long cone, and measurements of concentration taken at fixed points are to some extent dependent upon the period of sampling. This is because the natural wind is made up of fluctuations of all periods, and the "instantaneous" aspect of the cloud, being mainly influenced by the small-scale eddies, differs considerably from the "time-mean" aspect. The narrow "instantaneous cone" swings slowly over a wider front and is contained within an enveloping "time-mean cone." The properties which have been measured almost invariably refer to the "time-mean" aspect, that is, the concentrations at fixed points are averages over periods of time of not less than three minutes.

For a continuously generated cloud, the concentration of smoke at any point is directly proportional to the strength of the source and approximately inversely proportional to the mean wind speed, while the cross-wind and vertical distributions of concentrations are approximately of the "normal law of error" type. From the point of view of the mathematician, the results of greatest importance are: The central or peak concentration in the cloud from a continuous point-source decays as $1/x^{1.76}$ (x = distance downwind) and as $1/x^{0.9}$ in the cloud from an infinite cross-wind continuous source, and the maximum concentrations at $x = 100$ m in a mean wind of 5 m sec^{-1} are

point-source of 1 g sec^{-1} ..... 2 mg m^{-3}

infinite line-source of 1 g sec^{-1} m^{-1}.... 35 mg m^{-3}.

The fundamental problem for the mathematical physicist is to find means whereby these properties can be derived from measurements of the relevant meteorological factors, such as the profile of mean velocity, gustiness and temperature gradient.

Theoretical Aspects. The general equation of diffusion in a turbulent medium is

$$\frac{\partial \bar{\chi}}{\partial t} + \bar{u}\frac{\partial \bar{\chi}}{\partial x} + \bar{v}\frac{\partial \bar{\chi}}{\partial y} + \bar{w}\frac{\partial \bar{\chi}}{\partial z} = \frac{\partial}{\partial x}\left(K_x \frac{\partial \bar{\chi}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial \bar{\chi}}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial \bar{\chi}}{\partial z}\right), \tag{12}$$

where $\bar{\chi}$ is the mean concentration of smoke, the density of the air being supposed constant. It is convenient to take axes in which x is measured in the direction of the mean wind, y across wind and z vertically, so that $\bar{v} = \bar{w} = 0$. For these localized problems the mean wind may be supposed to vary only with height.

Two-Dimensional Problems. In the case of a long line source across wind, or an area source in which lateral edge effects are negligible, equation (12), neglecting downwind diffusion in comparison with vertical diffusion, reduces to

$$\bar{u}(z)\frac{\partial \bar{\chi}}{\partial x} = \frac{\partial}{\partial z}\left(K_z \frac{\partial \bar{\chi}}{\partial z}\right)$$

for the steady state. To deal further with this equation means that $\bar{u}(z)$ and K_z must be known explicitly. We shall consider here only power-law formulations; so far no solutions have been published in which $\bar{u}$ is expressed in terms of ln z. Most investigations assume the Schmidt "conjugate power-law" relation, that is, if $\bar{u} = \bar{u}_1(z/z_1)^p$, then $K = K_1(z/z_1)^{1-p}$, but since this is equivalent to the assumption that the eddy shearing-stress $\tau = K\rho\, d\bar{u}/dz$ is invariable with height, the resulting equation can hold only in a relatively shallow layer ($z \leqq 25$ m) near the surface. This limitation is of importance in questions dealing with the variation of humidity profile in air passing from sea to land (*v.* Booker), and results obtained in investigations which assume the conjugate power-law relation to hold in deeper layers must be regarded with considerable doubt, and any close agreement with observation to be a matter of chance rather than of sound reasoning.

With the conjugate power-law relation, the two-dimensional steady-state equation becomes

$$\frac{\bar{u}_1}{K_1} z_1^{1-2p}\frac{\partial \chi}{\partial x} = z^{-p}\frac{\partial}{\partial z}\left(z^{1-p}\frac{\partial \chi}{\partial z}\right). \tag{12a}$$

By a suitable change of variables this equation may be converted to the form

$$\frac{\partial \chi}{\partial x} = \frac{1-2p'}{z}\frac{\partial \chi}{\partial z} + \frac{\partial^2 \chi}{\partial z^2}; \qquad p' = p/(2p+1),$$

(W. G. L. Sutton [70]), which will be recognized as a generalization of the classical equation for the conduction of heat in a solid. This is probably the most useful way of regarding the equation, for the boundary conditions which occur in problems of atmospheric diffusion are entirely analogous to those familiar in the theory of heat flow. Solutions of this equation for various types of boundary conditions (chiefly those which arise in the problem of evaporation) have been discussed in considerable detail by W. G. L. Sutton [70], who based his analysis on Goursat's treatment of the equation of conduction of heat. Jaeger [27] has recently given an elegant and concise discussion of the diffusion equation by the method of the Laplace transform.

The boundary conditions for the problem of the steady continuous infinite cross-wind line-source of smoke are easily specified; the concentration tends to zero at infinity and increases without limit as the line of emission is approached, while the effect of the earth's surface is represented by the condition that there is no net flow across the plane $z = 0$. The solution is then easily obtained as

$$\bar{\chi}(x, z) = \frac{\text{const}}{\bar{u}_1^{1/(1+2p)}\, x^{(p+1)/(2p+1)}} \cdot \exp\left\{\frac{-\text{const}\ \bar{u}_1^{2p/(1+p)}\, z^{(1+2p)}}{x}\right\}, \tag{13}$$

in which the various "constants" are easily determined by the "continuity" condition.

In the problem of evaporation from a free-liquid or saturated area on the plane $z = 0$, of infinite extent across wind and of finite length downwind, the boundary conditions are not as obvious. In his treatment of the problem O. G. Sutton [68] introduced the condition that the vapour concentration attains the saturation value (χ_s) at all points on the wetted area, so that the problem of evaporation becomes that of finding the strength of the area source which will maintain this constant concentration on the surface despite the removal of vapour by the turbulent air stream above. The solution of the problem constituted by equation (12a) and the boundary conditions can then be obtained as an incomplete gamma function, namely,

$$\bar{\chi}(x, z) = \chi_s\left[1 - \text{const} \cdot \Gamma\left(\frac{\text{const}\ \bar{u}_1^{2p/(1+p)}\, z^{1+2p}}{x},\ \frac{p}{2p+1}\right)\right], \tag{14}$$

from which the total rate of evaporation can be found by integrating the local rate of evaporation $\left[K_z \frac{\partial \bar{\chi}}{\partial z}\right]_{z=0}$ over the area (see W. G. L. Sutton [70], Jaeger [27], Frost [21], Pasquill [40]).

Three-Dimensional Problems. For problems involving point or short line sources, or areas finite across wind, the term $\frac{\partial}{\partial y}\left(K_y \frac{\partial \bar{\chi}}{\partial y}\right)$ must be retained. So far very little progress has been made with these problems. In the first place, the form for K_y cannot be settled as easily as that for K_z, since there is no resultant shear-stress across wind and hence no counterpart of the "conjugate power-law" theorem. It is hardly possible to consider K_y as a function of y, that is, dependent upon distance from an arbitrary axis, and K_y must therefore be made a function of height (z). This, however, introduces considerable mathematical difficulties. It is known, for example, that a solution satisfying the boundary conditions for a continuous point-source can easily be found (in the usual exponential form) if both K_y and $\bar{u}$ are constant or if K_y and $\bar{u}$ are proportional to the same power of z, but these solutions do not agree with the experimental data. In the problem of evaporation, a formal solution for K_y, K_z and $\bar{u}$ constant, applicable to a rectangular area, can be found in terms of Mathieu functions, and Davies [11] has investigated the case in which K_y and $\bar{u}$ are proportional to the same power of z for an area of parabolic shape, but the solution for K_y proportional to an arbitrary power of z, applicable to a closed area, has not been found. There is need for a detailed pure mathematical examination of this type of equation, and progress in this branch of atmospheric diffusion is bound to be slow until such an investigation has been made.

Formulas for K_z, Statistical Theory. In applying the analysis described above to specific problems, it is usually assumed that the velocity profile is known to a high degree of accuracy, and that from this, the value of the friction velocity u_* can be found for any given mean velocity. Since

$$K_z = \frac{\tau/\rho}{d\bar{u}/dz} = \frac{u_*^2}{d\bar{u}/dz},$$

it follows that K_z is also known explicitly at all points in the layer in which the shearing stress is constant. Thus the two-dimensional problem of diffusion depends essentially on the accurate determination of the velocity profile near the surface, the underlying assumption being that the diffusion of mass by eddy motion is identical with that of the diffusion of momentum.

This method, initiated by Sutton, has been followed by Calder [7] in a very exact study of two-dimensional diffusion over surfaces covered by both short and long grass. Calder uses the logarithmic profile to determine the roughness length, friction velocity and zero-plane displacement (for the long-grass case), but in the equation (12a) he replaces the logarithmic profile by an approximate power-law profile and uses solutions (13) and (14) above. The theoretical results are in very good agreement with the observational data. The method cannot be used for three-dimensional problems, and for such problems recourse must be had to the statistical approach devised by Sutton [67].

This theory is based upon Taylor's "Diffusion by Continuous Movements" [73], using as the starting point the correlation coefficient $R(\xi)$ between eddy velocities at different times. For flow over a smooth surface, dimensional arguments indicate that $R(\xi)$ may be represented by

$$R(\xi) = \left(\frac{\nu}{\nu + \overline{u'^2}\,\xi}\right)^n, \tag{15}$$

where ν is the kinematic viscosity of the air and n is a parameter expressing the degree of turbulence in the fluid. The value of n is found from observations of the velocity profile. Using mixing-length concepts, an explicit power-law form for K_z can then be found, which when inserted in the equation of diffusion leads to results in excellent agreement with observation from a smooth saturated surface (Sutton [68], Pasquill [40]). For diffusion over rough surfaces, the kinematic viscosity in (15) must be replaced by the macroviscosity $N = u_* z_0$; the resulting expressions are then in good agreement with observation (Sutton [65]).

For three-dimensional problems Sutton abandons the mixing length *cum* differential equation approach and proceeds by finding expressions which satisfy Taylor's equation

$$\sigma^2 = 2\overline{u'^2} \int_0^T \int_0^t R(\xi)\, d\xi\, dt,$$

the boundary conditions and the equation of continuity. For this purpose it is necessary to introduce generalized diffusion coefficients C_y and C_z, which are found as explicit functions of the mean gustiness, the parameter n occurring in the velocity profile, and either the kinematic viscosity or the macroviscosity, according as the underlying surface is smooth or rough. The expression for the concentration from a line-source obtained in this way is not quite as accurate as that derived from the differential equation, but is adequate for many purposes, and the method has the advantage of providing a simple solution for the point-source problem where, as yet, the mixing length *cum* differential equation approach has failed. Recently, Sutton [69] has extended this work to cover the case of an elevated source such as a factory chimney.

Detailed Study of Evaporation. The work described above has shown that the rate of evaporation from a small saturated or free-liquid area can be calculated with considerable accuracy, at least in conditions of neutral vertical equilibrium and provided that edge losses can be disregarded. Since the mean rate of evaporation decreases with the length of the area downwind, and also because it is uncertain if a small area can reflect adequately the effect of changing stability (Sutton [67]), it is extremely doubtful if the conventional "evaporimeter" or "evaporation tank" is of any real use in assessing the rate of evaporation from areas of moderate size, such as reservoirs or small lakes.

For the estimation of evaporation from natural surfaces without the aid of evaporimeters, two methods are possible. One, initiated by Ångström for lakes and later applied by Penman [42] to land, depends upon an accurate evaluation of the remaining terms in the equation for heat balance. No knowledge of the process of turbulent transport is required. The second method, used by Sverdrup and others for evaporation from the sea and by Thornthwaite and Holzman [77] for land, uses the theory of eddy diffusion to evaluate the flux of water vapour from the wind and water-vapour profiles. This procedure has been subjected to a critical examination by Pasquill [41], who concludes that the Thornthwaite-Holzman formula

$$E = \frac{\rho k^2 (q_1 - q_2)(u_2 - u_1)}{(\ln z_2/z_1)^2}$$

(E = rate of evaporation, q_1 , q_2 and u_1 , u_2 = specific humidities and mean velocities at heights z_1 and z_2 , ρ = air density, k = Kármán's constant ≈ 0.4) is valid for conditions of neutral equilibrium and may be used in other conditions if errors up to about 20 per cent in the rate of evaporation can be tolerated.

Space does not permit more than a brief mention here of the large amount of work which has been done on evaporation from the ocean, but some reference should be made to the "evaporation coefficient" of Montgomery [35], defined as

$$\Gamma = -\frac{1}{e_s - e_a} \frac{de}{d(\ln z)},$$

where e_s = vapour pressure at the surface and e_a = vapour pressure at a standard height. The rate of evaporation from the sea surface is then

$$E = \rho k \gamma_a \Gamma (q_s - q_b) W_a,$$

where k is Kármán's constant, $\gamma_a = u_*/W_a$, where W_a = velocity at height a, and q_s and q_b are the specific humidities at heights a and b. Sverdrup [71] has shown that with certain assumptions $\Gamma = 1/\ln\left(\frac{b + z_0}{z_0}\right)$, where z_0 is the roughness length, and has made a critical examination of the validity of Montgomery's concept.

Richardson's Diffusion Theory and Its Relation to Recent Developments. We conclude this short summary of the diffusion problem by mention of some work by Richardson which may prove to be of considerable significance. In 1926, Richardson [50], showed in his "Distance-Neighbour Graph" theory, that if l is the projection of the separation of a pair of marked particles on a fixed direction, and q is the "number of neighbours per unit of l,"

$$q(l) = \int_{-\infty}^{\infty} \chi(x)\chi(x + l)\, dx,$$

where χ is the concentration. The equation of diffusion then becomes

$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial l}\left\{F(l) \frac{\partial q}{\partial l}\right\},$$

where $F(l)$ is a kind of diffusion coefficient. If l_0 is the initial value of l, and l_1 its value t seconds later,

$$F(l_0) = \frac{(\text{mean of } (l_1 - l_0)^2 \text{ for all pairs})}{2t}.$$

Observations by Richardson and Stommel [53] on small objects floating in water indicate that

$$F(l) \approx 0.07 l^{1.4}.$$

This is a result of considerable importance, because the recent statistical theories of Weizsäcker and Heisen-

berg (page 497) predict a similar law of diffusion. It is possible that here is the first indication of developments of considerable importance for meteorology.

Further progress in the study of small-scale processes, so important for many aspects of civilized life, depends on a number of factors. Micrometeorology demands measurements whose accuracy approaches that attained in the laboratory, and for which the development of specialized instruments, highly sensitive and yet sufficiently robust to be used in the open, is imperative. On the mathematical side it must be admitted that present theories, although often astonishingly successful for limited classes of problems, are unsatisfactory in that they involve a large element of empiricism, and rigour has often been sacrificed for simplicity of treatment. There is, in particular, a need to examine both practically and theoretically the basic assumptions of the various theories, such as the form of the various correlation functions and of the eddy spectrum over a wider range of conditions. The characteristic meteorological problem, that of diffusion in large density gradients, is still unsolved and is likely to remain so until further progress has been made in the dynamics of stratified fluids. Some of the attempts to solve this, perhaps the most difficult problem in turbulence, are discussed in the next section.

PROBLEMS ARISING FROM CHANGES IN THE DENSITY GRADIENT

Much, if not most, of the preceding work could be fairly described as a straightforward application of the aerodynamical theory of the turbulent boundary layer to meteorology. The feature which sharply differentiates the meteorological problem from those of normal aerodynamics is undoubtedly the influence of the density gradient on the nature of the flow. In other words, the meteorologist must take into account the effects of the gravitational field, and there is very little in the way of wind-tunnel investigation to guide him here.

The Richardson Criterion. The basic result in the theory of turbulence in a gravitational field is that of Richardson [51], who enunciated the criterion that the kinetic energy of the eddying motion will increase or decrease if the rate at which energy is extracted by the Reynolds stresses exceeds or falls below that at which work has to be done against gravity by the turbulence. From this, it is easy to show that turbulence will increase or die away according as the Richardson number

$$Ri = \frac{g\left(\frac{\partial T}{\partial z} + \Gamma\right)}{T\left(\frac{\partial \bar{u}}{\partial z}\right)^2}$$

(T = absolute temperature of the environment) is less or greater than the ratio of the eddy viscosity to the eddy conductivity. If these two coefficients are supposed identical, we have

turbulence increases if $Ri < 1$,

turbulence decreases if $Ri > 1$,

that is, the critical value (Ri_{crit}) of the Richardson number is unity.

In his derivation of this criterion Richardson limited its application to a system possessing a very small amount of turbulence ("just-no-turbulence"). The matter has since been the subject of intensive investigations, chiefly theoretical. Taylor [76] and Goldstein [23] showed that in the case of an inviscid incompressible fluid with a linear velocity profile and a continuous density distribution, $Ri_{crit} = 0.25$, but the most detailed and realistic investigation is that of Schlichting [57] on the decay of turbulence in the boundary layer of a smooth flat plate. Schlichting found that Ri_{crit} varied between 0.041 and 0.029, depending on the inertia effects of the density distribution, a result which was later confirmed by Reichardt in the Göttingen "hot-cold" wind tunnel. Recently, a detailed investigation of the mathematical aspects of Richardson's derivation has been made by Calder [8], who finds that the inclusion of certain terms, neglected by Richardson because of his assumption of "just-no-turbulence," changes the form of the criterion to $Ri_{crit} = 1 - \delta$, $\delta > 0$, but he was unable to give a definite value for δ, except that δ must be small if the initial degree of turbulence is small.

There have been several attempts to ascertain the value of Ri_{crit} in the atmosphere [63]. Durst agrees with Richardson in finding $Ri_{crit} = 1$, but Paeschke [38] concludes that the Schlichting value, $Ri_{crit} = 0.04$, is appropriate. Some of the best evidence is that assembled by Deacon [12] who proposes $Ri_{crit} = 0.15$ for conditions near the ground, while for the free atmosphere Petterssen and Swinbank [43] suggest $Ri_{crit} = 0.65$. It is evident from these results that at the present time the question of the exact value of the critical Richardson number (if indeed it exists) is one of the most open in meteorology.

The Heat Flux Equation and the Problem of Convection. It has been known for some time that the early assumption of the identity of the coefficients of eddy viscosity (K_M) and eddy conductivity (K_H) is open to considerable doubt; thus although in the lower layers of the atmosphere the transfer of momentum can be explained by taking $K_M \propto z^m$, where $0 \leqq m \leqq 1$, the propagation of the diurnal temperature wave in warm weather seems to require $K_H \propto z^{m'}$, where $1 \leqq m' \leqq 2$. These conclusions are based upon the assumption that the eddy flux of heat is proportional to the gradient of potential temperature (Taylor) or to the difference between the observed gradient and the adiabatic lapse rate (Brunt). Ertel [18], in a series of papers, questions this assumption because it neglects buoyancy effects, and concludes that the flux is more likely to be proportional to the gradient of temperature itself, a result which in turn has been criticized by Prandtl [45] on the grounds that the analysis can only be valid for occasions of calm or light winds and large lapse rates.

Ertel's main conclusion is supported by the later work of Priestley and Swinbank [47], who consider that the flux of heat has the form

$$q = \rho c_p \left\{ -\overline{w'l}\left(\frac{\partial T}{\partial z} + \Gamma\right) + \overline{w'T''} \right\},$$

where w' is the vertical component of the eddy velocity and T'' is the temperature anomaly of the eddy at the beginning of its upward motion. The first term in the brackets is that derived in the classical theory, but the second term is independent of the gradient of potential temperature and depends essentially on the existence of a correlation between the vertical velocity and the initial temperature of the eddy, which need not be that of the layer from which it came. The second term is therefore due entirely to buoyancy, whereas the first term could arise if the air were forced upwards by purely dynamical means. Unfortunately, the term $\overline{w'T''}$ cannot be isolated from the general heat flux and measured separately, and its magnitude can only be inferred from arguments of a general nature, but the theory does give a fairly satisfactory explanation of certain large-scale phenomena, such as the superadiabatic lapse rates of the higher regions of the troposphere in clear non-subsiding air. The same problem has recently been considered by Montgomery [36].

The problem of natural convection, that is, of heat transport due to the upward motion of the air in calm or very light wind conditions has been considered by Sutton [64], who, starting from the assumption that the vertical flux of heat is proportional to the difference between the observed temperature gradient and the adiabatic lapse rate, shows that the observations then necessarily imply a very rapid increase of K_H with height (as $z^{1.75}$) in hot, clear weather. On the assumption that the intensity of the turbulent upward currents is determined by the balance between their dissipation into smaller eddies and their rate of loss of potential energy, Sutton shows that certain simple relations must exist between the eddy velocity, the mixing length for heat transfer, the temperature gradient, the temperature oscillations and the eddy conductivity, provided that the upward flux of heat does not vary with height. These relations are in harmony with Johnson and Heywood's observations of the temperature field during the mid-hours of a clear summer day. Sutton also showed that a model of convection in which the ground is assumed to act as a plane instantaneous source of heat at intervals of the order of half-a-minute or so could provide an explanation of the temperature field found near the ground in warm weather. This implies that the typical convectional eddy behaves like a "bubble" which rises because of its buoyancy and grows by entraining cold air by turbulent mixing as it moves.

From this brief survey it will be seen that the solution of the problem of the transfer of heat from the ground to the air, or vice versa, is still far from complete. At the present time the main need is for accurate observations of elements such as the temperature and wind oscillations and their correlations (and hence the eddy flux) and the radiative flux. On the credit side, it appears to be fairly well established that in conditions of high lapse rate there is a significant difference between the rate of transport of heat and momentum near the ground, even when radiation is taken into account (Pasquill), but a complete explanation of the temperature field in both unstable and stable conditions is still to seek. When buoyancy effects are small compared with purely frictional effects, that is, in the problem of forced convection, it seems that there is little difference in the transport of heat and momentum by eddy motion.

The Diffusion of Matter and Other Studies in Non-adiabatic Gradients. Among the early attempts to extend the theory of diffusion to nonadiabatic gradients is that of Rossby and Montgomery [55], who, in effect, take the mixing length to be given by

$$l = kz/\sqrt{1 + \sigma Ri},$$

where σ is a "constant." This expression was used by Sverdrup in a study of turbulence over a snow field; his observations indicate that the value of σ is about 11. In a later contribution Sverdrup examined Best's velocity profiles in the light of this relation and found a large increase in σ as conditions changed from instability to stability. This work was later extended by Holzman [26], who suggested that

$$l = kz\sqrt{1 - \sigma_1 Ri},$$

where σ_1 is another constant so that $l = 0$ when $Ri = 1/\sigma_1$, when turbulence should be extinguished.

The Rossby-Montgomery and Holzman formulations have been subjected to a critical examination by Deacon [12], who finds that the Holzman expression agrees well with observations, σ_1 being independent of both roughness and stability, with a mean value about 7. This implies that the turbulent mixing-length becomes negligibly small if Ri is about 0.14.

Deacon [12] has applied these studies to an examination of two-dimensional diffusion in nonadiabatic gradients (chiefly superadiabatic lapse rates), using data on the travel of gas obtained at the Experimental Station, Alberta, Canada. The results show a certain measure of agreement with the theory, except in very large lapse rates, in which the experimental data are rather irregular.

To sum up, it is clear that a satisfactory theory of diffusion in large inversions and large lapse rates has yet to emerge, and this is particularly true for three-dimensional problems. It is unfortunate that the theory is most deficient where the need is greatest, that is, in conditions of large inversions, when even small sources of pollution can produce high concentrations, and the problem is one which should receive urgent attention from meteorologists if only because of its importance in the life of highly industrialized communities.

REFERENCES

1. Batchelor, G. K., "Kolmogoroff's Theory of Locally Isotropic Turbulence." *Proc. Camb. phil. Soc.*, 43:533–559 (1947).
2. Best, A. C., "Transfer of Heat and Momentum in the Lowest Layers of the Atmosphere." *Geophys. Mem.*, No. 65 (1935).

3. Boussinesq, J., *Théorie de l'écoulement tourbillonant et tumultueux des liquides* .... Paris, Gauthier-Villars, 1897.
4. Brunt, Sir David, *Physical and Dynamical Meteorology*, 2nd ed. Cambridge, University Press, 1939.
5. Burrows, C. R. (chairman), and Attwood, S. S. (editor), *Radio Wave Propagation*. New York, Academic Press, 1949. (Consolidated Summary Tech. Rep. of the Committee on Propagation of the National Research Committee.)
6. Calder, K. L., "A Note on the Constancy of Horizontal Turbulent Shearing Stress in the Lower Layers of the Atmosphere." *Quart. J. R. meteor. Soc.*, 65:537–541 (1939).
7. —— "Eddy Diffusion and Evaporation in Flow over Aerodynamically Smooth and Rough Surfaces." *Quart. J. Mech. appl. Math.*, 2:153–176 (1949).
8. —— "The Criterion of Turbulence in a Fluid of Variable Density with Particular Reference to Conditions in the Atmosphere." *Quart. J. R. meteor. Soc.*, 75:71–88 (1949).
9. Cowling, T. G., and White, A., "The Eddy Diffusivity and the Temperature of the Lower Layers of the Atmosphere." *Quart. J. R. meteor. Soc.*, 67:276–286 (1941).
10. Craig, R. A., "Measurements of Temperature and Humidity in the Lowest 1000 Feet of the Atmosphere over Massachusetts Bay." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 10, No. 1, 47 pp. (1946).
11. Davies, D. R., "Turbulence and Diffusion in the Lower Atmosphere with Particular Reference to the Lateral Effect." *Proc. roy. Soc.*, (A) 190:232–244 (1947).
12. Deacon, E. L., "Vertical Diffusion in the Lowest Layers of the Atmosphere." *Quart. J. R. meteor. Soc.*, 75:89–103 (1949).
13. Defant, A., "Die Bestimmung der Turbulenzgrössen der atmosphärischen Zirkulation aussertropischer Breiten." *S. B. Akad. Wiss. Wien.*, Abt. IIa, 130:383–403 (1921).
14. Dryden, H. L., "Recent Advances in the Mechanics of Boundary Layer Flow," in *Advances in Applied Mechanics*, R. v. Mises and T. v. Kármán, ed., pp. 2–78. New York, Academic Press, 1948.
15. —— "A Review of the Statistical Theory of Turbulence." *Quart. appl. Math.*, 1:7–42 (1943).
16. Emmons, G., "Vertical Distribution of Temperature and Humidity over the Ocean between Nantucket and New Jersey." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Inst.*, Vol. 10, No. 3, 89 pp. (1947).
17. Ertel, H., "Beweis der Wilh. Schmidtschen konjugierten Potenzformeln für Austausch und Windgeschwindigkeit in den bodennahen Luftschichten." *Meteor. Z.*, 50: 386–388 (1933).
18. —— "Thermodynamische Begründung des Richardsonschen Turbulenzkriteriums." *Meteor. Z.*, 56:109–111 (1939); "Der vertikale Turbulenz-Wärmestrom in der Atmosphäre." *Meteor. Z.*, 59:250–253 (1942); "Der Turbulenz-Wärmestrom in der Atmosphäre und das Entropieprinzip." *Meteor. Z.*, 60:246 (1943); "Die hydrothermodynamischen Grundgleichungen turbulenter Luftströmungen." *Meteor. Z.*, 60:289–295 (1943); "Über die Richtung des troposphärischen Turbulenz-Wärmestroms." *Meteor. Z.*, 61:8–11 (1944); "Zur statischen Theorie des vertikalen Turbulenz-Wärmestroms." *Meteor. Z.*, 61:207–209 (1944).
19. Flower, W. D., "An Investigation into the Variation of Lapse Rate of Temperature in the Atmosphere near the Ground at Ismailia, Egypt." *Geophys. Mem.*, No. 71 (1937).
20. Frost, R., "The Velocity Profile in the Lowest 400 Feet." *Meteor. Mag.*, 76:14–17 (1947).
21. —— "Turbulence and Diffusion in the Lower Atmosphere." *Proc. roy. Soc.*, (A) 186:20–35 (1946).
22. Goldstein, S., *Modern Developments in Fluid Dynamics*, Vol. I. Oxford, Clarendon Press, 1938.
23. —— "On the Stability of Superposed Streams of Fluids of Different Densities." *Proc. roy. Soc.*, (A) 132:524–548 (1931).
24. Haurwitz, B., "The Daily Temperature Period for a Linear Variation of the Austauschkoeffizient." *Trans. roy. Soc. Can.*, 3rd ser., III, 30:1–12 (1936).
25. Heisenberg, W., "On the Theory of Statistical and Isotropic Turbulence." *Proc. roy. Soc.*, (A) 195:402–406 (1948).
26. Holzman, B., "The Influence of Stability on Evaporation." *Ann. N. Y. Acad. Sci.*, 44:13–18 (1943).
27. Jaeger, J. C., "Diffusion in Turbulent Flow between Parallel Planes." *Quart. appl. Math.*, 3:210–217 (1945).
28. Johnson, Sir Nelson, "A Study of the Vertical Gradient of Temperature in the Atmosphere near the Ground." *Geophys. Mem.*, No. 46 (1929).
29. —— "Some Meteoroloical Observations Made at Sea." *Quart. J. R. meteor. Soc.*, 53:59–63 (1927).
30. —— and Heywood, G. S. P., "An Investigation of the Lapse-Rate of Temperature in the Lowest 100 m of the Atmosphere." *Geophys. Mem.*, No. 77 (1938).
31. Kármán, T. v., "Progress in the Statistical Theory of Turbulence." *J. mar. Res.*, 7:252–264 (1948) (Sverdrup 60th Anniversary Number).
32. —— and Howarth, L., "On the Statistical Theory of Isotropic Turbulence." *Proc. roy. Soc.*, (A) 164:192–215 (1938).
33. Köhler, H., "Meteorologische Turbulenzuntersuchungen." *K. svenska VetenskAkad. Handl.* (Stockholm), Vol. 13, No. 1 (3rd ser.). See also *Beitr. Phys. frei. Atmos.*, 19:91–104 (1932).
34. Kolmogoroff, A. N., "The Local Structure of Turbulence . . . ." *C. R. (Doklady) Acad. Sci. URSS*, 30:301–305 (1941) and 32:16–18 (1941).
35. Montgomery, R. B., "Observations of Vertical Humidity Distribution above the Ocean Surface and their Relation to Evaporation." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 4 (1940).
36. —— "Vertical Eddy Flux of Heat in the Atmosphere." *J. Meteor.*, 5:265–274 (1948).
37. Onsager, L., "The Distribution of Energy by Turbulence." *Phys. Rev.*, 68:286 (1945).
38. Paeschke, W., "Experimentelle Untersuchungen zum Rauhigkeits- und Stabilitätsproblem in der bodennahen Luftschicht." *Beitr. Phys. frei. Atmos.*, 24:163–189 (1937).
39. Pasquill, F., "Eddy Diffusion of Water Vapour and Heat near the Ground." *Proc. roy. Soc.*, (A) 198:116–140 (1949).
40. —— "Evaporation from a Plane, Free-liquid Surface in a Turbulent Air Stream." *Proc. roy. Soc.*, (A) 182:75–95 (1943).
41. —— "Some Estimates of the Amount and Diurnal Variation of Evaporation from a Clayland Pasture in Fair Spring Weather." *Quart. J. R. meteor. Soc.*, 75:249–256 (1949).
42. Penman, H. L., "Natural Evaporation from Open Water, Bare Soil and Grass." *Proc. roy. Soc.*, (A) 193:120–145 (1948).
43. Petterssen, S., and Swinbank, W. C., "On the Applica-

tion of the Richardson Criterion to Large-Scale Turbulence in the Free Atmosphere." *Quart. J. R. meteor. Soc.*, 73:335–345 (1947).

44. Prandtl, L., "The Mechanics of Viscous Fluids," in *Aerodynamic Theory*, W. F. Durand, ed., Vol. III, Division G, pp. 34–208. Berlin, J. Springer, 1935.
45. —— "Zur Frage des vertikalen Turbulenz-Wärmestrom." *Meteor. Z.*, 61:12–14 (1944), and "Nochmals der vertical Turbulenzwärmestrom." *Meteor. Z.*, 61:169–170 (1944).
46. Priestley, C. H. B., "Heat Transport and Zonal Stress between Latitudes." *Quart. J. R. meteor. Soc.*, 75:28–40 (1949).
47. —— and Swinbank, W. C., "Vertical Transport of Heat by Turbulence in the Atmosphere." *Proc. roy. Soc.*, (A) 189:543–561 (1947).
48. Ramdas, L. A., *India*, Meteor. Dept. Tech. Note No. 3.
49. Reynolds, O., "An Experimental Investigation of the Circumstances Which Determine Whether the Motion of Water Shall Be Direct or Sinuous and of the Law of Resistance in Parallel Channels," in *Papers on Mechanical and Physical Subjects*, Vol. II, pp. 51–105. Cambridge, University Press, 1901. (Reprinted from *Phil. Trans. roy. Soc. London*, 174:935–982 (1883).)
50. Richardson, L. F., "Atmospheric Diffusion Shewn on a Distance-Neighbour Graph." *Proc. roy. Soc.*, (A) 110:709–737 (1926).
51. —— "Turbulence and the Vertical Temperature Difference near Trees." *Phil. Mag.*, 49:81–90 (1925).
52. —— and Proctor, D., "Diffusion over Distances Ranging from 3 to 86 km." *Mem. R. meteor. Soc.*, 1:1–16 (1925).
53. Richardson, L. F., and Stommel, H., "Note on Eddy Diffusion in the Sea." *J. Meteor.*, 5:238–240 (1948).
54. Roberts, O. F. T., "The Theoretical Scattering of Smoke in a Turbulent Atmosphere." *Proc. roy. Soc.*, (A) 104:640–654 (1923).
55. Rossby, C.-G., and Montgomery, R. B., "The Layer of Frictional Influence in Wind and Ocean Currents." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 3, No. 3, 101 pp. (1935).
56. Schlichting, H., "Experimentelle Untersuchungen zum Rauhigkeitsproblem." *Ingen.-Archiv.*, 7:1–34 (1936).
57. —— "Turbulenz bei Wärmeschichtung." *Z. angew. Math. Mech.*, 15:313–338 (1935).
58. Schmidt, W., "Der Massenaustausch in freier Luft und verwandte Erscheinungen." (*Probl. kosm. Phys.*, Bd. 7.) Hamburg, H. Grand, 1925.
59. Scrase, F. J., "Some Characteristics of Eddy Motion in the Atmosphere." *Geophys. Mem.*, No. 52 (1930).
60. Sheppard, P. A., "The Structure and Refractive Index of the Lower Atmosphere," in *Meteorological Factors in Radio Wave Propagation*, pp. 37–79. Phys. Soc., London, 1947. (Report of a conference held at the Royal Institution, London, 8 April 1946.)
61. —— "The Aerodynamic Drag of the Earth's Surface and the Value of von Kármán's Constant in the Lower Atmosphere." *Proc. roy. Soc.*, (A) 188:208–222 (1947).
62. Sutcliffe, R., "Surface Resistance in Atmospheric Flow." *Quart. J. R. meteor. Soc.*, 62:3–14 (1936).
63. Sutton, O. G., *Atmospheric Turbulence*. London, Methuen, 1949.
64. —— "Convection in the Atmosphere near the Ground." *Quart. J. R. meteor. Soc.*, 74:13–30 (1948).
65. —— "The Application to Micrometeorology of the Theory of Turbulent Flow over Rough Surfaces." *Quart. J. R. meteor. Soc.*, 75:335–350 (1949).
66. —— "A Theory of Eddy Diffusion in the Atmosphere." *Proc. roy. Soc.*, (A) 135:143–165 (1932).
67. —— "The Problem of Diffusion in the Lower Atmosphere." *Quart. J. R. meteor. Soc.*, 73:257–276 (1947).
68. —— "Wind Structure and Evaporation in a Turbulent Atmosphere." *Proc. roy. Soc.*, (A) 146:701–722 (1934).
69. —— "The Theoretical Distribution of Airborne Pollution from Factory Chimneys." *Quart. J. R. meteor. Soc.*, 73:426–436 (1947).
70. Sutton, W. G. L., "On the Equation of Diffusion in a Turbulent Medium." *Proc. roy. Soc.*, (A) 182:48–75 (1943).
71. Sverdrup, H. U., "The Humidity Gradient over the Sea Surface." *J. Meteor.*, 3:1–8 (1946).
72. Taylor, Sir Geoffrey, "The Transport of Vorticity and Heat through Fluids in Turbulent Motion." *Proc. roy. Soc.*, (A) 135:685–702 (1932).
73. —— "Diffusion by Continuous Movements." *Proc. Lond. math. Soc.*, (2) 20:196–212 (1920).
74. —— "Statistical Theory of Turbulence." *Proc. roy. Soc.*, (A) 151:421–478 (1935), and *Proc. roy. Soc.*, (A) 156:307–317 (1936).
75. —— "Skin-Friction of the Wind on the Earth's Surface." *Proc. roy. Soc.*, (A) 92:196–199 (1916).
76. —— "Effect of Variation of Density on the Stability of Superposed Streams of Fluid." *Proc. roy. Soc.*, (A) 132:499–523 (1931).
77. Thornthwaite, C. W., and Holzman, B., "Measurement of Evaporation from Land and Water Surfaces." *U. S. Dept. Agric. Tech. Bull.* No. 817 (1942).
78. Thornthwaite, C. W., and Kaser, P., "Wind-Gradient Observations." *Trans. Amer. geophys. Un.*, Pt. I, pp. 166–182, 24th Annual Meeting, April 23–25, 1943.
79. Weizsäcker, C. F. v., "Das Spektrum der Turbulenz bei grossen Reynoldschen Zahlen." *Z. Phys.*, 124:614–627 (1948).
80. Wüst, G., "Temperatur und Dampfdruckgefälle in den untersten Metern über der Meeresoberfläche." *Meteor. Z.*, 54:4–9 (1937).

ATMOSPHERIC TIDES AND OSCILLATIONS

By SYDNEY CHAPMAN

Queen's College, Oxford, England

The oscillations considered in this article are those of world-wide character, on a scale greater than that of the ordinary local or regional weather distributions. It is convenient to call them "tides" even when their origin is thermal and not gravitational.

Most of our information concerning these tides comes from the barometer, which shows an excess or deficit of pressure as the motion heaps up the air or draws it away above a locality.

GRAVITATIONAL TIDES

The sea tides, with their twice-daily rise and fall, have been known since the dawn of history. The time of high tide advances from day to day in evident association with the moon's moiton. Our understanding of this goes back to Newton [85*a*], who showed how, on the basis of his laws of mechanics, a universal inverse-square "gravitational" attraction between all material particles would explain the weight that makes bodies near the earth fall towards it, explain the planetary motions, and also, less simply, the sea tides.

The gravitational attraction exerted by the sun on a particle of the earth is GS/r^2 per unit mass, where G denotes the gravitational constant, S the sun's mass, and r the distance from the sun's centre O to the terrestrial particle. On the average this force provides the orbital centripetal acceleration $\omega^2 r$ towards the sun, where ω denotes the orbital angular velocity of the earth (2π per year). The average values of these accelerations are those at the earth's centre C, for which $r = r_0$, where $r_0 = OC$. Hence $GS/r_0^2 = \omega^2 r_0$ or $\omega^2 = GS/r_0^3$.

But the gravitational attraction is greater, and the centripetal acceleration less, over the hemisphere nearer the sun, than at the earth's centre, leaving a distribution of unbalanced force over this hemisphere, directed sunwards (at most places obliquely upwards, partly acting against the earth's pull). Over the opposite hemisphere the gravitational acceleration is too weak to supply the whole centripetal acceleration, and there is a distribution of unbalanced acceleration away from the sun, here also obliquely upward, partly against the earth's pull. Particles free to move, like those of the sea water or air, will do so under the action of these unbalanced forces; the water surface will tend to become spheroidal, with its long axis along the line OC joining the centres of the sun and earth.

If the earth did not rotate relative to this line, such a steady distribution of level would actually be attained. It is called the *equilibrium solar tide*. This is proportional to the gradient along OC, at C, of the unbalanced acceleration $\omega^2 r - GS/r^2$, that is, to $\omega^2 + 2GS/r_0^3$ or $3GS/r_0^3$.

The moon exerts a similar tidal influence, proportional to $3\ GM/r_1^3$, where M and r_1 denote the moon's mass and distance (from C); M/S and r_1/r_0 are both very small, but it turns out that the lunar tidal action is 2.4 times as great as the solar, and on this account the moon chiefly governs the sea tides, though the sun at different epochs in the month adds to or reduces the lunar tides.

In the eighteenth century Newton's successors extended his planetary theory to explain the motions in the solar system in almost every detail. They began also to develop the theory of the tides. Laplace [74*a*] recognized the powerful influence of the earth's rotation on the tides, which are a dynamical, not statical, phenomenon. The "equilibrium tide" is only a theoretical conception providing a convenient standard of comparison for the real tides.

Laplace developed his tidal theory in a series of remarkable memoirs, later summarized in his *Mécanique céleste*. He restricted his analysis for the most part to the idealized case of an ocean of uniform depth covering the whole earth. Hence his results have only a limited application to the actual sea tides, which are complicated by the irregular outlines and interconnections of the oceans, and by their nonuniform depth. These complications present problems with which tidal theorists still wrestle arduously.

Laplace showed that his ideal ocean tides might be greater or less than the equilibrium tides, according to the depth of the ocean; and also that they might not be "direct," with high tide "under" the tide-producing body, and also on the other hemisphere: they might be "reversed," with low tides at points on the earth nearest to and farthest from the tide-producing body, and high tides intermediate between them.

THE ATMOSPHERIC TIDES S_n, L_n

Newton realized that the tidal forces must act on the atmosphere as well as on the seas. The lack of lateral boundaries renders the atmosphere a better subject for Laplace's idealized theory. This theory treated the sea water as incompressible, but he showed that the theory could be adapted surprisingly well to the compressible air if its temperature were assumed uniform, and unchanged throughout the compressions and rarefactions accompanying the air tides—that is, if the changes of air density were supposed to take place isothermally. On this basis he calculated the air tides for an atmosphere agreeing in total mass and average temperature with the actual atmosphere, and showed that the tides would be *direct*, that is, the air would be heaped up with high pressure under the tide-producing body, and on the opposite side of the earth.

Thus the tide should produce a twice-daily variation

of atmospheric pressure as recorded by the barometer, invented by Torricelli at about the time of Newton's birth. Already in the seventeenth century [21] the value of this instrument in the study of weather began to be realized, and in Newton's lifetime it became known that the barometric changes in the tropics are quite different from those in temperate latitudes. Instead of the mercurial height varying through several centimetres, as it does in temperate latitudes because of the irregular weather changes, the tropical barometer is almost steady except when hurricanes sweep the region. It shows, however, a small rise and fall twice daily, with a range of about 2 mm, almost as simple and regular as the tidal changes of sea level; but unlike these, the times of high pressure show no perceptible change throughout the month; they recur daily at nearly the same local *solar* time, about 10:00 A.M. and 10:00 P.M. This is illustrated in Fig. 1, which shows the barometric changes at Batavia and Potsdam (on different scales) for the same few days [14].

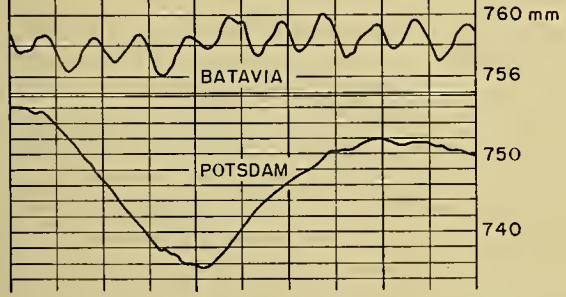


FIG. 1.—Barometric variations (on different scales) at Batavia and Potsdam, November 5–9, 1927.

This regular twice-daily barometric change is clearly due to the sun, so the "tidal" régime of the atmosphere differs greatly from that of the oceans, where the moon is the controlling agent. The moon's effect on the barometer is very small, though it can be determined with much labor from sufficient data.

The solar semidiurnal barometric variation is our chief indication of the greatest world-wide oscillation of the atmosphere. For brevity, this oscillation, with its associated barometric and wind changes, will be denoted by the symbol S_2, in which S signifies *solar*, and the suffix 2 indicates the number of its cycles or swings per day. Similarly, S_n will denote an oscillation whose barometric and other changes at each place are repeated n times each mean solar day, and L_n one with n repetitions in each mean *lunar* day.

The moon's capacity to exert any appreciable dynamical influence on the atmosphere lies solely in its tidal force, so potent on the oceans. The harmonic development of the lunar tidal potential [58] clearly indicates the values of n to be expected, and their relative importance in the lunar tidal force; for the chief term n equals 2, as in the sea tides, and there are two other terms for which n differs only slightly from 2; these increase or decrease the main semidiurnal tide according as the moon is nearer to or farther from the earth. There is also a term for which n is nearly equal to unity, but this depends on the moon's declination (namely, its angular "distance" from the equatorial plane) and is reversed in sign twice monthly as the moon moves northward or southward across the equator. This oscillation L_1 has not been detected in the barometric records [30].

THE SEARCH FOR THE LUNAR ATMOSPHERIC TIDE L_2

The Tide at Paris. Laplace's theory [74*a*] indicated a "direct" lunar tidal variation of the barometer with a range in the tropics of half a millimetre. He considered that such a change, though small, should be determinable from a considerable series of tropical readings; but no such series was available to him in 1823, when he developed an active interest in the air tide, an interest which he maintained for the rest of his life. Instead, he was able to obtain from Bouvard, of the Paris observatory,[1] an eight-year series (October 1, 1815 to October 1, 1823) of barometric readings made daily at 9, 12, 15, and 21 hours. He used only part of the three daytime readings, 4752 in all, and from these, by a well-designed method, he sought to compute the lunar semidiurnal tide L_2. He calculated its range as 0.054 mm, and the lunar times of maximum pressure as 3^h19^m after upper or lower transit. He attached limited significance to this result, after calculating the probability [74*c*] that it was not merely due to chance, that is, to the continual irregular "weather" variations of the barometer. He asserted that to determine so small a variation from such data with adequate certainty would require at least 40,000 observations.

Bouvard [74*f*] in 1827 repeated Laplace's calculation with four more years' data (in all, twelve years, January 1, 1815 to January 1, 1827), and from 8940 readings found a lunar tidal range of 0.0176 mm, with maxima at 2^h8^m and 14^h8^m. The great difference from Laplace's result confirmed the inference that the data used were still far too few.

In 1843 Eisenlohr [60] renewed the attempt with twenty-two years' data (1819–1840), using all the four daily readings. Unfortunately he departed from Laplace's excellent method of computation, which involved only *differences* between readings on the same day, thus eliminating the influence of the large changes of pressure from day to day. Eisenlohr rearranged his data according to the nearest lunar hour (0 to 23) at the time of each reading; with unlimited data this method would be satisfactory, showing the complete average change of the barometer according to lunar time. But with his limited data, the number of readings per lunar hour ranged from 1302 to 1377, and the hourly means

1. In references [28, 29] the source of Laplace's data for the air tide was wrongly given as Brest, through confusion with his use of Brest sea-level data for comparison with his theory of sea tides. It may also be added that in the translation of (Vols. 1–4 only, of) Laplace's great work, by N. Bowditch, ref. [74*a*] is to be found on pp. 793–801 of Vol. 2 (Boston, 1832).

were unequally affected by the great weather variations of pressure. They showed a quite irregular variation from hour to hour. Thus his laborious effort, in many ways well planned, was fruitless. He concluded that his data were insufficient to determine L_2, and hence that the attempts by Laplace and Bouvard were still less adequate. He urged that *hourly* readings of the barometer should be taken, so that in time, from a long series, L_2 might be determined.[2]

The Tropical Lunar Air Tide. When Eisenlohr wrote (1843), L_2 had already been determined from tropical pressure data, though the result was not published until 1847 [91]. Around 1840, several British colonial observatories, magnetic and meteorological, were set up under Sabine's leadership. In 1842 the director (Lefroy) of the St. Helena observatory successfully used his seventeen months' (August 1840 to December 1841) bihourly week-day barometric readings to determine L_2. His successor Smythe, with Sabine, confirmed the determination from three more years' data, October 1842 to September 1845. Sabine even attempted to find how the air tide varies with the lunar distance [91].

In 1852 the director (Elliot) of the Singapore observatory determined L_2 there [61] from five years' data, 1841 to 1845.

Later, when the Batavia observatory was established in 1866, these results stimulated its directors [22, 23, 96–99] to determine L_2 from their hourly barometric data, recorded photographically. Bergsma published the first results, for 1866 to 1868, in 1871. By 1905, L_2 at Batavia was really well determined from 350,000 observations covering forty years.

The Lunar Air Tide Outside the Tropics. Laplace [74*a*, *c*] stressed the need, in deriving results from observations, to determine the probability that the error lies within narrow known limits; without so doing, he said, one risks presenting the effects of irregular causes as laws of nature, "as has often happened in meteorology." This need has been overlooked or neglected by a multitude of those who, before and since his time, have vainly sought for lunar *monthly* meteorological variations. Eisenlohr, from 1833, was among these; but only a few of them have, like him, engaged in the more hopeful but still perilous search for lunar *daily* meteorological variations, in particular for L_2. Of these few, some, like Kreil in 1841, or later Bouquet de la Grye, used quite inadequate data[3]—for one year only, or even for five years, like Neumayer [84], who in 1867 failed to obtain consistent results from five years' hourly data (1858 to 1863) for Melbourne (lat 38°S). Even Airy [1], who used 180,000 hourly values for Greenwich (for twenty years, 1854 to 1873) unwisely concluded in 1877 that "we can assert positively that there is no trace of lunar tide in the atmosphere." Börnstein [24][4] in 1891 used only four years' data for Berlin and Vienna and Hamburg; for Keitum (lat 55°) he used ten years' hourly data (1878 to 1887), but found "no trace of a semidiurnal variation such as a lunar tide would produce"; he thought, however, he had found a definite lunar *diurnal* variation (not, like L_1, reversed fortnightly; see p. 511). Bartels [12] in 1927 showed that these conclusions completely misinterpreted the actual results that Börnstein had obtained, and that his supposed lunar diurnal variation was a purely chance effect, whereas L_2 was contained in his curves, though it was ill-determined. The misinterpretation sprang directly from the neglect of Laplace's advice to consider the probable accuracy of the results.

Morano [83] in 1899 used four years' data (1891 to 1894) for Rome (lat 42°N), although Neumayer [84] had failed, by the same method applied to five years' data for Melbourne, in a rather lower latitude, to obtain a reliable result. The validity of Morano's result, which is probably near the true value of L_2 at Rome, remained uncertain.

In 1918 a new attempt [28] succeeded in determining L_2 from the Greenwich hourly data, by then available for sixty-four years. Two-thirds of the material was rejected; the only days used were those on which the barometric range did not exceed 0.1 in. This was the first certainly valid nontropical determination of L_2.

This investigation was planned and undertaken with guidance [27], in a very simple way, from the theory of random errors. According to this theory, if a single observation is subject to an accidental error e, the probable random error of the sum of N such (independent) observations is $e\sqrt{N}$, and that of their mean is $e/\sqrt{N}$. The moon produces a systematic (though very small) semidiurnal variation of the barometer; if this is combined from N lunar days' observations (each in the form of a sequence of lunar hourly values), by forming the sum of the values for each lunar hour, the sequence of sums will contain N times the lunar daily variation. It will, however, be affected also by the other causes of variation, particularly, outside the tropics, by the succession of cyclones and anticyclones. If these produce an average random departure e of any hourly value from the long-term barometric mean, they will contribute to each lunar hourly sum of N hourly values a random contribution of the order $e\sqrt{N}$. As N is increased, the regular lunar daily variation in the lunar hourly sequence of sums will increase proportionately to N, and the random contributions will increase, but proportionately only to $\sqrt{N}$. Thus, although e greatly exceeds the range of the lunar air tide at Greenwich, the systematic tidal effect will altogether overpower the random contribution if N is taken large enough. In the sequence of lunar hourly *means*, L_2 is independent of N, whereas the random errors are of the order $e/\sqrt{N}$.

As the Greenwich data were used only for days of barometric range 0.1 in. or less, the average random departures e from each day's *mean* might be estimated as 0.01 in. Since N was 6457, $e/\sqrt{N}$ would be about 0.00012 in.

2. Not until 1945 was a further and successful attempt made to determine L_2 at Paris. The results are not yet published.

3. The method of computation is also important, as well as the amount of data, for the success of a determination.

4. See also [12, pp. 39, 40].

Figure 2 (full line) shows the mean lunar daily variation of Greenwich pressure obtained [28] from these N days, by a method of rearrangement of solar hourly values according to lunar time. Happily, this method avoided a pitfall, then unsuspected but afterwards disclosed by Bartels [12], associated with the use of selected barometrically "quiet" days [41].

The total range of pressure in Fig. 2 is less than 0.001 in., and the change from one lunar hour to the next averages about 0.00015 in. This exceeds the average random error in Fig. 2, namely, 0.00010 in., and the *systematic* nature of the lunar daily variation is clearly manifest. Apart from its meteorological and dynamical interest, this determination has great *statistical* interest as a remarkable illustration of the "law of combination of random errors"—an example confirmed by many later air-tide determinations, most notably by that of the tidal variation of air temperature at Batavia [36]. (See p. 519.)

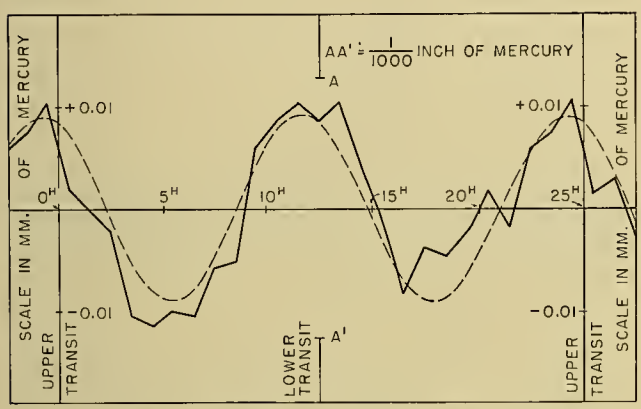


FIG. 2.—The average lunar daily variation (full line) of barometric pressure at Greenwich, computed from 6457 days' hourly data, 1854–1917; the broken line shows the lunar semidiurnal component of the variation.

Harmonic Analysis and the Harmonic Dial: Units. In Fig. 2 the broken curve represents the lunar semidiurnal harmonic (or Fourier) component obtained by harmonic analysis of the calculated variation (full line). This component curve represents the type of variation due to the moon, and the difference between the two curves must be ascribed to random variation due to weather.

Any variation which is periodic in an interval T can be harmonically analyzed and represented as a sum of harmonic terms,

$$\Sigma\, c_n \sin(n\theta + \gamma_n), \tag{1}$$

where $n = 1, 2, \ldots$ and θ denotes time reckoned in angle at the rate 360° per interval T. The integer n is the *order* of the harmonic, c_n and γ_n are its *amplitude* and *phase*. This harmonic has maxima at the time $\theta = (90° - \gamma_n)/n$ and at intervals T/n thereafter throughout the period T.

In the case of a solar or lunar daily variation, T signifies a (mean) solar or lunar day, and θ signifies solar time t at the rate 15° per mean solar hour, or lunar time τ at the rate 15° per mean lunar hour. It is convenient to reckon t from local mean midnight, and τ from local lower mean lunar transit. It is also convenient to denote the nth harmonic (S_n or L_n) by the distinctive notations:

$$s_n \sin(nt + \sigma_n) \tag{2}$$

for S_n, and for L_n,

$$l_n \sin(n\tau + \lambda_n). \tag{3}$$

The only harmonics yet detected in the lunar air tide (see p. 511) are the second ($n = 2$) and the two for which n differs only slightly from 2; the other harmonics obtainable by analysis of the full-line curve in Fig. 2 represent only residual accidental error.

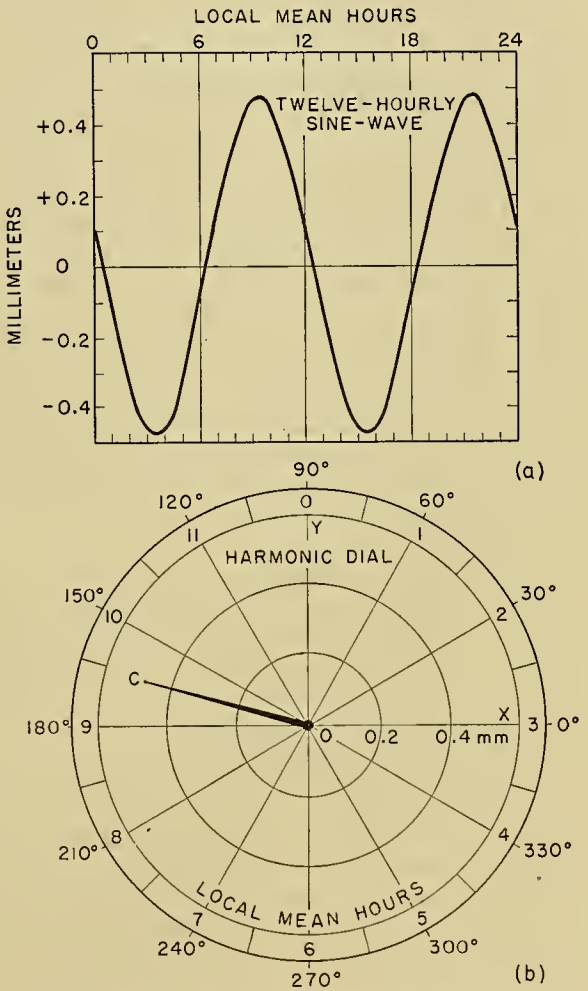


FIG. 3.—The solar semidiurnal component (S_2) of the daily barometric variation at Washington, D. C., *represented* (Fig. 3*a*, above) by a graph, *specified* (Fig. 3*b*, below) by a harmonic dial.

In this article l_2, for the lunar tidal variation of *barometric pressure*, will be expressed either in micrometres (μm) of mercury (1 μm = 10^{-6}m = 0.001 mm) or in microbars (1 microbar = 0.001 mb or 1 dyne cm^{-2}, or 0.00075 mm of mercury). The unit of *speed* used for l_2, the amplitude of the lunar tidal wind variation associated with L_2, will be 1 cm sec^{-1} (= 0.036 km hr^{-1}). In graphical illustrations showing both S_2 and L_2, the

scale values for the pressure and speed variations associated with S_2 will be ten times greater than for L_2.

A harmonic variation such as the nth term of (1) can be graphically represented by a curve, for example, the broken line for L_2 in Fig. 2 or the curve for S_2 in Fig. 3*a*;[5] or it may be specified by a diagram which indicates the amplitude c_n and phase γ_n, as in Fig. 3*b* [16]. This shows an origin O, a phase-reference line OX, and a line OC whose length, on the amplitude scale marked along OX, represents c_n (in this case s_2), and whose direction indicates, by the angle XOC, reckoned anticlockwise, the phase γ_n (in this case σ_2). A circle may be drawn with O as centre, and graduated, as in Fig. 3*b*, to show the phase angle.

A perpendicular axis OY may also be added, and the circle may also be graduated (clockwise from OY) in *time* measure $\varphi = n\theta$, at the rate 360° per interval T/n. For Fig. 3*b*, $n = 2$, so that T/n is half a solar day or twelve solar hours, and the time measure is therefore 30° per solar hour (or in a similar diagram for L_2, 30° per mean lunar hour, reckoned as 24 per lunar day). On this graduation, OC points to the time given by $n\theta = 90° - \gamma_n$, at which the harmonic variation has its first maximum in the interval T. When $n = 2$, the diagram corresponds to an ordinary clock face, on which each hour corresponds to 30°; OC points to the times of maxima, morning and afternoon. When $n = 1$, the diagram may be likened to a 24-hour dial. When $n = 3$ or $n = 4$, the diagram will show only eight or six hours on its face or rim, at intervals of 45° or 60° respectively. Because of this alternative interpretation of the diagram, as a clock face indicating the amplitude and time(s) of maximum, the diagram is appropriately called a *harmonic* dial [30].

Any part of a harmonic dial not needed in a particular case may be omitted; so also can the line OC if its end C is shown. This is useful when, as in Fig. 4, several determinations of a harmonic are to be shown on one dial. Figure 4 shows forty dial points [12], each representing the determination of L_2 in Batavia pressure for one of the forty years 1866 to 1905.

The Probable Error. More than a century after Laplace's pioneer attempt to assess the reliability of his determination of L_2 [74*c*, *d*], Bartels [11, 12] in 1926 applied the theory of errors in a plane to assess the uncertainty of the mean of a number of independent determinations of a periodic variation such as L_2, conveniently represented by dial points as in Fig. 4. He showed how to determine the *probable error ellipse* centred on the dial point C for the mean determination. When, as in Fig. 4, the individual dial points are symmetrically distributed around C, the ellipse becomes the *probable error circle*, as there shown. Its radius r is calculated, according to the theory of errors, from the distances d of the individual dial points from C. When, as in Fig. 4, the individual dial points are of equal weight, $r = 0.939\bar{d}$, where $\bar{d}$ denotes the mean value of d.

The probability that any individual dial point will fall within a distance d from C is $1 - e^{-(x/r)^2}$, where $x = 0.833d$. When $d = r$, this probability is ½, signifying that half the points are likely to fall within the probable error circle. This chance is reasonably well exemplified in Fig. 4, where nineteen of the forty points lie within the circle.

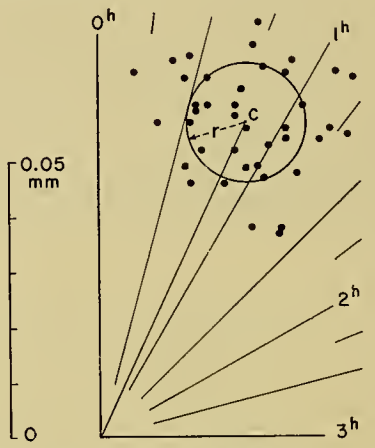


FIG. 4.—Harmonic dial showing the forty dial points for determinations of L_2, the lunar semidiurnal air tide in Batavia barometric pressure, for each of the years 1866–1905. Their centroid is the point C, specifying the forty-year mean of L_2. The circle with C as centre is the probable-error circle for any of the forty yearly dial points.

The circle indicates the probable error r of each of the forty yearly determinations of L_2; that of the 40-year mean is $r/\sqrt{40}$, as might be indicated on a separate dial showing only C and its own probable error circle.

A determination may be considered reasonably good if the length of its dial vector is at least three times its probable error. For L_2 at Batavia, each yearly determination satisfies this criterion, as there r is 0.011 mm, and l_2 ranges from about 0.045 to 0.078; for the 40-year mean value, $l_2 = 0.062$ and $r = 0.0016$, so that $l_2/r = 39$.

The uncertainty in the *amplitude* of a harmonic variation may be regarded as corresponding, in the sense explained above, to its r, and that of the *phase* to the angle $\sin^{-1}(r/l)$ subtended at O by half the probable error circle. Thus the phase uncertainty will be greater, for a given value of r, the smaller the amplitude l; for the "yearly" probable error circle shown in Fig. 4, the phase uncertainty is ±10°, or ±20 minutes of (lunar) time; for the 40-year mean it is only three minutes of time.

Methods of Computation of L_2. Several different methods have been used to determine L_2 from barometric records at various stations. In a few cases lunar hourly readings were taken, or measured from barographs, or interpolated between the solar hourly measures. This is quite unnecessary; the solar hourly values fully suffice if properly used. Where only a few readings

5. For the solar semidiurnal component variation of barometric pressure at Washington, D. C.; Fig. 3*a* is the graph of $s_2 \sin(2t + \sigma_2)$.

per day are available, a method depending on the differences between them, thus eliminating their absolute values, is desirable. Laplace [74], and in recent years also Bartels [17], used such a method. Almost all determinations other than those on the early Paris data have been based on hourly values, or (better, as Bartels urged [17]) on bihourly values, which give almost equal accuracy with much less labour [53].

In most of the determinations up to about 1935 the lunar day was taken as the basis, and the observations were rearranged according to lunar time (actual, not mean). Usually, though not always, the solar daily variation was removed before or after the retabulation. It is very important to remove, or allow for, any non-periodic change of pressure in the course of each day. This was only gradually realized [29, 32, 51, 54].

In 1917 reasonably certain determinations of L_2 had been made at only three stations [22, 23, 61, 91, 96–99]; it is now known at over sixty-five stations [12, 17, 28–46, 48–57, 88, 89]. More than half of these determinations have been made by the method [54] that now seems most suitable where hourly or bihourly data are available. The method is based on *solar* daily sequences of twenty-five hourly or (better) thirteen bihourly values, the last value for each day, which is also the first for the next day, being added so that the aperiodic change in the day can be removed. The daily sequences are separated in twelve lunar-phase groups, using tables [19, 20] constructed by Bartels and Fanselau for this purpose. The grouping is *facilitated* if each daily sequence of twenty-five or thirteen (three-figure) values, with its identifying and lunar-classification data, is entered on a punched card, using Hollerith sorting and adding machines in the later work; but these devices are not *necessary* for the application of the method. A practical description of the method, with examples, including the probable-error computation, is available [103]. The solar daily component variations S_n are computed in the course of the work, and are thus available for comparison with L_2 from the same material.

The determination of these variations, and especially of L_2, from the records of air pressure, wind, and temperature, may be likened to the extraction of a rare constituent from a great mass of crude ore—or of a needle from a haystack! It is an example of the unexhausted value of the long series of meteorological (as also of magnetic) data garnered at many observatories decade after decade. There is great need and scope for such work on many series of data not yet dealt with.

THE LUNAR ATMOSPHERIC TIDE L_2

Geographical Distribution: Annual Mean. Figure 5 indicates the annual mean value of L_2 so far as results are now available [47, 57]. The world map shows arrows which in direction and length (on the scales given) represent the dial vector for L_2 at the point corresponding to the *centre* of the arrow shaft. All the stations lie between 60°N and 40°S; there are large areas in this belt, however, where L_2 is not known—notably Central Asia, the western Pacific, and parts of South America. Some determinations made for Russian stations may have been omitted because the literature is not accessible. It is to be hoped that these gaps will gradually be filled, and the limits of our knowledge pushed polewards, using the improved computing methods [54, 103] mentioned in the foregoing section.

If the lunar air tide were in phase with the tidal force, all the arrows in Fig. 5 would be upright; actually most of them point to the right, implying a lag of high atmospheric tide (by about half an hour on the average) after lunar transit; but some well-determined arrows point leftward, as at Mauritius and Kimberley, and at the five north European stations, where high tide definitely precedes the lunar transit. Later diagrams (*e.g.*, Figs. 8, 10) indicate the uncertainties of amplitude and phase at several stations.

The L_2 component of lunar tidal *force* decreases steadily polewards from the equator, and on the whole so do the arrow-lengths representing l_2 in Fig. 5; but there are several departures from this regularity, definitely not due to errors in the determinations. The abnormalities in the distribution of L_2 are illustrated (rather tentatively, so far as the data will allow) in Fig. 6, which shows lines of equal amplitude l_2; where the lines are ill-determined they are drawn "broken." There is a belt of specially *high* tidal amplitude across the South Indian Ocean; and along the west coast of North and South America l_2 is abnormally low. Also at Buenos Aires, on the east coast, l_2 is much less than at Melbourne, in nearly the same latitude.

The remarkable anomaly of the small lunar air tide near the northwestern American coast is illustrated in Fig. 7 [46, 47, 57], which gives the distribution of L_2 over North America as in Fig. 5, but on a larger scale and with additional details (*cf.* the next section).

The Annual Variation of L_2. The lunar tidal *force* undergoes no regular *annual* variation, though it includes a semimonthly change of L_2 inseparable [30] from a semiannual change of S_2. Nevertheless, L_2, unlike the sea tides, varies notably in the course of each year. This must be due to a large-scale annual change in our atmosphere, the system on which the lunar tidal force acts.

World meteorological charts of isotherms and isobars show marked seasonal changes of distribution—seasonal in the sense that on the whole, as the sun crosses the equator northward or southward, these changes alternate, the summer state of one hemisphere being approximately reproduced in the other hemisphere in its summer. The lunar air tide is not seasonal in this sense; its main changes of l_2 and λ_2 occur simultaneously in both hemispheres.

Figure 8 (*a*, *b*) illustrates this in one way, 8*a* (above) showing l_2, and 8*b* (below) showing λ_2, for the *D* (December solstitial) group of four months November to February. The black dots indicate l_2 and λ_2 for about fifty stations, in the latitudes to be read on the scales of the abscissas; the lines centred on these dots indicate the uncertainty of l_2 and λ_2, in the manner described in the foregoing sections. The curved lines indicate the

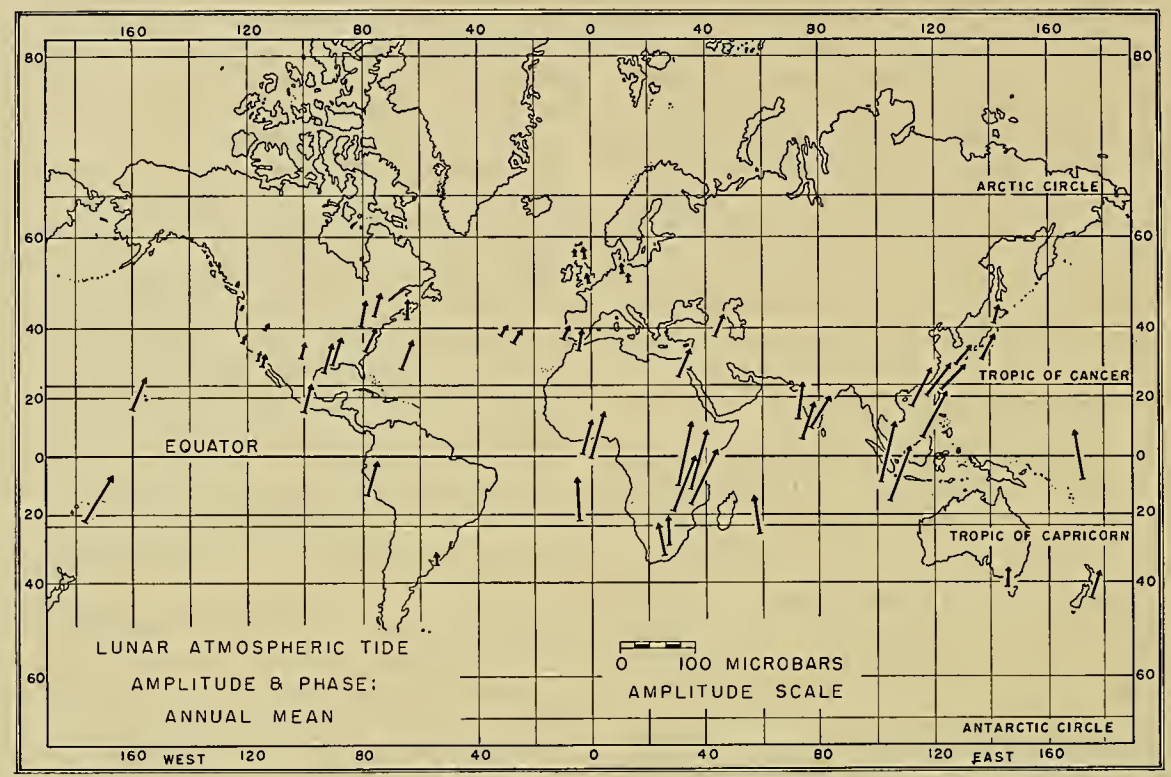


FIG. 5.—Geographical distribution of the annual mean lunar semidiurnal air tide L_2 in barometric pressure, indicated by dial vectors, each referring to the place at the mid-point of the arrow, whose length gives the amplitude l_2 on the scale shown, and whose direction gives the time of high air tide, as shown on a local mean lunar clock.

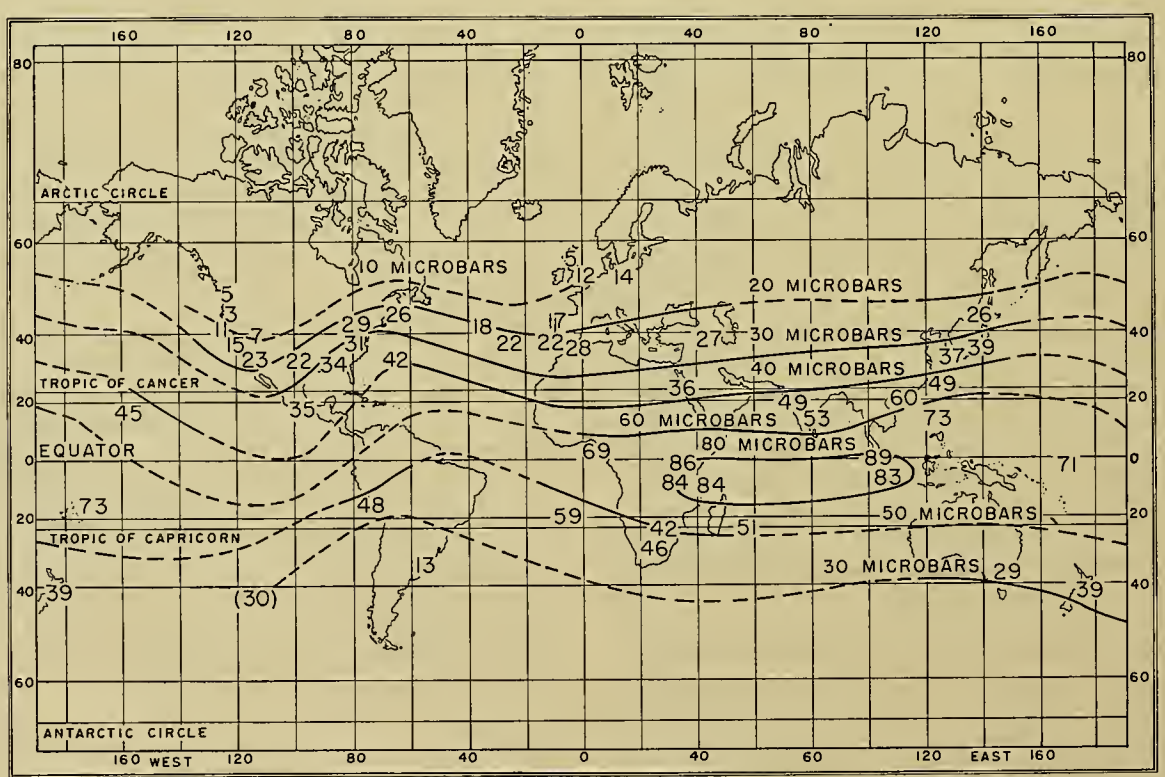


FIG. 6—Tentative lines (broken where only weakly determined) of equal amplitude l_2 (in microbars) of the annual mean lunar semidiurnal air tide L_2 in barometric pressure. The numbers (other than those, 10, 20, . . . , 80, followed by the word "microbar," which show the value of l_2 for each line) show local values of l_2.

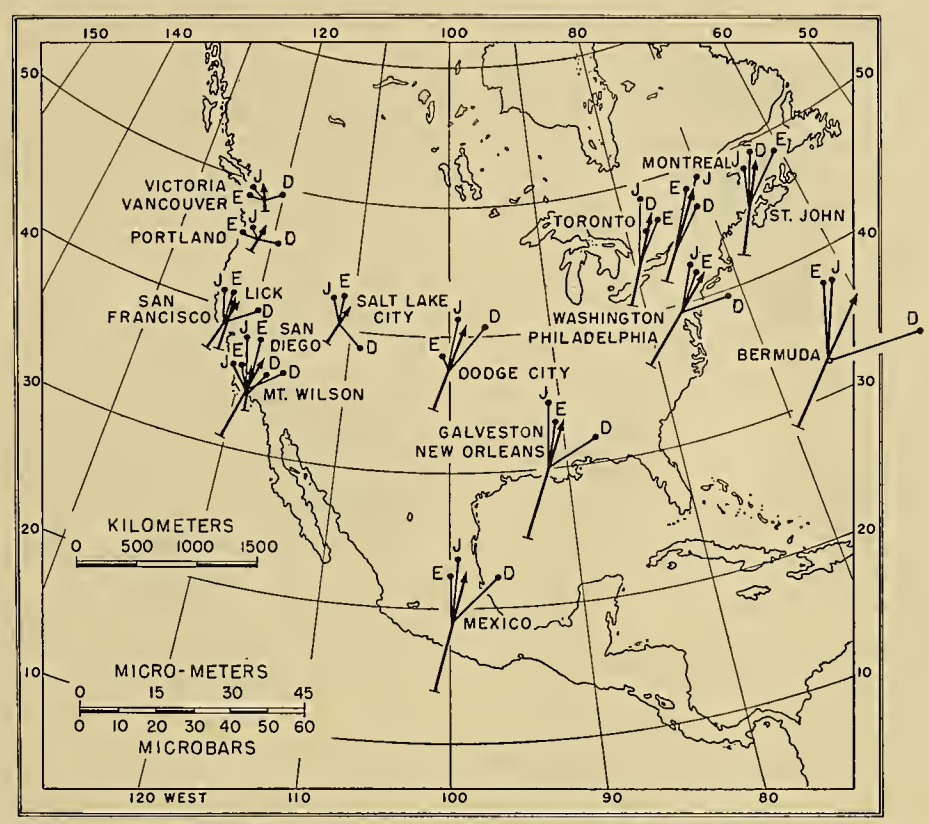


FIG. 7.—The distribution of L_2, the lunar semidiurnal air tide in barometric pressure, over North America, as shown by arrows, each of which specifies the annual mean L_2 at the point at its centre. The lines diverging from the centres of the arrows are the forward halves of corresponding arrows specifying the mean L_2 for groups of four months, J (May to August), D (November to February), and E (March, April, September, October).

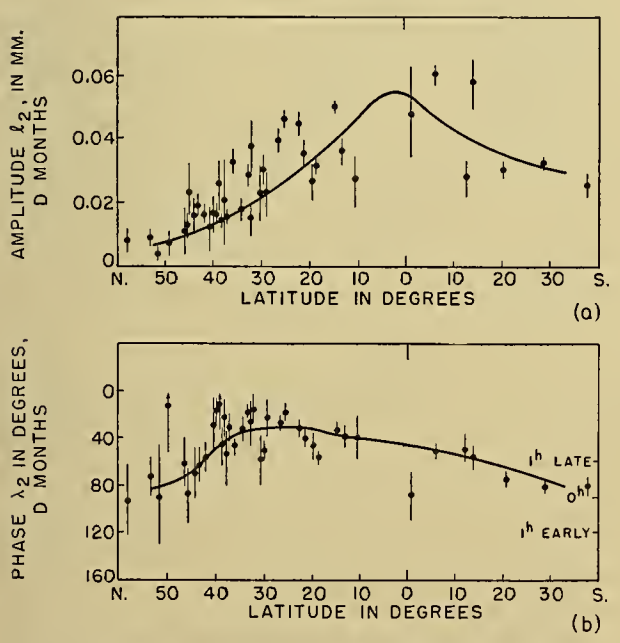


FIG. 8.—The dots indicate the latitude and the values of the amplitude l_2 (Fig. 8*a*, above) and phase λ_2 (Fig. 8*b*, below) of the lunar semidiurnal air tide in barometric pressure in the mean of the D months (November to February), at various stations. The vertical lines indicate the probable range of uncertainty of the determinations. The curves are drawn through points giving the mean l_2 or λ_2 for gronps of stations within narrow ranges of latitude.

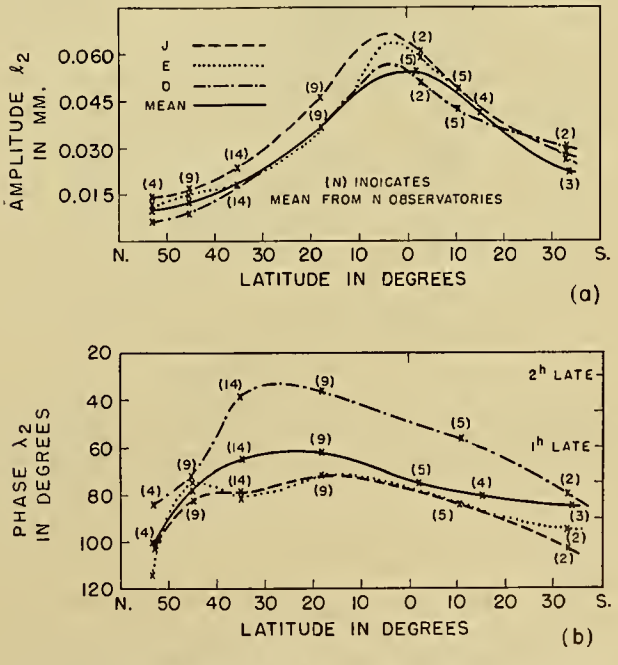


FIG. 9.—Curves showing the average dependence on latitude of the amplitude l_2 (Fig. 9*a*, above) and phase λ_2 (Fig. 9*b*, below) of the lunar semidiurnal air tide in barometric pressure, for the mean of the year and for groups of months J (May to August), D (November to February), and E (March, April, September, October). The curves are drawn through (or near to) points (×) each giving the mean l_2 or λ_2 for a group of stations (the number of stations is indicated beside each point) within a moderate range of latitude.

average l_2 or λ_2 as a function of latitude. They are obtained by forming means of l_2 or λ_2 for groups of stations within fairly small ranges of latitude. For several stations the black dot is "off" the mean curve by more than the length of its uncertainty line. Some such departures are to be expected from the laws of chance, but some certainly indicate that L_2 does not depend on latitude only, as was pointed out in the preceding section.

Similar mean curves of l_2 and λ_2 as functions of latitude are given in Fig. 9 for the annual mean of L_2, and for the group means for the D group of months (as in Fig. 8), the J (June solstitial) group May to August, and the E (equinoctial) group March, April, September, October. The individual values of l_2 and λ_2 with their uncertainty lines are not shown.

The outstanding feature of the annual change is the increased lag of high tide, by about an hour, in the D group as compared with the J and E groups. The amplitude is on the whole greatest in the J group of months.

Both these features of L_2 are illustrated in Fig. 7, for the region of North America. Besides the annual mean dial arrow for several United States and Canadian stations, Fig. 7 gives the forward halves of the dial arrows for the J, E, D groups of months. Not every station shows the two features just described, but this may be partly due to the accidental errors affecting all these determinations—errors about $\sqrt{3}$ times as great as for the annual means.

Figure 10*a* shows this annual change of L_2 in another way for a single station, Taihoku, Formosa [43]. It shows the dial points and their probable error circles for the annual mean L_2 (marked Y for year) and for the three groups of months J, E, D.

The annual change is, however, not fully shown by such means for four-monthly groups. It is better, where adequate data are available, to determine L_2 for each

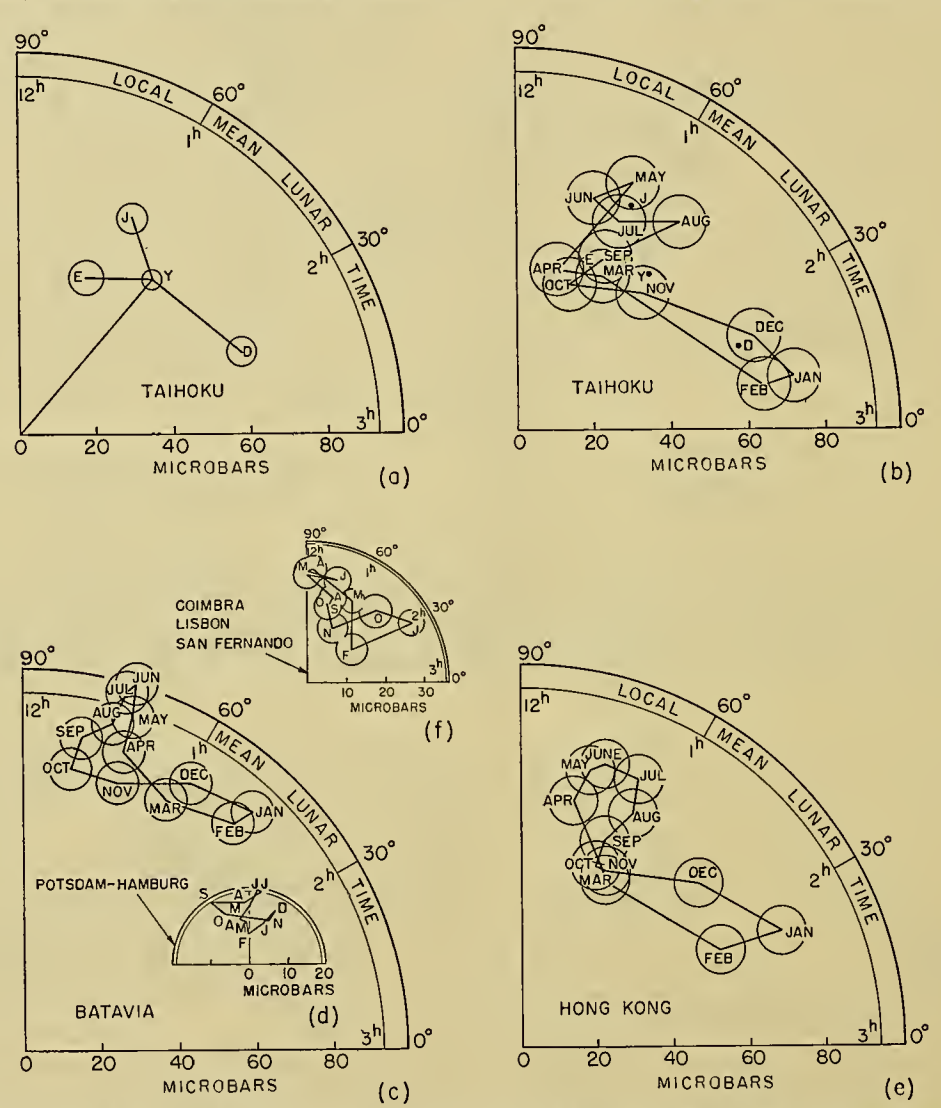


Fig. 10.—Harmonic dials (with probable error circles) indicating the annual change of the lunar semidiurnal air tide in barometric pressure. (*a*) Annual (Y) and J, E, D four-monthly determinations for Taihoku, Formosa (1897–1932). Also sets of twelve monthly mean dial points for (*b*) Taihoku (1897–1932), (*c*) Batavia (1866–1895), (*d*, inset) mean of Potsdam (1893–1922) and Hamburg (1884–1920), (*e*) Hong Kong (1885–1912), and (*f*, in centre) the mean of Coimbra, Lisbon, and San Fernando.

of the twelve calendar months (for many years). Figures 10*b*, *c*, *d*, *e*, *f* show sets of twelve dial points, one for each month, for (*b*) Taihoku [43], (*c*) Batavia [22, 23, 96–99], (*d*, inset in *c*) Potsdam and Hamburg [12] combined, (*e*) Hong Kong [29], and (*f*) the three Iberian-peninsular stations Coimbra, Lisbon, and San Fernando combined (using 112 years' data in all) [57]. Similar diagrams have been drawn for several other stations. Their most remarkable feature is the large lag of high tide, by nearly two hours, in January and February as compared with some of the *J* and *E* months. This is shown as well by the southern station Batavia as by the other (northern) stations.

The Lunar Tidal Variation of Temperature. The heating of the atmosphere by moonlight is quite negligible, but the moon nevertheless does produce a lunar semidiurnal variation of the air temperature, as a secondary consequence of its mechanical tidal action. The changes of air density accompanying the tidal variation of pressure will be different according as they take place isothermally or adiabatically, or in some intermediate way between these extremes. A similar question arises in the theory of sound waves. Newton [85*b*], who first calculated the speed of sound, assumed that the density variations are isothermal, and obtained a result that disagreed with his measurements. Laplace [74*f*] realized that the density variations are too rapid to allow the heat of compression to be conducted away during the brief period of each oscillation, and the assumption that the variations are adiabatic led him to the correct formula for the speed of sound.

The lunar atmospheric tide is a double tidal wave travelling round the earth each lunar day; the period is long—half a lunar day—but the distance between the places of high and low pressure, or condensation and rarefaction, is also great (except near the poles). Calculation shows that the density changes must be adiabatic as regards heat flow (either horizontal or vertical) *in the atmosphere* (except at very great heights—of the order 120 km—where the thermal conductivity of air is much increased owing to the long molecular free paths). The only possibility of the tidal oscillation being nearly isothermal is by interchange of heat with the liquid under surface (the solid earth does not conduct sufficiently to modify the tidal adiabatic temperature variations near ground level).

The adiabatic nature of the tidal air wave has been tested by determining the lunar semidiurnal variation of air temperature at Batavia, from sixty-two years' bihourly observations. Figure 11 shows the resulting dial vector with its probable error circle, within which lies the point *C* that corresponds to the adiabatic temperature variation, calculated from the known pressure variation L_2. Thus the determined and calculated temperature variations agree within the margin of accuracy of the determination [36].

It would be of interest to compute the lunar semidiurnal variation of air temperature from a long series of records from the windward side of some small flat tropical island in a great ocean. This would throw light on the degree of interchange of heat between air and sea. Bartels has suggested that it might prove advantageous, in such a study, to use only the night variations of air temperature, if these were less variable from day to day than the daytime values.

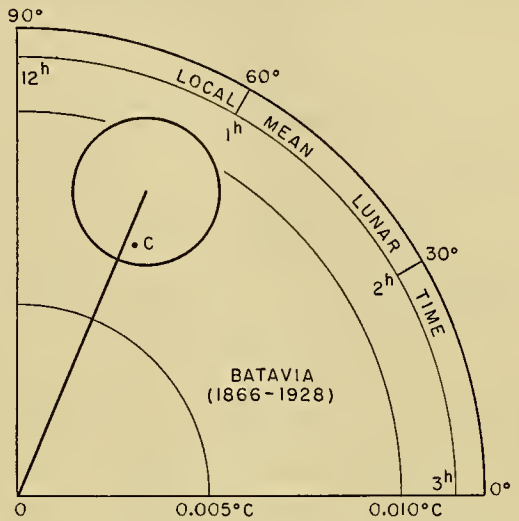


FIG. 11.—Harmonic dial specifying (with probable-error circle) the lunar semidiurnal tidal variation of *air temperature* at Batavia. The point *C* represents the variation calculated from the lunar semidiurnal variation of *barometric pressure* at Batavia, on the assumption that the density variations are adiabatic.

The Lunar Tidal Wind Currents. The tidal variations of sea level are accompanied by tidal currents (superposed on any other motions present, such as those in the Gulf Stream). Similarly in the atmosphere the L_2 pressure variation must be accompanied by tidal *wind* variations. These are best determined from bihourly values of the east-west and north-south component wind speeds. Only a few observatories, such as Mauritius and Bombay, have published such data, calculated from the usual wind records of direction and total speed. For this reason but little work has been done on the daily wind variations, whether solar or lunar. The only available *lunar* results are illustrated in Fig. 12*a* [47], which shows the L_2 dial vectors, with probable error circles, for the eastward and northward wind speeds at Mauritius, from sixteen years' observations (1916, 1917, 1920–33). The ratio l/r (see p. 514) is less than three, so that the determinations should be strengthened by using more data (which are available for Mauritius); similar investigations should be made also for other stations.

Figure 12*b* [47] shows in a similar way the dial vectors for the S_2 (solar semidiurnal) variations of east and north wind speed at Mauritius, derived from the same data, the amplitude scale being ten times less open than in Fig. 12*a*.

The ratios of the solar to the lunar amplitudes are of the order 20, rather greater than for the S_2 and L_2 pressure variations, for which at Mauritius the ratio is 17. The amplitude of the L_2 wind variations, about 1 cm sec^{-1}, is of the right order of magnitude according to the mathematical theory of these oscillations.

Figures 13*a*, *b* [47] show the actual wind-speed vectors due to the S_2 and L_2 wind variations at each hour of either half of the solar or lunar day (this diagram, of course, is not a dial diagram like Fig. 12). These wind velocities are superposed on any other local winds pres-

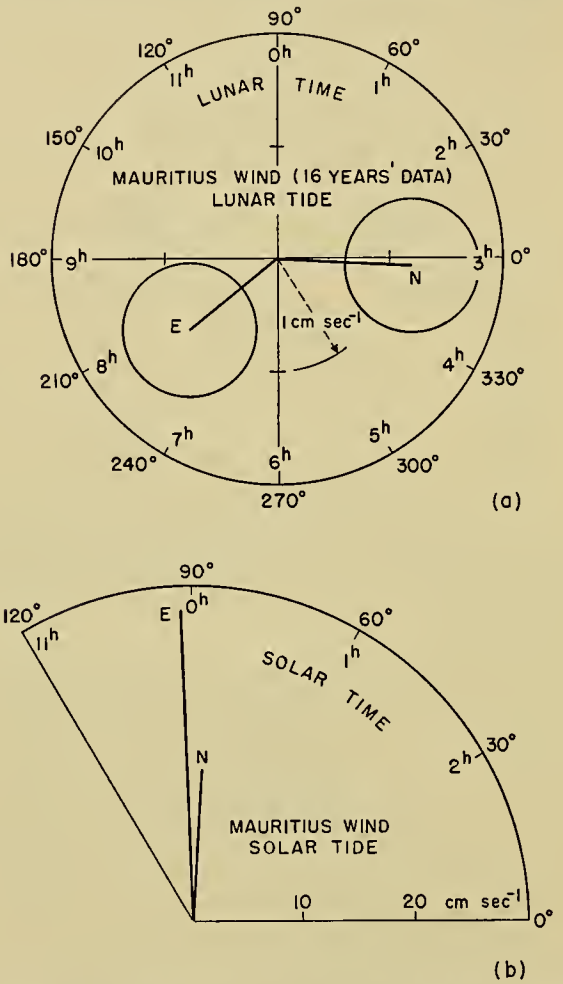


Fig. 12.—(*a*) Harmonic dial (with probable-error circles) for the annual mean *lunar* semidiurnal variations in the northward and eastward components of wind velocity at Mauritius, from about 16 years' bihourly data. (*b*) Harmonic dial for the corresponding annual mean *solar* semidiurnal variations. Note the tenfold scale difference between the two diagrams.

ent. The speed scale in Fig. 13*a* (lunar) is ten times more open than that for Fig. 13*b* (solar). The diagram Fig. 13*a* is, of course, not well determined.

The corresponding semidiurnal paths of Mauritius air particles due to these oscillations are similar in form and orientation to the ellipses of Fig. 13, which will represent these paths if all the time marks are advanced by three hours, and if the speed-scale is changed to a length-scale $T/2\pi$ times as great, where T denotes the duration in seconds of the (solar or lunar) half day. This factor is 6876 for the solar diagram 13*b*, and 7114 for the lunar diagram Fig. 13*a*. The distance scales are indicated on the left of each diagram. The extreme departure of any air particle from its mean position, at Mauritius, owing to these oscillations, is about 23 km for S_2 and about 1 km for L_2.

The Lunar Tidal Rise and Fall of the Ionospheric Layers. It has long been inferred from the evidence of the geomagnetic variations that there are horizontal lunar tidal currents in the ionosphere, the ionized electrically conducting region of the high atmosphere, containing at least two distinct layers, the E-layer at about 100 km height, and the F-layer at about 250 km. The first direct determination of the lunar tidal rise and fall of the high atmosphere was made in 1939 by Appleton

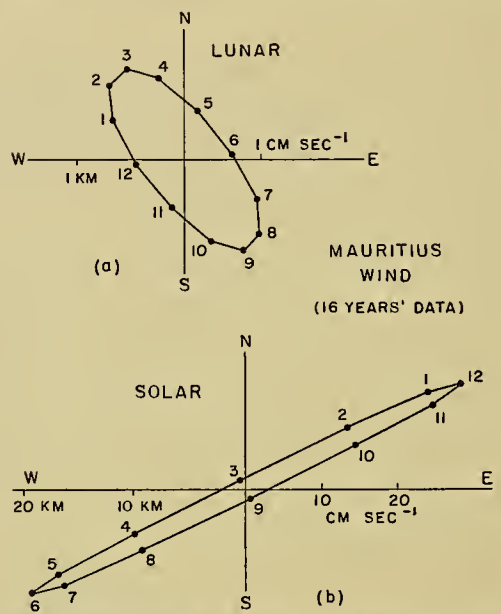


Fig. 13.—Diagrams based on Fig. 12, showing in plan the wind velocities at each lunar or solar hour (morning and afternoon) associated with the lunar and solar semidiurnal variations of wind at Mauritius. The velocity at each hour is represented (on the scale shown at the right) by the line (not drawn) from the center of the diagram to the numbered point for that hour. The diagrams also illustrate the corresponding *paths* of an air particle at Mauritius due to these wind variations, if the hour-numbers are increased by three; the distance scales are shown on the left of each diagram. Note the tenfold scale difference between the two diagrams.

and Weekes [6], who found from hourly radio measurements of the height of the E-layer above Cambridge, England, throughout several weeks, a lunar semidiurnal (L_2) variation of height of the E-layer amounting to 1 km above and below the mean level. This remarkable result is illustrated in Fig. 14, which shows eleven dial points each representing a determination from a period of twelve to fourteen days, between August 1937 and July 1938. The cross shows the mean dial point, and the probable error circle indicates the uncertainty of any one of the eleven points. The radius r for the mean point is $1/\sqrt{11}$ times less than this uncertainty, so that the determination is a good one.

Martyn [76–79] has lately examined the lunar tidal variations in the heights of the E- and F-regions, using the less accurate data available from routine recording at various ionospheric observatories throughout the

world.[6] Martyn finds at latitudes 35°S and 27°S a lunar variation in the height of the E-region which is opposite in phase to that found by Appleton and Weekes in the higher latitude of England. Martyn also finds lunar variations in both the heights and electron densities of the F-region. Near the magnetic equator these variations are very much larger than the already large variations found in the E-region. At Huancayo (Peru) the total tidal variation at certain hours and seasons amounts to some 60 km in height and 20 per cent in electron density. These variations concern the layers of electrons and ions interspersed among the neutral molecules of the high atmosphere. Their relation to the tidal oscillations of the main body of air requires consideration of electrodynamic as well as of hydrodynamic forces. There is no doubt that when the theory of these variations is fully understood the observational results will provide important and interesting information about the lunar atmospheric tide at high levels.

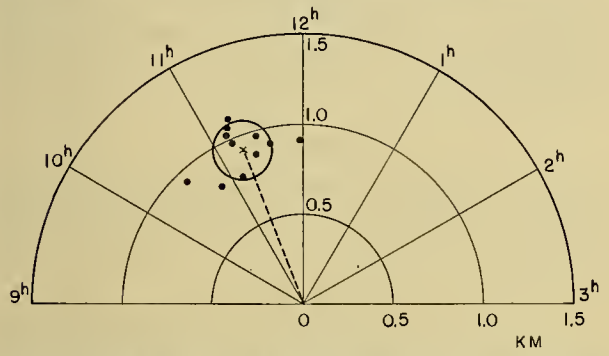


FIG. 14.—Harmonic dial [6] for the lunar tide in the E-region of the ionosphere. The circle shows the probable error for any one of the eleven separate dial points, each determined from 12–14 days' data.

Lunar Tidal Variations of Cosmic Rays. Cosmic ray observations provide, in a surprising and most interesting way, information as to the lunar air tide at a level of eighteen or twenty kilometres above the ground, though a precise interpretation of the data awaits further study. Among the cosmic rays received at the ground are mesons, supposed to be generated (by the primary rays) at this level; being unstable, a proportion of them are transformed on their way to the ground. If the lunar tide raises or lowers the level of the mean air pressure at which the mesons are generated, their path to the ground will be lengthened or shortened, and the number of survivors at ground level will be reduced or increased. Duperier [59] has made a reliable determination of L_2 in his recorded amounts of cosmic ray reception (mainly of mesons) at London, and has inferred therefrom that the lunar tide at about 18 km height is considerably magnified (about tenfold as compared with the equilibrium tide), though much less than in the ionospheric E-layer.

The Geomagnetic Lunar Tide. Nearly a century ago a small lunar daily variation was detected in the records of the components (or "elements") of the earth's magnetic field. This variation is produced (like the corresponding solar daily geomagnetic variation) mainly above the earth's surface, though these varying "primary" magnetic fields of external origin induce electric currents in the conducting body of the earth—mainly deep down, but also, to a lesser degree, near the surface, where they can be measured and recorded. The analysis of these earth-current records reveals solar and lunar daily variations, which form yet another curious by-product, as in the cosmic rays, of the high-level solar and lunar tidal atmospheric oscillations.

The external source of these solar and lunar daily geomagnetic variations consists of systems of electric currents flowing in some layer or layers (not yet clearly identified) in the ionosphere, and these currents are induced by mainly horizontal oscillatory large-scale motions of the ionospheric air. The process is similar to that in a dynamo, as first suggested by Balfour Stewart [42, 49]. The moving air corresponds to the armature, the conducting ionospheric layers to the armature windings, and the earth's main magnetic field to the field of the dynamo pole pieces.

From the determinations of S_n and L_n in the magnetic records of many stations it is possible to determine the distribution and intensity of these solar and lunar daily-varying electric current systems in the ionosphere, and also the *type* of the inducing atmospheric motions at those levels. To infer the *intensity* of these motions requires a knowledge of the electric conductivity of the layers in which the known electric currents flow. Until the precise situation of the currents is ascertained, and their electric conductivity, the intensity of the solar daily and lunar daily oscillations in the ionosphere cannot be precisely inferred from the geomagnetic data. The present indication is that the lunar tidal horizontal movements, like the lunar tidal rise and fall of the E-layer, are very greatly magnified as compared with what the barometric L_2 data would suggest. It is, however, of interest to note that the *ratio* of S_2 to L_2 in the magnetic records is about the same as that in the barometric variations. There is much scope for further investigation, both by observation and theory, of the bearing of the geomagnetic data on the solar and lunar daily atmospheric oscillations.

THE SOLAR SEMIDIURNAL OSCILLATION S_2

The Components S_n of the Solar Daily Barometric Variations. In middle and high latitudes the barometric variations are large, and mainly connected with weather changes. By averaging over many days selected in any way—days of a given season or calendar month, or days of high barometer or of rain—characteristic daily barometric variations can be determined. In the tropics, where large irregular barometric changes are rare, S_2

6. For other lunar ionospheric tidal determinations see [4, 5, 26, 81], and a forthcoming report by D. F. Martyn in the Zürich (1950) Proceedings of the International Union for Scientific Radio (U.R.S.I.). The lunar *diurnal* variation in the thickness of the F_2-layer in Alaska, reported by M. W. and J. G. Jones, in *J. Meteor*, 7: 14–20 (1950), is not real.

can be perceived (as also often at some European stations and others in moderate latitudes) even in the records for individual days.

By harmonic analysis the components S_n of the solar daily variation ($n = 1, 2, \ldots$) can be determined and separately studied.

The diurnal component S_1 [68, 69] differs remarkably from S_2, being (unlike S_2) much affected by local weather (cloud or sunshine) and topography. At the bottom of deep valleys it is greatly magnified. Its amplitude s_1 is much greater in summer than in winter; its phase σ_1 is about 90°, the maximum of S_1 thus occurring near local noon. It is not a world-wide oscillation, but a thermal effect [68, 71] sensitive to local influences. It will not be further considered here.

The components S_3 and S_4, with periods of eight and six hours, are also thermal effects (the sun's tidal action has no appreciable components with these periods). They result from the corresponding components T_3 and T_4 of the daily variation of air *temperature*, which they resemble in their geographical distribution and seasonal changes. For example, S_3 and T_3 have opposite phases in opposite hemispheres, and these phases are reversed from summer to winter. The S_3 barometric variation manifests a world-wide atmospheric oscillation [93], less simple and regular than the S_2 oscillation; the S_4 oscillation is still less regular [87]. These S_3 and S_4 oscillations should be further studied to fill in the framework of the whole subject of atmospheric oscillations [12–16, 18, 55, 102, 111].

The Solar Semidiurnal Oscillation S_2. In modern times Hann [63–67], Angot [3], Schmidt [92, 94], and Simpson [95] have taken a leading part in collecting and discussing the data for S_2. The literature is too vast to be cited here, but further references may be found in the papers quoted, particularly those of Hann.

Simpson's study was based on data from 214 stations. He illustrated the regularity of phase of S_2 in low latitudes by showing that at seventeen stations between latitudes ±10°, the local time of maximum of S_2 lay between 9.5^h and 10^h (A.M. and P.M.) at all but one, at which it was 10.3^h.

In the polar regions this uniformity of *local* time of maximum gives place to a different uniformity, that of *absolute* (*e.g.*, Greenwich) time of maximum [2, 9, 62]. At ten out of fifteen stations north of latitude 70°, this Greenwich time lay between 11.5^h and 12.5^h. As Schmidt [92] indicated, this shows that S_2 is a combination of a regular double wave travelling westward round the earth like the sun, and a semidiurnal oscillation of the air, symmetrical about the earth's axis, between the poles and the equator.

Simpson expressed S_2 at each station (in longitude φ east of Greenwich, expressed in time units, 15° per hour) as the sum of two terms corresponding to two such oscillations:

$$s_2 \sin(2t + \sigma_2) = b \sin(2t + B) + c \sin(2t - 2\varphi + C),$$

wherein t denotes *local* solar time and $t - \varphi$, *Greenwich* time. From 190 stations divided into eight latitude groups, from 10°S to 85°N, he determined b, B, c, C for each group, as given in Table I.

Table I. Constants for the Representation of S_2 at Various Latitudes [95]

Group		Travelling wave		Symmetrical wave	
Mean lat.	No. of stations	b (mm)	$B°$	c (mm)	$C°$
0	17	0.920	156.8	0.068	−4.0
18	15	0.835	155.3	0.082	−23.2
30	12	0.628	149.1	0.059	10.4
40	46	0.387	153.9	0.043	91.1
50	60	0.240	153.0	0.041	104.4
60	18	0.096	158.0	0.062	108.4
70	14	0.022	152.9	0.072	98.6
80	8			0.080	116.4
		Mean	154.1		

The phase B of the travelling wave has a remarkably small range (9° or ±9 minutes of time) in the eight zones. The amplitude b, decreasing steadily polewards from the equator, is well represented by the formula (l = latitude):

$$b = 0.937 \cos^3 l.$$

Up to about 60° latitude the other wave is of minor importance. Wilkes [111] has represented it by the formula:

$$[0.07 - 0.1 |\sin l|] \sin 2(t - \varphi) + 0.075 |\sin l| \cos 2(t - \varphi),$$

where $|\sin l|$ denotes the positive magnitude of $\sin l$.

Figure 15 shows dial vectors for S_2 (in barometric pressure) which relate to the points at their *thick ends* (not their centres, as in Figs. 5 and 7 for L_2). It illustrates

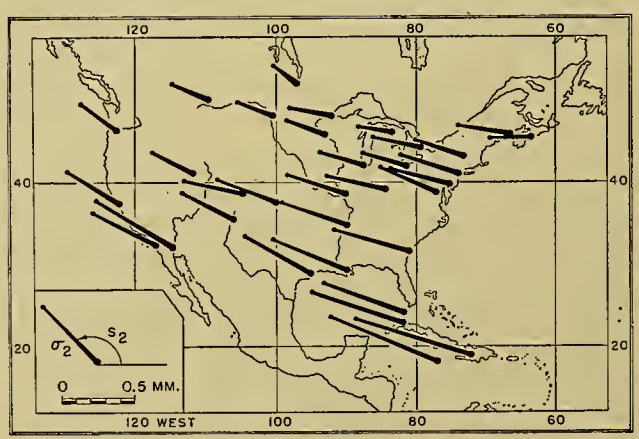


Fig. 15.—The distribution of the annual mean solar semidiurnal barometric variation (S_2) over North America. Each dial vector refers to S_2 at the point at its thick end (*cf.* with the lunar Figs. 5, 7, where each vector refers to L_2 at its centre).

the regularity of S_2 in phase and amplitude (decreasing northward) over North America; but it shows also that S_2 like L_2 (Fig. 7), though to a much less extent,

is reduced near the Pacific Coast [16]. This was noted by Hann, who also found that s_2 is less on the east Adriatic Coast than in Italy, and in the West Indies as compared with the East Indies, where indeed s_2, like l_2, is abnormally large.

The Annual Variation of S_2. The solar semidiurnal barometric variation also shows, as Hann indicated [67], considerable regularity in its change throughout the calendar year. This is illustrated in Fig. 16 by dial diagrams for four widely spaced stations in temperate latitudes (N or S), namely, (*a*) Washington, D. C. [16]; (*b*) Kumamoto (33°N, 131°E) [43]; (*c*) the mean of Coimbra, Lisbon, and San Fernando [57] (as in Fig. 10*f* for L_2); and (*d*) Montevideo, Uruguay [67].

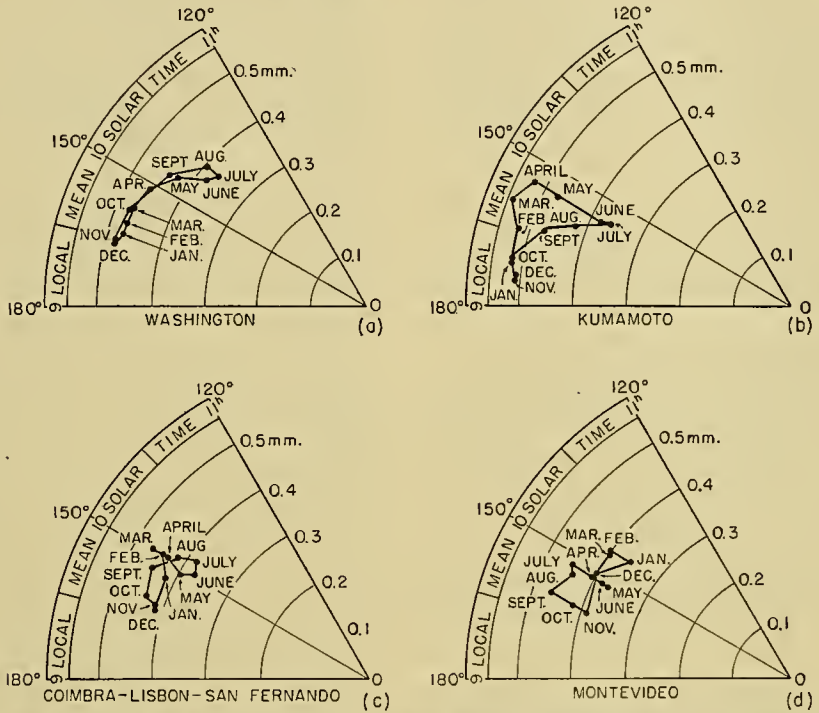


FIG. 16.—Harmonic dials indicating the annual change in the solar semidiurnal barometric variation (S_2) at four widely spaced points in middle latitudes, (*a*) Washington, D. C., (*b*) Kumamoto (33°N, 131°E), (*c*) mean of Coimbra, Lisbon, and San Fernando, (*d*) Montevideo, Uruguay.

The annual paths of the S_2 monthly dial points are much better determined than the L_2 paths in Fig. 10, yet they seem to differ more from one another, suggesting greater regularity of annual variation for L_2 than for S_2, which if established would be very remarkable.

Hann [67] considered separately the annual variations of s_2 and σ_2, though it would certainly be better to treat them together, that is, to discuss the annual movement of the S_2 dial point. He expressed the yearly variation of s_2 as the sum of a twelve-monthly and semiannual term:

$$s_2 = \alpha_0 + \alpha_1 \sin(x + \beta_1) + \alpha_2 \sin(2x + \beta_2),$$

reckoning x from mid-January at the rate 360° per year. He found that α_2 decreases polewards from the equator rather regularly, from 0.075 mm at the equator to 0.035 at 30° and 0.026 at 60° latitude, and that β_2 has a mean value of 311° from 0° to 40°S latitude, and 299° from 0° to 40°N. The corresponding means of α_1 and β_1 are 0.020, 0.078 and 134°, 94°. He discussed in much detail the regional irregularities in the distribution of α_1 and β_1.

As regards σ_2, Hann concluded that from 14°S to 50°N it has its maximum in January, and south of 14°S, in July. Figure 16 does not altogether confirm this. The many other details of Hann's discussion of S_2 cannot be summarized here. The data now available for S_2 and L_2 call for a more comprehensive comparative discussion than has yet been attempted.

THE THEORY OF THE ATMOSPHERIC OSCILLATIONS

Newton to Kelvin. Newton, in his *de Mundi Systemate* (the third part of his *Principia*) remarked that universal gravitation implies a tidal ebb and flow in the atmosphere as well as in the oceans. He rightly considered that it would be inappreciable—though it has been seen (on p. 519) that this flux and reflux *can* be determined by extensive computation from sufficient wind data.

Laplace in his dynamical theory of tides laid the foundation for all subsequent work on the subject. He first determined the tides of a liquid ocean completely covering a spherical rotating earth. In most of his work he took the ocean to be of uniform depth. Later he showed that he could associate the theory for such an ocean with the tidal oscillation of the (compressible) atmosphere, provided that the vertical accelerations are neglected (which we now know is permissible), that the

atmosphere is isothermal and of constant composition, and that the density variations accompanying the tidal motion occur isothermally.

Let p, ρ, T denote the pressure, density, and absolute temperature of the air at height Z above the ground, and let the suffix zero added to these and other symbols distinguish the values for $Z = 0$, that is, the ground values. The equation of static equilibrium of the air is

$$d(\ln p)/dZ = -1/H,$$

where $H = p/g\rho$ and the logarithm is to the natural base e. Clearly H is the height of the column of air of *uniform* density ρ required to give the pressure p (if the variation of g with height is ignored, though g is reduced by 3 per cent at 100 km). Hence H_0 is called the height of the homogeneous atmosphere; H is called the *scale height* of the atmosphere at the height Z, and is in general a function of Z.

In a perfect gas

$$p = knT = k\rho T/m = RT/M,$$

where

k denotes Boltzmann's constant (1.380×10^{-16} ergs per degree C),

n the number of molecules per cubic centimetre,

m the mean molecular mass,

$R = kN$, where N is Loschmidt's (less appropriately, Avogadro's) number (6.023×10^{23}) = 8.313×10^{7}.

M is the mean (chemical) molecular weight of the air (about 29); $M = Nm$. Hence

$$H = kT/mg = RT/Mg.$$

In the atmosphere considered by Laplace, H had the constant value H_0, because T and m are the same at all heights. Hence

$$p/p_0 = \rho/\rho_0 = e^{-z/H}.$$

Laplace showed that the tidal oscillations in such an atmosphere could be inferred from those for a liquid ocean of uniform depth H_0, which is therefore in this case called the (tidally) equivalent depth of the atmosphere. His theory, however, is not applicable to the actual atmosphere, in which T is not the same at all heights. Moreover, his condition that the density changes occur isothermally was doubtful. Newton had made the same assumption in his theory of sound waves, leading to an erroneous value for the speed c of sound, namely, $c^2 = p/\rho = gH$. Laplace had corrected this to $c^2 = \gamma gH$, where γ denotes the ratio of the specific heat at constant pressure to that at constant volume, to allow for the *adiabatic* character of such rapid density changes. It might well seem that for such slow oscillations as S_2 and L_2 the density changes would be isothermal, but calculation [34; 36; 111, p. 36] shows that they too must be adiabatic, because of the long wave length; and this is confirmed by the determination of the lunar tidal variation of air temperature at Batavia (p. 519).

It was to test his tidal theory that Laplace attempted, without success, to compute the lunar atmospheric tide L_2 for Paris. He realised that the magnitude of L_2 was very small, and as S_2 is so much larger he concluded that S_2 is due mainly to the sun's thermal action, the tidal contribution being insignificant. He seems also to have thought that there was little hope of constructing a theory of the oscillations produced in the atmosphere by its daily heating and cooling.

In 1882 Kelvin gave his attention to the subject, and in a presidential address to the Royal Society of Edinburgh [102] he quoted a table showing the 24-, 12-, and 8-hour periodic components (S_1, S_2, and S_3) of the solar daily barometric variation for thirty different stations. He pointed out that the 12-hour component exceeds the 24-hour component, especially in the higher latitudes, although in the daily variation of air temperature the 24-hour component is the larger. He suggested that the cause lies in the existence of an atmospheric free oscillation, of period nearer to 12 than to 24 hours, so that the 12-hourly thermal influence is magnified by *resonance*. His words were as follows:

> The cause of the semi-diurnal variation of barometric pressure cannot be the gravitational tide-generating influence of the sun, because if it were there would be a much larger lunar influence of the same kind, while in reality the lunar barometric tide is insensible, or nearly so. It seems, therefore, certain that the semi-diurnal variation of the barometer is due to temperature. Now, the *diurnal* term, in the harmonic analysis of the variation of *temperature*, is undoubtedly much larger in all, or nearly all, places than the *semi-diurnal*. It is then very remarkable that the *semi-diurnal term of the barometric effect* of the variation of temperature should be greater, and so much greater as it is, than the diurnal. The explanation probably is to be found by considering the oscillations of the atmosphere, as a whole, in the light of the very formulas which Laplace gave in his *Mécanique céleste* for the ocean, and which he showed to be also applicable to the atmosphere. When thermal influence is substituted for gravitational, in the tide-generating force reckoned for, and when the modes of oscillation corresponding respectively to the diurnal and semi-diurnal terms of the thermal influence are investigated, it will probably be found that the period of free oscillation of the former agrees much less nearly with 24 hours than does that of the latter with 12 hours; and that, therefore, with comparatively small magnitude of the tide-generating force, the resulting tide is greater in the semi-diurnal term than in the diurnal.

The Development of the Resonance Theory. The periods of free atmospheric oscillation, on a plane or spherical earth, were investigated by Lord Rayleigh [90], in 1890, with conclusions somewhat in favour of the resonance theory; but they could not be relied upon, partly because he did not take the earth's rotation into account.

Margules [75], with the explicit object of testing the resonance theory, investigated in much detail the free and forced oscillations of the atmosphere on the basis of Laplace's theory, and concluded that Kelvin's expectation was closely fulfilled. But Margules' conclusions cannot be relied upon either, partly because he did not take account of the true distribution of temperature with height in the atmosphere, which, indeed, at the time he wrote, was quite inadequately known. He also attempted to calculate the forced oscillations due to the daily variation of air temperature, on various hypotheses

as to its distribution with height, the most realistic being that its amplitude decreases exponentially upwards, as it would do if the heat were supplied at the ground and transmitted upwards by uniform conductivity; but he did not take into account the linear retardation of phase with height, which in this case should accompany the decrease of amplitude.

In 1910 Lamb [72] made a most important extension of Laplace's theory. His work related to an atmosphere on a plane base, thus abstracting from the problem the sphericity and rotation of the earth; but later [73] he removed these last two restrictions, though only for an atmosphere in convective equilibrium. His main discussion referred to an atmosphere in which H varies uniformly with the height ($H = H_0$ being a special case). He showed that the propagation of long waves in such an atmosphere is similar to that of long waves in a liquid ocean of depth H_0 in two special cases, namely, (1) Laplace's case, in which $H = H_0$ (or $T/m = T_0/m_0$) at all heights, and the density variations occur isothermally; and (2) for an atmosphere in adiabatic equilibrium (so that its height is $\gamma H_0/(\gamma - 1)$, the temperature T decreasing uniformly upwards to zero at the rate $(\gamma - 1)\, T_0/\gamma\, H_0$), and in which the density variations occur adiabatically (Laplace's case can be considered as corresponding to $\gamma = 1$).

The atmosphere is not in adiabatic equilibrium, so that it cannot be supposed, at least without further proof, that the tidally equivalent (liquid ocean) depth for the atmosphere is H_0. Lamb in fact showed that when H varies linearly with height, but not adiabatically, there is an infinite series of speeds for long waves, with the *implication* that there is a similar series of values of the equivalent depth h. However, the impression persisted for over twenty years that for any type of atmosphere there is just *one* value of h.

Lamb briefly discussed the resonance hypothesis of S_2 in his 1910 paper and in subsequent editions of his *Hydrodynamics*. He estimated from the improved form of Laplace's theory given by Hough [70], in terms of spherical harmonic functions, that if the atmosphere is resonant with a free oscillation similar to S_2 in its geographical distribution, h must be about 8 km, whereas for the actual atmosphere H_0 varies from about 7.3 km at the poles to 8.7 km at the equator. He continued:

> Without pressing too far conclusions based on the hypothesis of an atmosphere uniform over the earth, and approximately in convective equilibrium, we may, I think, at least assert the existence of a free oscillation of the earth's atmosphere, of "semi-diurnal" type, with a period not very different from, but probably somewhat less than, 12 mean solar hours.

He continued further:

> At the same time, the reason for rejecting the explanation of the semi-diurnal barometric variation as due to a gravitational solar tide seems to call for a little further examination. The amplitude of this variation at places on the equator is given by Kelvin as 0.032 inch. The amplitude given by the "equilibrium" theory of the tides is about 0.00047 inch. Some numerical results given by Hough in illustration of the kinetic theory of oceanic tides indicate that in order that this amplitude should be increased by dynamical action some seventy-fold, the free period must differ from the imposed period of 12 solar hours by not more than 2 or 3 minutes. Since the difference between the lunar and solar semi-diurnal periods amounts to 26 minutes, it is quite conceivable that the solar influence might in this way be rendered much more effective than the lunar. The real difficulty, so far as this point is concerned, is the *a priori* improbability of so very close an agreement between the two periods. The most decisive evidence, however, appears to be furnished by the *phase* of the observed semi-diurnal inequality, which is accelerated instead of retarded (as it would be by tidal friction) relatively to the sun's transit.

In 1924 Chapman [31] stressed the argument for strong resonance of S_2, based on the regularity of its geographical distribution, as compared with the considerable nonuniformity of the solar semidiurnal variation of air temperature (especially as between land and sea areas), which according to Lamb's last-quoted remark must be at least an important part of the cause of S_2. This argument he strengthened by contrasting the regularity of the geographical distribution of S_2, due partly to an irregular cause, with the degree of irregularity shown by L_2, whose cause is certainly distributed very regularly.

Chapman also extended Margules' calculation of the oscillations produced by the semidiurnal component of the daily variation of air temperature (T_2, see p. 522) taking account of the variation of phase (later discussed, in this connection, by Bjerknes [23*a*]), as well as of amplitude, with height. He concluded that the phase of the part of S_2 which is of thermal origin must be about 135° in advance of the phase of T_2, which he tried to estimate from the temperature data collected by Hann and others, using Taylor's estimate of the thermal conductivity due to eddy motion.

He also compared the magnitudes of the thermal and tidal contributions to S_2, which is possible because (substantially) both are affected by the same resonance magnification. He was able to show that they were of roughly equal order of magnitude (the inadequacy of the data regarding T_2 precludes a more accurate statement as yet), and on this basis he was able to explain the observed phase of S_2 from the phase of the thermal part, inferred from T_2, and on the assumption that the tidal part is in phase with the sun. This further enabled him to estimate the factor of resonance magnification as about 100. He was unable to prove that the atmosphere has a free period of oscillation (of the right geographical distribution) which would give this magnification. As the resonance magnification (when considerable) would be proportional to $1/(t_i - t_f)$, where t_i denotes the imposed period and t_f the free period, he concluded that despite the a priori improbability, $t_i - t_f$ cannot exceed 2 or 3 minutes, and is positive.

Whipple in 1918 [109], and also in 1924 in the discussion on [31], found great difficulty in accepting the resonance theory, on account of the possibility that such accurate "tuning" of the forced to the free oscillation might be upset by the large changes in air pressure and

temperature associated with weather and annual variations, and also on account of the difficulty the S_2 wave would have in twice daily surmounting the heights of Central Asia and the Rocky Mountains without losing a material fraction of its energy. These irregularities, however, are on a relatively small scale. The possibility that the variations of mean temperature from year to year might affect the tuning was examined by Bartels [12], but no such effect was found.

Taylor also was led to doubt the validity of the resonance hypothesis, despite the strength of the general arguments in its favour. Lamb had assumed, on the basis of the work already mentioned, that free oscillations of the atmosphere can exist which are identical, in distribution and period, with those of an ocean of such depth that long waves are propagated in it with the speed that he calculated for plane atmospheric waves. This assumption was used by Taylor [100] to estimate the period of the oscillation of S_2 type to be expected in an atmosphere in which the speed of propagation of long waves was that of the waves produced by the Krakatoa volcanic eruption of 1883. This great atmospheric pulse was propagated more than once round the entire earth, with a speed of 319 m sec^{-1}, which corresponds to an equivalent depth $h = 10.4$ km, a value markedly too great to give a free period (for an oscillation of S_2 type) nearly equal to 12 hours.

Lamb's assumption may be regarded as an extension of Laplace's theory of waves in an isothermal atmosphere. As realised later by Taylor, it involves the possibility that the atmosphere may have many equivalent depths, a contingency not possible in the atmospheres of the special type considered by Laplace and Lamb, for which $h = H_0$. Lamb's assumption is not obviously true, but in 1936 Taylor [101] proved its validity, and further developed Lamb's investigation of oscillations that are distributed in a similar way geographically (that is, as functions of longitude φ and colatitude θ), but have different height distributions of motion. In this work he was the first to take account of the cessation, at the tropopause, of the upward decrease of temperature.

The following year Pekeris [86] applied Taylor's methods to determine the free periods of an atmosphere in which the stratospheric temperature increases upwards above a certain height. This temperature distribution had been inferred from studies of the abnormal propagation of sound to great distances (beyond a zone of silence surrounding the source of sound), as well as from the heights of occurrence of meteors. Pekeris showed that, subject to a certain condition, the atmosphere could oscillate in ways corresponding to two equivalent oceanic depths. One of these was about 10 km, associated with a speed of propagation equal to that of the Krakatoa wave; the other gave a period (for a geographical distribution of S_2 type) of very nearly 12 hours, though the uncertainty of the upper atmospheric data precluded an exact calculation of the free period. The condition referred to was that the atmospheric temperature, after increasing upward above the stratosphere, should reach a maximum and thereafter decrease upwards to a low value. This was in agreement with the temperature distribution proposed by Martyn and Pulley [80] in 1936.

An important conclusion reached by Pekeris on this basis was that at high levels the pressure variation may be reversed in phase and highly magnified. This fitted well with the dynamo theory of the solar and lunar daily geomagnetic variations, the phases of which disagree with those of S_2 and L_2 at the ground, and the amplitudes of which, in the light of present knowledge of the electrical conductivity of the ionosphere, must be much greater than those of S_2 and L_2 at the ground. The magnified amplitude of L_2 there is confirmed by the determination [6] of the lunar tide in the E-layer, though the phase in the E-layer (over England) does not show the predicted reversal.

Pekeris [86] in 1939 re-examined the barometric traces for the Krakatoa waves and found evidence of a minor component wave propagated with the speed of 280 m sec^{-1}, corresponding to an equivalent depth of 7.9 km, accordant with a free oscillation (of S_2 type) with a period nearly equal to 12 hours. He showed that as the explosion occurred at a low level most of the energy of the pulse should go into the faster-travelling wave. (He estimated this energy as about 10^{24} ergs, roughly 1000 times that of the waves set up by the Nagasaki atomic bomb explosion.) The waves set up by the 1908 Siberian meteorite have been similarly examined [110] by Whipple.

The discussion by Pekeris has been extended by Weekes and Wilkes [107, 111], who have shown that according to theory the height distribution of the amplitude and phase of L_2 may differ materially in the ionosphere from that of S_2, if there is an upward rise of temperature in the E-layer followed by an upward decrease above some height in that layer, where the temperature has another maximum (with a further rise of temperature in the F-layer). The observational studies of the tidal motions and tidal changes of electron density in the different ionospheric layers, now being actively pursued [4, 5, 26, 76–81], will throw light on this possibility. Reliable measurements of ionospheric temperatures by rocket-borne instruments naturally provide a valuable additional basis for a detailed theory of the oscillations.

In the free oscillations of a liquid ocean of uniform depth h, the velocity components u, v, and w, and the departure Δp of the pressure from its mean value p (at each depth), have a relative geographical distribution (or variation with θ and φ) that is independent of depth. These dependent variables u, v, w, and Δp are taken to be proportional to (the real part of) $\exp i(s\varphi + \sigma t)$, where $2\pi/\sigma$ is the free period t_f. Laplace obtained a "tidal" equation which, for given values of s and h, determines a series of values of σ, and a corresponding series of functions representing the variation of u, v, w, and Δp with the colatitude θ.

The same equation is applicable to the *forced* oscillations of an atmosphere, whatever its temperature-height distribution (supposed uniform over the globe); in this case σ (as well as s) is known ($2\pi/\sigma$ is the imposed

period t_i), and the equation determines a series of values of h and a corresponding series of functions representing the latitudinal variation.

The height distribution of u, v, w, Δp, and $\Delta\rho$ (the departure of the density at any height from its mean at that height) is determined by a separate equation, conveniently expressed in terms of the pressure at each height, as an independent variable, by taking $x = -\ln(p/p_0)$. If we write

$$\text{div } \mathbf{v} = -D \ln(\rho + \Delta\rho)/Dt = (p_0/p)^{\frac{1}{2}}y,$$

where $\mathbf{v}$ denotes the vector velocity, and D/Dt the "mobile operator" [73], the equation is

$$\frac{d^2y}{dx^2} + \left[-\frac{1}{4} + \frac{1}{h}\left(\frac{\gamma - 1}{\gamma} H + \frac{dH}{dx}\right)\right] y = 0,$$

in which the height-distribution of the atmosphere (depending on the temperature T and the mean molecular mass m), is involved through H. In so far as y can be considered as of fairly constant order of magnitude (and this is a matter for examination by means of this equation), the expression for div $\mathbf{v}$ above indicates an upward increase of div $\mathbf{v}$ inversely proportional to $p^{\frac{1}{2}}$, that is, by 1000-fold at about the height of the E-layer.

Weekes and Wilkes have given an interesting interpretation of this equation by analogy with the propagation of electromagnetic waves in a medium having a variable refractive index. In the atmospheric case the expression for the analogous "equivalent" refractive index is

$$\mu^2 = -\frac{1}{4} + \frac{1}{h}\left(\frac{\gamma - 1}{\gamma} H + \frac{dH}{dx}\right).$$

If the height-distribution of H, for any value of h, makes μ^2 negative at a certain level, upward propagation of the energy, mainly put into the atmosphere by tidal or thermal causes in the lower layers, is effectively blocked. The air at that height acts as a barrier, total or partial, trapping the energy, and building up the amplitude in the whole spherical shell between the ground and the barrier, giving rise to resonance. If μ^2 is negative, not for all heights above the level Z at which it first becomes zero, but only for an interval of height, the barrier is partially transparent, and some of the oscillatory energy passes through it, either to a second (or third) barrier where there is a height interval of negative μ^2, or to the high levels at which thermal conductivity and dissipation of the energy into heat by viscosity become important. At these high levels the condition that $\Delta p/p$ or $\Delta\rho/\rho$ is small, as assumed in the equation, ceases to hold, and the modified differential equations will become nonlinear.

The conditions favouring negative μ^2 are that H should be small and that dH/dx should be either positive and small, or negative, corresponding to an upward *decrease* of temperature, because x increases upwards. The number of barriers to energy flow depends on the number of such regions of upward-decreasing temperature, but they alone are not sufficient to give a barrier, unless the value of h is appropriate, which in turn depends on the mode and period of the oscillation under consideration.

The boundary conditions in the equation for y are that at high levels the energy flow is upwards, and that at the ground ($Z = 0$ and $x = 0$) the vertical velocity is zero—unless the influence of the tidal motions of the under surface of the atmosphere (the tides in the oceans and in the solid earth) is being taken into account [12, 55], in which case at $Z = 0$ the vertical velocity w must have the corresponding distribution of values.

Wulf and Nicholson [112] have made a bold and imaginative attempt to explain the main irregularities of the geographical distribution of L_2, and its remarkable annual changes, stressing in particular the much greater and more widespread surface irregularities over the earth's Northern Hemisphere than over the Southern Hemisphere. Their suggestions need to be formulated and analysed mathematically, and tested also by reference to S_2, with its somewhat different geographical pattern and annual change.

The part of S_2 that is symmetrical about the earth's axis must be due to some inequality in longitude in the distribution of the semidiurnal component of the daily variation of air temperature, but no detailed study of this has yet been made.

SUGGESTIONS FOR FUTURE WORK

There is still great scope for useful extensions of our knowledge of the *facts* of the atmospheric oscillations, both solar daily and lunar daily. The newest and richest field of such observational and computational study is offered by the ionosphere with its several layers, each of which needs separate examination as regards the solar and lunar daily changes in its height, electron density, and other properties. Another new and almost untilled field of study is offered by the continuous records of cosmic radiation, whose components with different penetrative powers likewise deserve independent treatment. An older but far from exhausted field is that of the daily magnetic variations, which are causally due to the atmospheric oscillations in the ionosphere. Meteoric data may also add to our knowledge of the oscillations, though their more sporadic nature renders them less convenient for statistical treatment; this disadvantage may possibly be mitigated in the future if practical methods of radio observation of meteors, continuous throughout the day and night, are developed.

Even as regards the manifestation of the oscillations at ground level, where they have been studied for over a century, there remains much useful work to be done; the daily variations of pressure still need further study, and an improved treatment of the solar daily variation of air temperature is required to elucidate the thermal part of the solar half-daily tide. The lunar tide in air temperature found at Batavia might usefully be confirmed by a similar reduction for some continental station, and a reduction for some station on a small isolated island in mid-ocean would throw light on the systematic heat interchange between the air and the ocean.

The study of the lunar tidal winds has barely been

begun, and there is much scope for study of the periodic winds associated with both the solar and the lunar daily atmospheric oscillations.

In all these fields it is desirable to extend the search to the minor periodic components, such as the lunar diurnal tide, and the components associated with the changing distance of the moon.[7]

On the theoretical side also there is much scope for further work; the existing theory will need to be revised as our knowledge of the temperature-height distribution in the upper atmosphere advances through rocket investigations and in other ways; it will also need to be extended to take account of the nonuniform geographical distribution of temperature and of its height distribution in different regions. Moreover, the regional anomalies in the distribution of the lunar tide, as shown by the barometric variations, require quantitative explanation, including calculations of the effects of land and sea distribution as suggested by Wulf and Nicholson.

REFERENCES

1. Airy, G. B., *Greenwich Meteorological Reductions, 1854–1873, Barometer*. London, 1877. (See pp. 10, 14, 30)
2. Alt, E., "Die Doppeloszillation des Barometers, insbesondere im arktischen Gebiete." *Meteor. Z.*, 26: 145–164 (1909).
3. Angot, A., "Étude sur la marche diurne du baromètre." *Ann. Bur. météor. Fr.* (1887).
4. Appleton, E. V., and Beynon, W. J. G., "Lunar Tidal Oscillations in the Ionosphere." *Nature*, 162: 486 (1948).
5. —— "Lunar Oscillations in the D-Layer of the Ionosphere." *Nature*, 164: 308 (1949).
6. Appleton, E. V., and Weekes, K., "On Lunar Tides in the Upper Atmosphere." *Proc. roy. Soc.*, (A) 171: 171–187 (1939).
7. Bartels, J., "Zur Berechnung der täglichen Luftdruckschwankungen." *Ann. Hydrogr., Berl.*, 51: 153–160 (1923).
8. —— "Der lokale Anteil in der täglichen Luftdruckschwankung." *Beitr. Phys. frei. Atmos.*, 11: 51–60 (1923).
9. —— "Zur täglichen Luftdruckschwankung im Südpolargebiet." *Veröff. preuss. meteor. Inst., Ber., 1920–3*, Nr. 320, SS. 101–119 (1924).
10. —— "Barometrische Messung der Hochseegezeiten." *Ann. Hydrogr., Berl.*, 54: 222–227, 270–273 (1926).
11. —— "On the Determination of Minute Periodic Variations." *Quart. J. R. meteor. Soc.*, 52: 173–176 (1926).
12. —— "Über die atmosphärischen Gezeiten." *Veröff. preuss. meteor. Inst., Abh.*, Bd. 8, Nr. 9, 51 SS. (1927).
13. —— "Schwingungen der Atmosphäre." *Naturwissenschaften*, 15: 860–865 (1927).
14. —— "Gezeitenerscheinungen in der Atmosphäre." *Z. Geophys.*, 4: 1–17 (1928).
15. —— "Gezeitenschwingungen der Atmosphäre." *Handb. Exper. Physik*, Leipzig, W. Wien, F. Harms, Hsgbr., Bd. 25, Teil 1, SS. 161–210 (1928).
16. —— "Tides in the Atmosphere." *Sci. Mon.*, 35: 110–130 (1932).
17. —— "Berechnung der lunaren atmosphärischen Gezeiten aus Terminablesungen am Barometer." *Beitr. Geophys.*, 54: 56–75 (1939).
18. —— "Sonnen- und mondentägige Luftdruckschwankungen," *Lehrbuch der Meteorologie*. Hann-Süring, 5. Aufl., Bd. I, Leipzig, W. Keller, 1938. (See pp. 276–306)
19. —— und Fanselau, G., "Geophysikalischer Mond-Almanach." *Z. Geophys.*, 13: 311–328 (1937).
20. —— "Geophysikalische Mondtafeln 1850–1975." *Geophys. Inst. Potsdam*, Abh. Nr. 2, 44 SS. (1938). Also "Geophysical Lunar Almanac." *Terr. Magn. atmos. Elect.*, 43: 155–158 (1938).
21. Beale, J., "An Account of Some Mercurial Observations Made with a Barometer, and Their Results." (A letter to H. Oldenburg.) *Phil. Trans. roy. Soc. London*, 1: 153–159, 163–166 (1665–66).
22. Bergsma, P. A., "Lunar Atmospheric Tide." *Obsns. magn. meteor. Obs. Batavia*, 1: 19–25 (1871). (Contains meteorological observations for 1866 to 1868.) Calculations of this type, for individual years, carried out in routine fashion, have been published for the 40 years 1866–1905. See also summary for these years in Vol. 28 (for 1905), pp. 102 ff., published in 1907.
23. —— "Lunar Atmospheric Tide," "Influence of the Moon's Phases on the Mean Daily Barometric Pressure," and "Influence of the Moon's Phases on the Temperature of the Air." *Obsns. magn. meteor. Obs. Batavia*, 3: 37*-40*, 69*-71* (1878). (Contains meteorological observations for 1874 and 1875.)

23a. Bjerknes, J., "Atmospheric Tides." *J. mar. Res.*, 7: 154–162 (1948).

24. Börnstein, R., "Eine Beziehung zwischen dem Luftdruck und dem Stundenwinkel des Mondes." *Meteor. Z.*, 8: 161–170 (1891).
25. —— "Bemerkungen über die halbtägige Luftdruckschwankung im arktischen Gebiete." *Meteor. Z.*, 26: 519–521 (1909).
26. Burkard, O., "Gezeiten in der oberen Ionosphäre." *Terr. Magn. atmos. Elect.*, 53: 273–277 (1948).
27. Chapman, S., "An Example of the Determination of a Minute Periodic Variation, as Illustrative of the Law of Errors." *Mon. Not. R. astr. Soc.*, 78: 635–638 (1918).
28. —— "The Lunar Atmospheric Tide at Greenwich, 1854–1917." *Quart. J. R. meteor. Soc.*, 44: 271–280 (1918).
29. —— "The Lunar Tide in the Earth's Atmosphere." *Quart. J. R. meteor. Soc.*, 45: 113–137 (1919).
30. —— "The Lunar Atmospheric Tide at Mauritius and Tiflis." *Quart. J. R. meteor. Soc.*, 50: 99–112 (1924).
31. —— "The Semidiurnal Oscillation of the Atmosphere." *Quart. J. R. meteor. Soc.*, 50: 165–193 (1924).
32. —— "On the Determination of the Lunar Atmospheric Tide." *Z. Geophys.*, 6: 396–420 (1930).
33. —— "The Lunar Atmospheric Tide at Kimberley (1896–1915)." *Mem. R. meteor. Soc.*, 4: 29–33 (1932).
34. —— "On the Theory of the Lunar Tidal Variation of Atmospheric Temperature." *Mem. R. meteor. Soc.*, 4: 35–40 (1932).
35. —— "Tides in the Atmosphere." *J. Lond. math. Soc.*, 7: 68–80 (1932).
36. —— "The Lunar Diurnal Variation of Atmospheric Temperature at Batavia, 1866–1928." *Proc. roy. Soc.*, (A) 137: 1–24 (1932).
37. —— "The Lunar Atmospheric Tide at Honolulu, 1905–1924." *Quart. J. R. meteor. Soc.*, 61: 189–194 (1935).
38. —— "The Lunar Atmospheric Tide over Canada, 1897 to 1932." *Quart. J. R. meteor. Soc.*, 61: 359–365 (1935).

7. Such studies of magnetic and meteorological data are being fostered by a joint committee of the International Association of Meteorology, and the International Association of Terrestrial Magnetism and Electricity (of the International Union for Geodesy and Geophysics).

39. —— "The Lunar Tide in the Earth's Atmosphere." *Proc. roy. Soc.*, (A) 151: 105–117 (1935).
40. —— "The Lunar Atmospheric Tide in the Azores, 1894–1932." *Quart. J. R. meteor. Soc.*, 62: 41–45 (1936).
41. —— "The Lunar Atmospheric Tide at Glasgow." *Proc. roy. Soc. Edinb.*, (A) 56: 1–5 (1935–36).
42. —— *The Earth's Magnetism.* London, Methuen, 1936, 1951. (See Chaps. 3, 4)
43. —— "The Lunar Atmospheric Tide at Five Japanese Stations." *Quart. J. R. meteor. Soc.*, 63: 457–469 (1937).
44. —— "Tides in the Atmosphere." *Observatory*, 60: 154–165 (1937).
45. —— "The Lunar Atmospheric Tide at Accra, Gold Coast, Africa." *Quart. J. R. meteor. Soc.*, 64: 523–524 (1938).
46. —— "Tides in the Air." *P. V. Météor. Un. géod. géophys. int.*, Washington, 1939, 2: 3–40 (1940).
47. —— "Some Meteorological Advances Since 1939." *P. V. Météor. Un. géod. géophys. int.*, Oslo, 1948, in press.
48. —— and Austin, M., "The Lunar Atmospheric Tide at Buenos Ayres, 1891–1910." *Quart. J. R. meteor. Soc.*, 60: 23–28 (1934).
49. —— and Bartels, J., *Geomagnetism*, Vol. I. Oxford, Clarendon Press, 1940. (See Chaps. 7, 8)
50. —— and Falshaw, E., "The Lunar Atmospheric Tide at Aberdeen, 1869–1919." *Quart. J. R. meteor. Soc.*, 48: 246–250 (1922).
51. —— and Hardman, M., "The Lunar Atmospheric Tide at Helwan, Madras, and Mexico." *Mem. R. meteor. Soc.*, 2: 153–160 (1928).
52. —— "The Lunar Atmospheric Tide at Ocean Island." *Quart. J. R. meteor. Soc.*, 57: 163–166 (1931).
53. —— and Miller, J. C. P., "The Lunar Atmospheric Tide at Melbourne, 1869–1892, 1900–1914." *Quart. J. R. meteor. Soc.*, 62: 540–551 (1936).
54. Chapman, S., and Miller, J. C. P., "The Statistical Determination of Lunar Daily Variations in Geomagnetic and Meteorological Elements." *Mon. Not. R. astr. Soc.*, Geophys. Supp. 4: 649–669 (1940).
55. Chapman, S., Pramanik, S. K., and Topping, J., "The World-Wide Oscillations of the Atmosphere." *Beitr. Geophys.*, 33: 246–260 (1931).
56. Chapman, S., and Thomson, A., "The Lunar Atmospheric Tide at Apia, Samoa (1903–1927)." *Mem. R. meteor. Soc.*, 4: 21–25 (1932).
57. Chapman, S., and Tschu, K. K., "The Lunar Atmospheric Tide at Twenty-Seven Stations Widely Distributed over the Globe." *Proc. roy. Soc.*, (A) 195: 310–323 (1948).
58. Doodson, A. T., "The Harmonic Development of the Tide-Generating Potential." *Proc. roy. Soc.*, (A) 100: 305–329 (1922).
59. Duperier, A., "A Lunar Effect on Cosmic Rays." *Nature*, 157: 296 (1946).
60. Eisenlohr, O., "Untersuchungen über das Klima von Paris und über die vom Monde bewirkte atmosphärische Ebbe und Fluth." *Pogg. Ann. Phys. Chemie*, 60: 161–212 (1843).
61. Elliot, C. M., "On the Lunar Atmospheric Tide at Singapore." *Phil. Trans. roy. Soc., London*, 142: 125–129 (1852).
62. Greely, A. W., *Three Years of Arctic Service; An Account of the Lady Franklin Bay Expedition of 1881–1884.* New York, Scribner, 1894.
63. Hann, J. v., "Untersuchungen über die tägliche Oscillation des Barometers." *Denkschr. Akad. Wiss. Wien*, Abt. I, 55: 49–121 (1889).
64. —— "Weitere Untersuchungen über die tägliche Oscillation des Barometers." *Denkschr. Akad. Wiss. Wien*, 59: 297–356 (1892).
65. —— *Lehrbuch der Meteorologie.* Leipzig, C. H. Tauchnitz, 3. Aufl., 1915; 4. Aufl. (with R. Süring), 1926. Leipzig, W. Keller, 5. Aufl., 1938.
66. —— "Untersuchungen über die tägliche Oszillation des Barometers. Die dritteltägige (achtstündige) Luftdruckschwankung." *Denkschr. Akad. Wiss. Wien*, 95: 1–64 (1918).
67. —— "Die jährliche Periode der halbtägigen Luftdruckschwankung." *S. B. Akad. Wiss. Wien*, Abt. IIa, 127: 263–365 (1918).
68. —— "Die ganztägige (24 stündige) Luftdruckschwankung in ihrer Abhängigkeit von der Unterlage (Ozean Bodengestalt)." *S. B. Akad. Wiss. Wien.*, Abt. IIa, 128: 379–506 (1919). Reference [9, p. 110] gives corrections.
69. Hergesell, H., "Der tägliche Gang des Luftdruckes und der Temperatur in der freien Atmosphäre." *Meteor. Z.*, 36: 212–217 (1919). "Die täglichen Druck- und Temperaturwellen der Atmosphäre und die mit ihnen zusammenhängenden Vertikalbewegungen." *Beitr. Phys. frei. Atmos.*, 8: 179–193 (1919). "Die täglichen Wellen der meteorologischen Elemente in der Atmosphäre." *Beitr. Phys. frei. Atmos.*, 9: 30–66 (1920).
70. Hough, S. S., "On the Application of Harmonic Analysis to the Dynamical Theory of the Tides." *Phil. Trans. roy. Soc. London*, Ser. A, Part I, 189: 201–257 (1897); Part II, 191: 139–185 (1898).
71. Kleinschmidt, E., "Über die Ursache der halbtägigen Barometerschwankung, und über die mondtägige Wind- und Luftdruckschwankung auf Berggipfeln." *Beitr. Phys. frei. Atmos.*, 10: 151–168 (1922).
72. Lamb, H., "On Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 84: 551–572 (1910).
73. —— *Hydrodynamics.* Cambridge, University Press, 4th ed., 1916; 5th ed., 1924; 6th ed., 1932.
74. Laplace, P. S. (later Marquis de la Place), *Mécanique céleste*, Paris. (*a*) 2 (iv): 294–298 (1799); (*b*) 5 (xiii): 145–167 (1825); (*c*) 5 (xiii): 237–243 (1825); (*d*) 5 (Supp.): 20–35 (dated 1827, but published after Laplace's death in 1830); (*e*) 5 (xii): 95 (1825). The substance of (*b*) and (*c*) was taken from "De l'action de la lune sur l'atmosphère," *Ann. Chim.* (*Phys.*), 24: 280–294 (1823), and was reviewed, together with 74 (*f*) below, and partly translated, in "Berechnung der von dem Monde bewirkten atmosphärischen Fluth." *Pogg. Ann. Phys. Chem.*, 13: 137–149 (1828). Laplace also contributed "Additions" on this subject to the *Connaissance des Tems* [*sic*] for 1826 and 1830. See also 74 (*f*) Bouvard, A., "Mémoire sur les observations météorologiques faites à l'Observatoire Royal de Paris." *Mém. Acad. roy. Sci., Paris*, 7: 267–341 (1824). (Read April 1827, published after Laplace's death in 1830.)
75. Margules, M., "Über die Schwingungen periodisch erwärmter Luft." *S. B. Akad. Wiss. Wien*, Abt. IIa, 99: 204–227 (1890).
76. Martyn, D. F., "Atmospheric Tides in the Ionosphere. I. Solar Tides in the F_2 Region." *Proc. roy. Soc.*, (A) 189: 241–260 (1947).
77. —— "Atmospheric Tides in the Ionosphere. II. Lunar Tidal Variations in the F Region near the Magnetic Equator." *Proc. roy. Soc.*, (A) 190: 273–288 (1947).
78. —— "Atmospheric Tides in the Ionosphere. III. Lunar Tidal Variations at Canberra." *Proc. roy. Soc.*, (A) 194: 429–444 (1948).
79. —— "Lunar Variations in the Principal Ionospheric Regions." *Nature*, 163: 34–36 (1949).

80. —— and Pulley, O. O., "The Temperatures and Constituents of the Upper Atmosphere." *Proc. roy. Soc.*, (A) 154: 455–486 (1936).
81. McNish, A. G., and Gautier, T. N., "Theory of Lunar Effects and Midday Decrease in F_2 Ion-Density at Huancayo, Peru." *J. geophys. Res.*, 54: 181–185 (1949).
82. Miller, J. C. P., "On a Special Case in the Determination of Probable Errors." *Mon. Not. R. astr. Soc.*, 94: 860–866 (1934).
83. Morano, F., "Marea atmosferica." *R. C. Accad. Lincei*, 8 (v): 521–528 (1899).
84. Neumayer, G., "On the Lunar Atmospheric Tide at Melbourne." *Proc. roy. Soc.*, 15: 489–501 (1867).
85. Newton, I., *Philosophiae Naturalis Principia Mathematica*. London, 1687. (*a*) Bk. 1, Prop. 66, Cor. 19, 20; Bk. 3, Prop. 24, 36, 37 (also *de Mundi Systemate*, London, 1727, Sections 38–47, 49–54). (*b*) Bk. 2, Prop. 48–50.
86. Pekeris, C. L., "Atmospheric Oscillations." *Proc. roy. Soc.*, (A) 158: 650–671 (1937); "The Propagation of a Pulse in the Atmosphere." *Proc. roy. Soc.*, (A) 171: 434–449 (1939).
87. Pramanik, S. K., "The Six-Hourly Variations of Atmospheric Pressure and Temperature." *Mem. R. meteor. Soc.*, 1: 35–37 (1926).
88. —— "The Lunar Atmospheric Tide at Bombay (1873–1922)." *Mem. Indian meteor. Dept.*, 25: 279–289 (1931).
89. —— Chatterjee, S. C., and Joshi, P. P., "The Lunar Atmospheric Tide at Kodaikanal and Periyakulam." *Sci. Notes meteor. Dept. India.*, Vol. 4, No. 31, pp. 1–5 (1931).
90. Rayleigh, Lord (Strutt, J. W.), "On the Vibrations of an Atmosphere." *Phil. Mag.*, 5th Ser., 29: 173–180 (1890).
91. Sabine, E., "On the Lunar Atmospheric Tide at St. Helena." *Phil. Trans. roy. Soc. London*, 137: 45–50 (1847).
92. Schmidt, A., "Über die doppelte tägliche Oscillation des Barometers." *Meteor. Z.*, 7: 182–185 (1890).
93. —— "Zur dritteltägigen Luftdruckschwankung." (From a letter to von Hann.) *Meteor. Z.*, 36: 29 (1919).
94. —— "Die Veranschaulichung der Resonanztheorie." (In a review of Hann's paper [68].) *Meteor. Z.*, 38: 303–304 (1921).
95. Simpson, G. C., "The Twelve-Hourly Barometer Oscillation." *Quart. J. R. meteor. Soc.*, 44: 1–18 (1918).
96. Stok, J. P. van der, "On the Influence of the Moon upon the Cloudiness of the Sky." *Obsns. magn. meteor. Obs. Batavia*, Vol. 6 (for 1881 and 1882), Appen. 1, 2 pp. (1885).
97. —— "On the Lunar Atmospheric Tide." *Obsns. magn. meteor. Obs. Batavia*, 6: (3)–(8), Appen. 2 (1885).
98. —— "On the Influence of the Moon upon the Cloudiness of the Sky and the Temperature of the Air." *Obsns. magn. meteor. Obs. Batavia*, 9: (1)-(8) (for 1886), Appen. 1 (1887).
99. —— "An Inquiry into the Influence of the Moon's Phases on Rainfall" in *Rainfall in the East Indian Archipelago*. Vol. 4, 1882, Tables V-VII (cited in *Obsns. magn. meteor. Obs. Batavia*, 50: 20 (1927)).
100. Taylor, G. I., "Waves and Tides in the Atmosphere." *Proc. roy. Soc.*, (A) 126: 169–183, 728 (1929–30).
101. —— "The Oscillations of the Atmosphere." *Proc. roy. Soc.*, (A) 156: 318–326 (1936).
102. Thomson, W., (later Lord Kelvin) "On the Thermodynamic Acceleration of the Earth's Rotation." *Proc. roy. Soc. Edinb.*, 11: 396–405 (1882).
103. Tschu, K. K., "On the Practical Determination of Lunar and Luni-Solar Daily Variations in Certain Geophysical Data." *Aust. J. sci. Res.*, (A) 2: 1–24 (1949). (Some errors in this paper are corrected, *Aust. J. sci. Res.*, (A), in press.)
104. Wagner, G., "Die Änderung des Luftdruckes im anomalistischen Monat." *Beitr. Geophys.*, 11: 276–313 (1912).
105. —— "Der Einfluss des Mondes auf das Wetter." *Beitr. Geophys.*, 12: 277–328, 528–587 (1913). Corrects (p. 552) an error in reference [104]; contains a copious bibliography (pp. 578–587) of many (but not all) of the early works on the lunar atmospheric tide, including those by Bouvard, Kreil Danckwortt, Bouquet de la Grye.
106. —— "Zusammenstellung der Barometer-Beobachtungen von Samoa aus den Jahren 1903–1908 zur Bestimmung der Gezeitenbewegungen der Atmosphäre." *Abh. Ges. Wiss. Göttingen*, Bd. 9, Nr. 4, 48 SS. (1913).
107. Weekes, K., and Wilkes, M. V., "Atmospheric Oscillations and the Resonance Theory." *Proc. roy. Soc.*, (A) 192: 80–99 (1947).
108. Wegener, A., "Zur Frage der atmosphärischen Mondgezeiten." *Meteor. Z.*, 32: 253–258 (1915).
109. Whipple, F. J. W., "A Note on the Propagation of the Semi-Diurnal Pressure Wave." *Quart. J. R. meteor. Soc.*, 44: 20–22 (1918).
110. —— "The Great Siberian Meteor and the Waves, Seismic and Aerial, Which It Produced." *Quart. J. R. meteor. Soc.*, 56: 287–303 (1930).
111. Wilkes, M. V., *Oscillations of the Earth's Atmosphere*. Cambridge, University Press, 1949.
112. Wulf, O. R., and Nicholson, S. B., "Terrestrial Influences in the Lunar and Solar Tidal Motions of the Air." *Terr. Magn. atmos. Elect.*, 52: 175–182 (1947).

APPLICATION OF THE THERMODYNAMICS OF OPEN SYSTEMS TO METEOROLOGY*

By JACQUES M. VAN MIEGHEM

University of Brussels

HOMOGENEOUS SYSTEMS

Introduction. The systems which are generally considered in thermodynamics exchange energy (*e.g.*, heat) with their environment; however, they neither give up nor receive mass. It is for this reason that they are called *closed systems*. An *open system*, on the other hand, is a system which exchanges not only energy but also matter with the surrounding medium [4]. A cloud from which precipitation is falling is thus an open system.

It will be assumed in this first section that the system is homogeneous, that is to say that its physical variables (pressure p and absolute temperature T) are uniform throughout the volume V under consideration. This will always be the case in the atmosphere, provided one takes a volume of air which is not too extensive.

Uniformity of the quantities p and T involves an important consequence: The system cannot be the site of irreversible transformations which bring about the equalization of temperature or pressure between different points of the system. Therefore it is natural to suppose that physical transformations of the system, involving no internal modifications of mass, are reversible; only transformations involving changes of the masses m_j of certain of its constituents j may be irreversible.

At the foundation of any thermodynamic investigation lie two functions of state: the *internal energy* E and the *entropy* S of the system. With these are generally associated the *enthalpy* $H \equiv E + pV$ and the Gibbs *thermodynamic potential* $G \equiv H - TS$. Any one X of these functions depends on p and T, and on the masses $m_1, m_2, \cdots$ of the constituents. According to Gibbs, the functions of state of a system are homogeneous functions of the first degree in variables $m_1, m_2, \cdots$; therefore one has $X = \sum_j m_j x_j$, where the specific functions of state $x_j \left(\text{where } x_j = \frac{\partial X}{\partial m_j}\right)$ of the constituents j are homogeneous functions of zero degree in the variables m_j ($j = 1, 2, 3, \cdots$).

The First Law. The specific internal energy (per unit mass) of the system will be designated by e, its specific enthalpy by h, its specific volume by v, the heat added to the system from time t to time $t + dt$ ($dt > 0$) by dQ, and the heat added to the unit mass during the same time interval by dq. If we let m represent the total mass of the system, the following relations then hold: $H = mh$, $V = m\text{v}$, $dQ = mdq$.

The intensive (or local) form

$$dq = de + pd\text{v} = dh - \text{v}dp \qquad (1)$$

of the first law still applies when the system under consideration is open, provided that we assume the heat dq added to a unit mass includes, in addition to the heat received by radiation and conduction (the case of closed systems), the heat received by convection (an exchange of matter with the surrounding medium). Upon multiplying the two sides of (1) by m, we obtain the extensive form

$$dQ + hdm = dE + pdV = dH - Vdp \qquad (2)$$

of the first law for the case of an open system [7, 9, 14]. *The sum of the quantity dQ of heat received by the open system plus the enthalpy increase hdm of the system, resulting from exchange of matter with the surroundings, is equal to the increase dE in internal energy E of the system plus the term pdV.*

If the masses m_j of the constituents of the system are introduced, it is clear that $m = \sum_j m_j$. It should be noted that the total increase dm_j of mass m_j of constituent j arises from an increase $d_i m_j$ due to *internal* mass modifications and from an *external* contribution $d_e m_j$ ($\gtrless 0$) during the same time interval dt. Therefore $dm_j \equiv d_i m_j + d_e m_j$, with the condition $d_i m \equiv \sum_j d_i m_j = 0$, and as a result $dm = d_e m \equiv \sum_j d_e m_j \neq 0$. Furthermore the differential of any function of state whatever, for example H, assumes the form $dH = d_i H + d_e H$, with

$$d_i H = \frac{\partial H}{\partial T} dT + \frac{\partial H}{\partial p} dp + \sum_j h_j d_i m_j,$$

and

$$d_e H = \sum_j h_j d_e m_j,$$

where h_j is the specific enthalpy of constituent j. It then turns out, as a consequence of (2), that dQ can be written $dQ = d_i Q + d_e Q$, with

$$d_i Q = d_i H - Vdp, \quad d_e Q = \sum_j (h_j - h) d_e m_j, \qquad (3)$$

where the quantity $d_i Q$ of heat received by the system is associated with internal, physico-chemical changes of state which it undergoes during the time interval dt, while the quantity $d_e Q$ of heat received is associated with exchanges of mass with the surroundings during the same lapse of time. However, the separation of dQ into $d_i Q$ and $d_e Q$ was not accomplished as a result of the different mechanisms of heat exchange (radiation, conduction, or convection) between the system and the surroundings, but rather as a result of the effects produced by the added heat on the physico-chemical variables T, p, $m_1, m_2, \cdots$ of the system [14]. We observe

* Translated from the original French.

moreover that $d_eQ \equiv 0$ when the open system contains only a single constituent.

The Second Law. Gibbs' fundamental relation [2, 3, 5]

$$dH = TdS + Vdp + \sum_j \mu_j dm_j, \tag{4}$$

where $\mu_j \equiv \dfrac{\partial G}{\partial m_j} = h_j - Ts_j$ stands for the chemical potential of constituent j, is valid whatever the changes dT, dp, dm_j in state variables T, p, m_j of the system; consequently it is applicable to open systems. By substituting (2) into (4) it follows [7, 9, 14] that

$$dS - sdm = dQ/T + dQ'/T, \tag{5}$$

where the uncompensated heat of Clausius dQ' is defined by $dQ' = d_iQ' + d_eQ'$, in which

$$\begin{aligned} d_iQ' &= -\sum_j \mu_j d_i m_j \geq 0, \\ d_eQ' &= \sum_j (\mu - \mu_j) d_e m_j \gtrless 0, \end{aligned} \tag{6}$$

and $\mu \equiv h - Ts$. The specific entropy of the system is denoted by s, and $S = ms$. The increase dS which the entropy S of an open system undergoes from time t to time $t + dt$ ($dt > 0$) includes [7, 14]:

1. *The entropy change* due to exchange of matter and energy during the time interval dt,

$$dQ/T + d_eQ'/T + sdm \gtrless 0.$$

2. *The entropy production* in the interior of the system during the same time interval,

$$d_iQ'/T \geq 0.$$

The second law states that *the entropy production d_iQ'/T due to the internal mass transformation is real*, that is to say that the irreversibility of this transformation never entails destruction of entropy. By deduction from (3), (6), and (4),

$$d_iQ + d_iQ' = Td_iS. \tag{7}$$

Attention should be drawn to the fact that mass exchanges d_em_j between the system and the surroundings must take place at the temperature T and the pressure p of the system. This condition is usually satisfied in the case of a system which loses mass; when, on the other hand, mass enters the system the condition is no longer necessarily fulfilled. In the latter case it is well to make sure that this restriction is fully satisfied before applying the two laws which we have just formulated.

Fundamental Differences Between Open and Closed Systems. In the case of a *closed* system, the analytical expression $dQ = dE + pdV$ for the first law consists of three terms for each of which there is an exact, well-known physical meaning. We have already seen what meaning is assumed by the term dQ in the case of an *open* system. It should be noted that in this more general case the other two terms no longer have physical meanings. For example, the term pdV is no longer "the mechanical work done by the system during time interval dt," as a hasty generalization might lead one to believe [14]. Indeed, dV can be broken down into two additive terms, one of which d_eV depends essentially on the arbitrary addition of mass from the outside. Furthermore, when the boundary of a system is opened to an exchange of matter with the environment ($dm \neq 0$), the increase dX of any function of state X of the system is deprived of all physical sense. Mathematically, one obviously has $dX = mdx + xdm = \sum_j m_j dx_j + \sum_j x_j dm_j$. However, since the specific state functions x and x_j are determined only up to an arbitrary additive constant [14], the terms xdm and $x_j dm_j$ of dX have indeterminate values. On the other hand the expression $dX - xdm = mdx$ has a perfectly well-defined value, and consequently this is the case with expressions (2) and (5) for the two laws. Moreover, since the arbitrary constant which may be added to functions x and x_j is the same for all these functions [14], the differences $x - x_j$ and $x_j - x_k$, ($j, k = 1, 2, 3$), have well-defined values, and this is the case with expressions (3) and (6) for d_iQ, d_eQ, d_iQ', and d_eQ'.

In the case of a *closed* system, the second law can be written $TdS - dQ = dQ' = -\sum_j \mu_j dm_j \geq 0$, where $dQ' = 0$ corresponds to reversible transformations and $dQ' > 0$ corresponds to irreversible transformations. The physical meaning of the uncompensated heat of Clausius dQ' appears in the simplest fashion when a closed isothermal cycle is considered. During such a change of state, the uncompensated heat of Clausius received by the system is equal to the excess of the heat effectively given up over that effectively received by the system. In the case of an open system, however, dQ' is composed of two terms: the first depends on arbitrary addition or removal of mass and consequently may sometimes be positive and sometimes negative; the second depends on the internal physico-chemical transformation and is always positive. It is this latter term which actually generalizes, in the more realistic case of an open system, the classical concept of the uncompensated heat of Clausius.

In order to make clearly evident the differences in the significances of dQ and dQ', first for a closed and then for an open system, it is enough to set down the two laws for the case of a closed system consisting of two phases (a liquid and its vapor, for example) which undergoes only isobaric and isothermal transformations (vaporization and condensation at constant p and T), then to regard this system as being composed of two open systems (the gaseous phase and the liquid phase) and to see how the relations (2), (3), (5), and (6) reduce in this case [7].

Pseudoadiabatic Transformations. When the heat received by an open system does not alter the internal transformation of which the open system is the site, the system is said to undergo a *pseudoadiabatic* transformation [14]. In this case $d_iQ = 0$, and as a result of (3) and (7),

$$\begin{aligned} d_iH &= dH - \sum_j h_j d_e m_j = Vdp, \\ d_iS &= dS - \sum_j s_j d_e m_j = -\sum_j \frac{\mu_j}{T} d_i m_j. \end{aligned} \tag{8}$$

Let us now consider a system containing mass m_a of dry air, mass m_v of water vapor, and mass m_w of water (aqueous cloud), and further let us suppose that the system receives water or gives it up (rain). In this case

$$dm_a \equiv d_i m_a \equiv d_e m_a \equiv 0, \quad dm_v \equiv d_i m_v, \quad d_e m_v \equiv 0, \quad dm_w \equiv d_i m_w + d_e m_w,$$
$$d_i m_v + d_i m_w = 0, \quad dm = d_e m = d_e m_w \not\equiv 0. \tag{9}$$

The equation for pseudoadiabatic transformations of this system can be deduced at once from (8); it follows that

$$d_i S = dS - s_w d_e m_w = \frac{A_v}{T} dm_v \geq 0, \tag{10}$$

where $A_v \equiv \mu_w - \mu_v$ is the affinity of vaporization [2]. If S in (10) is replaced by its value $S = m_a s_a(T, p_a) + m_v s_v(T, p_v) + m_w s_w(T)$, where p_a and p_v represent the partial pressures of dry air and water vapor, we obtain, after using the relation $T(s_v - s_w) = L_v + A_v$ from [14], the differential equation of pseudoadiabatic transformations of moist air,

$$d\left(m_a s_a + \frac{m_v L_v}{T}\right) + (m_v + m_w) c_w \frac{dT}{T} + m_v d\left(\frac{A_v}{T}\right) = 0. \tag{11}$$

Here L_v stands for the heat of vaporization and c_w for the specific heat of water [14]. We suppose that the water vapor is at saturation value ($A_v = 0$), and we assume that the water that is brought in evaporates as soon as it is introduced or that the water formed in the interior of the system leaves as soon as it is formed ($m_w \equiv 0$, $d_i m_w = -d_e m_w = -d_i m_v$). Under these conditions the pseudoadiabatic transformation of the air is said to be reversible and dry [9, 14]; following (11)

$$d\left(s_a + \frac{r_v L_v}{T}\right) + r_v c_w \frac{dT}{T} = 0, \tag{12}$$

where $r_v \equiv m_v/m_a$ represents the mixing ratio of the water vapor. This equation is immediately integrable; employing the theorem of the mean, we find [9, 14]

$$(c_{pa} + r_v^* c_w) \ln T - R_a \ln p_a + \frac{r_v L_v}{T} = \text{const}, \tag{13}$$

where r_v^* is a mean value of r_v, c_{Pa} is the specific heat of dry air at constant pressure, and R_a is the specific perfect gas constant for dry air.

The finite equation (13) of reversible, pseudoadiabatic transformations of air saturated with water vapor is one of the fundamental equations of atmospheric thermodynamics; we have just derived it rigorously.

Polythermic Systems. A polythermic system is one in which all the phases do not have the same temperature. Since, by hypothesis, the temperature is a quantity which is constant throughout the interior of each phase, we can treat each one of them as an open system whose state is defined by the temperature of the phase, by the pressure p which is assumed the same for all phases, and by the masses of the constituents of the phase. The thermodynamics of polythermic systems, therefore, is only a special case of the thermodynamics of open systems [7, 14].

In order to establish our ideas, let us treat a closed polythermic system consisting of two phases: a gaseous phase (*first* phase) and a liquid phase (*second* phase). The gaseous phase comprises a mass m_a of dry air and a mass m_v of water vapor at temperature T'; the liquid phase, a mass m_w of water at temperature T''; the two phases are assumed to be at atmospheric pressure p. In this case we have $d_i m_a = d_i m_v = d_i m_w = 0$, $d_e m_a = 0$, $d_e m_v + d_e m_w = 0$, $dm_a = 0$, $dm_v \equiv d_e m_v$, $dm_w \equiv d_e m_w$. Application of the first law (2) to each of the two phases [7, 14] leads to the relation

$$[(d'Q)'' + h' dm_v] + [(d''Q)' + h'' dm_w] = 0, \tag{14}$$

which shows that the heat $(d'Q)''$ received by the *first* phase (gas) from the *second* phase (liquid) is not equal and opposite in sign to the heat $(d''Q)'$ given up by the *first* phase to the *second* phase. Furthermore, by a similar application [7, 14] of the second law (5) we obtain

$$dS = \frac{(d'Q)^*}{T'} + \frac{(d''Q)^*}{T''} + \left(\frac{1}{T''} - \frac{1}{T'}\right)(d''Q)' + \left[\frac{\mu_w(T'')}{T''} - \frac{\mu_v(T', p_v)}{T'} - \left(\frac{1}{T''} - \frac{1}{T'}\right) h_w(T'')\right] dm_v, \tag{15}$$

where $(d'Q)^*$ and $(d''Q)^*$ represent the heats added to the *first* and *second* phases by the surroundings of the system. It is clear that $d'Q \equiv (d'Q)^* + (d'Q)''$ and $d''Q \equiv (d''Q)^* + (d''Q)'$, where $d'Q$ and $d''Q$ are the quantities of heat received by each of the two phases. Equation (15) shows that the increase dS in the entropy S of the system consisting of two phases includes:

1. The change of entropy due to influx of heats $(d'Q)^*$ and $(d''Q)^*$ from the surroundings, and
2. The production of entropy resulting from exchanges of heat $(d''Q)'$ and matter dm_v between the phases. The second law states that there definitely is production and never destruction of entropy in the interior of the system.

Finally, the first law applied to the closed system consisting of the two phases provides the relation

$$dQ = dH - V dp, \tag{16}$$

where $dQ \equiv (d'Q)^* + (d''Q)^*$, and where the enthalpy H depends among other things on the temperatures T' and T'' of the *first* (gas) phase and the *second* (liquid) phase, respectively.

The fundamental equations (14), (15), and (16) allow studies to be made of the "horizontal mixing" of two air masses of different temperature and of the evaporation of rain in the free atmosphere [14].

The Temperature of the Wet-Bulb Thermometer and the Equivalent Temperature. Let us consider a closed system consisting of a mass m_a of dry air, a mass

m_v of water vapor at temperature T', and a mass m_w of water at temperature T''. Both phases are presumed to be at atmospheric pressure p. We will assume that the system undergoes only adiabatic ($dQ = 0$) and isobaric ($dp = 0$) transformations; in this case we have, (16),

$$m_a = m_a^*, \qquad m_v + m_w = m_v^* + m_w^*,$$

and

$$H(T', T'', m_a, m_v, m_w) = H(T'_*, T''_*, m_a^*, m_v^*, m_w^*),$$

where (T', T'', m_a, m_v, m_w) and $(T'_*, T''_*, m_a^*, m_v^*, m_w^*)$ represent two states of the system at the same pressure p. It turns out [14] that

$$\begin{aligned}(m_a c_{pa} + m_v c_{pv})(T' - T'_*) &= (m_v^* - m_v)\,[L_v(T'_*) + h_w(T'_*) - h_w(T'')] \\ &\quad + m_w^*\,[h_w(T''_*) - h_w(T'')],\end{aligned} \tag{17}$$

where $L_v(T'_*) = h_v(T'_*) - h_w(T'_*)$ is the heat of vaporization of water and c_{pa} and c_{pv} are the specific heats at constant pressure of dry air and water vapor.

Let us first consider a mass $(m_a + m_v)$ of moist air at temperature $T = T'$. At pressure p, assumed constant, we evaporate into this system a mass of water m_w such that the vapor becomes saturated at temperature T'_*. This temperature, which is represented by T_w, is by definition the temperature of the wet-bulb thermometer if the evaporating water assumes the temperature T_w of the saturated gaseous phase. Thus we are led to put $T' \equiv T$, $T'_* \equiv T'' \equiv T_w$, and $m_w^* \equiv 0$ in (17) from which is obtained the psychrometric formula

$$T_w = T - \frac{[r_v(T_w) - r_v(T)]L_v(T_w)}{c_{pa} + c_{pv} r_v(T)}, \tag{18}$$

where $r_v(T)$ stands for the mixing ratio of water vapor at temperature T.

Now let us consider a mass $m_a^* + m_v^*$ of moist air at temperature $T = T'_*$. We will assume that at constant pressure p one can condense the mass m_v^* of water vapor contained in the mixture. The temperature T' which the system then reaches is the equivalent temperature T_e of the moist air under consideration, provided it is assumed that the vapor can be condensed at temperature T'_* of the gaseous phase. Thus we are led to put $T'_* \equiv T'' \equiv T$, $T' \equiv T_e$, and $m_w^* \equiv m_v \equiv 0$ in (17), from which results von Bezold's formula

$$T_e = T + \frac{r_v(T)L_v(T)}{c_{pa}}. \tag{19}$$

We note that this isobaric condensation is a fictitious transformation, since it can never be realized physically.

NONHOMOGENEOUS SYSTEMS

General Remarks. The open nonhomogeneous system of greatest interest to meteorologists consists of a turbulent fluid. The *physical* state of the fluid is defined by the specific mass ρ and the pressure p, its state of motion by the components (v^1, v^2, v^3) of velocity $\mathbf{v}$ in a system of rectangular Cartesian coordinates (x^1, x^2, x^3), fixed with respect to the earth.[1] Scalars ρ and p and vector $\mathbf{v}$ are functions of x^1, x^2, x^3 and of time t.

By hypothesis, a mass element[2] $\rho\delta x^1\delta x^2\delta x^3$, moving at velocity v^k, exchanges no mass with the surroundings (absence of *molecular* diffusion). Therefore the system satisfies the equation of continuity in ρ and v^k; its movement is governed by gravity, the Coriolis force, and the pressure gradient force. The equations of balance of momentum $\rho\mathbf{v}$ defined by the components ρv^k, and of kinetic energy $k \equiv \frac{1}{2}v^i v^i$, can easily be deduced from the Eulerian equations of motion [11].

We designate by $\bar{X}$ the Reynolds mean of the function X and by $\widehat{X} = \overline{\rho X}/\bar{\rho}$ the corresponding weighted mean. It should be recalled that $\bar{X}' = 0$ and $\overline{\rho X''} = \bar{\rho}\widehat{X''} = 0$, where $X' \equiv X - \bar{X}$ and $X'' \equiv X - \widehat{X}$.

The *mean* state of the fluid is defined by $\bar{\rho}$ and $\bar{p}$, and its mean motion by $\widehat{v^k}$, $(k = 1, 2, 3)$. Since $\overline{\rho v''^k} = 0$ there is no diffusion at the scale of the mean state and of the mean motion, and as a result the mass element $\bar{\rho}\delta x^1\delta x^2\delta x^3$, moving at the mean velocity $\widehat{v^k}$, does not exchange matter with its environment. Therefore the mean system satisfies the equation of continuity in $\bar{\rho}$ and $\widehat{v^k}$. We know that the movement is governed by gravity, the Coriolis force (velocity $\widehat{v^k}$), the force due to the gradient of mean pressure $\bar{p}$, and the resultant force due to the Reynolds stresses $R_j^k \equiv R_k^j \equiv -\overline{\rho v''^k v''^j}$ $(j, k = 1, 2, 3)$. The equations of balance of the momentum components $\bar{\rho}\widehat{v^k}$ and of the corresponding kinetic energy $k_m \equiv \frac{1}{2}\widehat{v^j}\widehat{v^j}$ of the mean motion can easily be deduced from the equations of motion [11]. We observe that the absence of diffusion of mass ($\overline{\rho v''^k} = 0$) at the scale of the mean state does not imply the absence, at this scale, of diffusion of the properties of the fluid. In fact, if f represents the specific magnitude of any property whatever, we have in general $\overline{\rho f v''^j} \neq 0$. For example, by substituting $f \equiv v^k$ we would obtain $\overline{\rho v^k v''^j} = \overline{\rho v''^k v''^j} = -R_j^k$. Thus it appears that the Reynolds stresses R_j^k result from turbulent diffusion of momentum ρv^k. By contrast, a mass element $\rho\delta x^1\delta x^2\delta x^3$ moving with mean velocity $\widehat{v^k}$ diffuses mass and for this reason constitutes an open system, which does not admit an equation of continuity.

Finally, by subtracting the equations of kinetic energies $\widehat{k}$ and k_m, we find the equation of balance of turbulent kinetic energy $k_t \equiv \frac{1}{2}v''^i v''^j$ [1, 11],

$$\frac{\partial}{\partial t}(\bar{\rho}\widehat{k_t}) + \frac{\partial}{\partial x^j}(\bar{\rho}\widehat{k_t}\widehat{v^j} + \overline{\rho k_t v''^j}) = \bar{\rho}(\Delta_t + \Delta_i) \equiv \bar{\rho}\kappa, \tag{20}$$

where

$$\Delta_t \equiv \frac{R_j^k}{\bar{\rho}}\frac{\partial \widehat{v^i}}{\partial x^k} \quad \text{and} \quad \Delta_i \equiv \widehat{\frac{dv^k}{dt}\,v''^k}, \qquad (j, k = 1, 2, 3). \tag{21}$$

1. In order to simplify the discussion we have neglected molecular viscosity. Extension of the results of this section to the case of a viscous fluid does not present any difficulty [11].

2. By hypothesis, $\delta t \equiv 0$.

Here, Δ_t represents the fraction of kinetic energy of mean motion dissipated by turbulence per unit mass and time, and Δ_i the work per unit mass and time done in the course of eddying motion v''^i by the resultant of all the effective forces (excluding Reynolds' apparent force $\partial R_k^i/\partial x^i$) applied to the unit of mass considered. Work Δ_i can be regarded as being energy released by instability, per unit mass and time, in the course of the eddying motion v''^k [11]. When $\Delta_i < 0$, the turbulent motion is stable; when $\Delta_i > 0$, it is unstable [11]. In the first case Δ_i represents the fraction of turbulent kinetic energy transformed into heat per unit mass and per unit time; in the second, it represents the quantity of heat transformed into turbulent kinetic energy [13].

Finally we observe that turbulence can maintain itself only if $\kappa > 0$ (generalization of the criterion of L. F. Richardson). .

The First Law. The unit mass of the fluid (ρ, p, v^k) being a closed system, we can apply to it the principle of conservation of energy, from which we obtain [10]

$$\frac{d}{dt}(e + k + \varphi) = \frac{dq}{dt} - \frac{1}{\rho}\frac{\partial p v^k}{\partial x^k}, \qquad (k = 1, 2, 3). \quad (22)$$

This equation states that, *per unit mass and time, the increase in the sum of internal energy e, kinetic energy k, and potential energy φ, is equal to the quantity of heat received dq/dt, plus the surface work done by the surroundings on the unit of mass under consideration during the same time interval.*

Because of the equation of kinetic energy k and the equation of continuity in ρ and v^k, equation (22) takes the form (1), the enthalpy h being defined by $\rho h = \rho e + p$.

By setting

$$\rho \frac{dq}{dt} \equiv \rho\alpha - \frac{\partial W^k}{\partial x^k}, \quad (23)$$

where α represents the rate of production of heat per unit mass and W^k the components of heat flux (conductivity and radiation), equation (22) assumes (taking account of the equation of continuity) the form of an equation of balance [1, 11]:

$$\frac{\partial}{\partial t}\rho(e + k + \varphi) + \frac{\partial}{\partial x^k}\left[\rho(e + k + \varphi)v^k + pv^k + W^k\right] = \rho\alpha. \quad (24)$$

The existence of an equation of continuity for the system $(\bar\rho, \bar p, \widehat{v^k})$ justifies application of the Reynolds mean operation to the two sides of (24). By simplifying the equation thus obtained, with the help of the equation of kinetic energy k_m and of equation (20), we finally obtain the equation of balance of mean internal energy [11],

$$\frac{\partial}{\partial t}(\bar\rho\widehat{e}) + \frac{\partial}{\partial x^k}[\bar\rho\widehat{e}\widehat{v^k} + \overline{\rho h''v''^k} + W^k] = -\bar p\frac{\partial\widehat{v^k}}{\partial x^k} - \bar\rho\Delta_i + \overline{\rho\alpha}. \quad (25)$$

This equation shows that the apparent diffusion due to turbulence causes a flux of convective heat, described by its components [6, 11]

$$W_t^k \equiv \overline{\rho h''v''^k}. \quad (26)$$

Finally, if the equation of continuity in $\bar\rho$ and $\widehat{v^k}$ is substituted in (25), it follows [11] that

$$\frac{\widehat{d}\widehat{e}}{dt} + \bar p\frac{\widehat{d}}{dt}\left(\frac{1}{\bar\rho}\right) + \Delta_i = \frac{\widehat{d}q_m}{dt}, \quad (27)$$

where $\widehat{d}/dt$ represents the rate of variation of any quantity following the mean motion and where

$$\bar\rho\frac{\widehat{d}q_m}{dt} \equiv \overline{\rho\alpha} - \frac{\partial}{\partial x^k}(W^k + W_t^k) \quad (28)$$

represents the heat received per unit volume and per unit time. This quantity of heat includes the heat produced in a unit volume and the heat added by radiation, conduction, and convection (in the sense of the physicists).

Equation (27) states that *the quantity of heat received, per unit time and per unit mass of fluid, is equal to the sum of the increase in internal energy of the unit mass considered during this time interval, the mechanical work of expansion performed by this unit mass during the same time, and the energy of instability which it releases in this time interval.*

The Second Law. The second law states that the degradation of "noble" energy (in this case, kinetic energy) into internal energy or into heat is always accompanied by an increase in entropy. The evolution of the system is determined in this way.

In the case of a turbulent fluid there occurs simultaneously with this degradation the transformation of a fraction of the kinetic energy of mean motion into kinetic energy of turbulence, and also, more generally, a transformation of the kinetic energy at one scale of turbulence into energy at the scale of turbulence immediately smaller. Thus in a turbulent fluid, kinetic energy of motion is degraded not only into heat (the case of viscous fluid [11]), but also into kinetic energy of "inferior quality." The second law requires that the entropy of the system is always increasing in the course of these transformations of energy.

In order to account for this degradation, let us suppose that the mean physical state of a turbulent fluid has a corresponding mean specific entropy, which is a function of the mean internal energy and of the mean specific mass,

$$s_m = s_m(\widehat{e}, \bar\rho). \quad (29)$$

This satisfies the Gibbs relation (4),

$$\frac{\widehat{d}s_m}{dt} = \frac{1}{T_m}\frac{\widehat{d}\widehat{e}}{dt} - \frac{\bar p}{(\bar\rho)^2 T_m}\frac{\widehat{d}\bar\rho}{dt}, \quad (30)$$

where T_m is the "temperature" which characterizes the mean thermal state of the fluid and where

$$\frac{\widehat{d}}{dt} \equiv \frac{\partial}{\partial t} + \widehat{v^k}\frac{\partial}{\partial x^k},$$

($k = 1, 2, 3$). Upon substituting (27) and (28) into (30) we find, after taking account of the equation of continuity of mean motion [11],

$$\frac{\partial}{\partial t}(\bar{\rho} s_m) + \frac{\partial}{\partial x^k}\left[\bar{\rho} s_m \widehat{v^k} + \frac{W^k + W_t^k}{T_m}\right] = \sigma, \qquad (31)$$

where

$$\sigma \equiv -\frac{W^k + W_t^k}{(T_m)^2}\frac{\partial T_m}{\partial x^k} + \bar{\rho}\frac{(\widehat{\alpha} + \Delta_t - \kappa)}{T_m} > 0 \qquad (32)$$

represents the rate of production of entropy. The second law states that the entropy production is real ($\sigma > 0$). We recall that $\Delta_t - \kappa = -\Delta_i$ from (20).

Equation (31) shows that *the flux of mean entropy includes a convective flux and a thermal flux due to radiation, conduction, and turbulence. The production of entropy results from the nonuniformity of the temperature, from the production of heat, from the dissipation of kinetic energy of mean motion by turbulence, and from the destruction of turbulent kinetic energy* [11].

Finally, it should be noted that in the case of a viscous fluid the expression Δ_i which appears in (25), (27), and (32) must be replaced by $\Delta_i - \Delta_v$ where $\Delta_v (> 0)$ represents Rayleigh's dissipation function, that is, the fraction of kinetic energy of mean motion which is dissipated by viscosity (internal friction on the molecular scale) into heat per unit mass and per unit time.

The Heat Flux Due to Atmospheric Turbulence in the Vertical Direction. If it is assumed that the mean motion of the air is horizontal, the adiabatic eddying motion in the vertical direction is governed by the differential equation [11]

$$\frac{dv_z''}{dt} = -\frac{\Gamma c_{pa}}{\widehat{T}}[(\Gamma - \gamma) r_z - T_*''], \qquad (33)$$

where v_z'' = the eddying velocity in the vertical z direction,
r_z = the mixing length,
$\widehat{T}$ = the mean temperature of the dry air,
γ = the vertical lapse rate $-\partial\widehat{T}/\partial z$,
Γ = the adiabatic lapse rate (g/c_{pa} where g represents gravity),
T_*'' = the excess of the temperature of the parcel above the mean temperature of the level from which it originated.

After multiplying (33) by $\rho v_z''$ we obtain the expression for the energy of instability Δ_i released during the eddying motion, (21),

$$\bar{\rho}\Delta_i = \overline{\rho v_z'' \frac{dv_z''}{dt}} = -\frac{\Gamma c_{pa}}{\widehat{T}}[A(\Gamma - \gamma) - \overline{\rho T_*'' v_z''}], \qquad (34)$$

where $A \equiv \overline{\rho r_z v_z''} > 0$ is the exchange coefficient (Schmidt's austausch) in the vertical direction.

As for the heat flux W_t resulting from the adiabatic eddying motion in the vertical direction, it is given by (26):

$$W_t = c_{pa}\overline{\rho T'' v_z''} = -Ac_{pa}(\Gamma - \gamma) + c_{pa}\overline{\rho T_*'' v_z''}. \qquad (35)$$

The first term of the last member of (35) represents the flux of Schmidt and the second, the thermo-convective flux (Ertel, Priestley, and Swinbank) due to differences between the temperatures of parcels from the same level. These differences are initiated and maintained by sources of heat whose distribution is determined by $\overline{\rho\alpha}$. Because of this fact, parcels which are displaced vertically experience an additional hydrostatic buoyant force per unit mass equal to $gT_*''/\widehat{T}$. According to whether $T_*'' <$ or > 0, that is to say according to whether the particle is more or less dense than its surroundings, we have $v_z'' <$ or > 0 and consequently $T_*'' v_z'' > 0$. This inequality brings out the manner of organization of thermal convection. It turns out that the energy of instability due to the excess T_*'' of the parcel's temperature above the mean temperature of its starting level is always positive and therefore corresponds to a positive production of turbulent kinetic energy. This production can result only from a transformation of heat into kinetic energy, and consequently

$$\overline{\rho\alpha} > \frac{\Gamma c_{pa}}{\widehat{T}}\overline{\rho T_*'' v_z''} > 0. \qquad (36)$$

Moreover we observe that the thermal-convective flux of heat is always upward.

Finally, by returning to equation (32), it can be shown that the intensity of the entropy source associated with heat flux (35) is given by

$$\sigma = \frac{Ac_{pa}(\Gamma - \gamma)^2}{(\widehat{T})^2} + \frac{c_{pa}\gamma}{(\widehat{T})^2}\overline{\rho T_*'' v_z''} + \frac{1}{\widehat{T}}\left[\overline{\rho\alpha} - \frac{c_{pa}\Gamma}{\widehat{T}}\overline{\rho T_*'' v_z''}\right] > 0, \qquad (37)$$

provided that $T_m \equiv \widehat{T}$. From (36) and the inequality $\gamma > 0$, the inequality of (37) is always verified. The flux W_t is therefore compatible with the second law of thermodynamics without the necessity of assuming a priori that it is upward (as Ertel claims). It should be noted[3] that the heat flux of Schmidt involves an effective production of entropy represented by the perfect square in the first term of the expression for σ.

Application to the General Circulation of the Atmosphere. The equations which we have considered above can be of use in the study of the general atmospheric circulation. The "mean motion" is then a zonal current (westerly or easterly), and the "eddying motion" cor-

3. Bibliographical information and further details can be found in [11].

responds to perturbations embedded in this current. In this case it is clearly a question of turbulence on a very large scale (synoptic scale), essentially different from the small-scale turbulence which is usually studied. At the scale of the general circulation it becomes difficult indeed, if not impossible, to admit the conservation of potential temperature and of momentum or angular momentum of the large eddies along their trajectories.

On the other hand nothing allows us to state a priori that a fraction of the kinetic energy of the zonal current is really transformed into kinetic energy of large-scale turbulence (macroturbulence); in this case, the opposite might very well occur and probably does occur. In short, a thermomechanical theory of macroturbulence (*Grossturbulenz* of A. Defant) remains to be worked out.

The equations of balance of momentum and kinetic energy can be used to obtain qualitative indications of the production and transport of these quantities in the atmosphere [8, 12]. Unfortunately, in the case of large-scale turbulence we do not have the exact values of the Reynolds stresses R^k_j, the fraction Δ_t ($>$ or $<$ 0) of the kinetic energy of mean motion dissipated by turbulence, and the energy of instability Δ_i released in the course of the eddying motion. It is therefore impossible to make a quantitative theoretical study of the distribution of sources of zonal momentum, of kinetic energy, of internal energy, and of entropy, as well as of their manner of redistribution in the atmosphere.

Knowledge of the tensor of large-scale turbulence would permit us to achieve a better comprehension of the processes and the evolution of the general circulation of the atmosphere. New investigations, both theoretical and synoptic, into the sources and the flux of momentum, of energy in its various forms [13], and of entropy appear desirable.

Final Remarks. When the system under consideration is homogeneous, the fundamental equations of thermodynamics can be written for the case of arbitrary addition of mass ($dm \gtrless 0$) to the system. These equations have manifold applications not only in meteorology (clouds with precipitation, evaporation of precipitation, mixing of air masses of different temperatures, psychrometry), but also in chemistry, in biology, and in industry.

The situation is not quite the same when the system is nonhomogeneous (gradients of T and p), for in this case the material points of the system are necessarily in motion. Now the notion of movement can be defined and studied only if the moving element retains its identity in the process of its displacement; in other words, a system of dynamical equations can give a true representation of motion only if it includes the equation of continuity of mass, expressing the invariance of the mass of an element along its trajectory. Furthermore, in the case of a nonhomogeneous system, it is necessary to adopt in succession two scales of observation: first the microscopic scale, that is, the scale of the individual elementary eddy (the scale of the molecule, if one envisions diffusion and molecular viscosity), then the macroscopic scale, that is, the scale of the mean motion. It must be observed that the latter scale cannot be adopted arbitrarily, for there must be no diffusion of mass at this scale ($\overline{\rho v''^k} \equiv 0$). If this were not true, there would be no equation of continuity on the macroscopic scale, and it would be impossible to define and study the mean motion. Therefore the extension of the laws of thermodynamics for nonhomogeneous systems, to which mass can be added or removed arbitrarily, encounters difficulty in the very beginning unless it is possible to satisfy the condition $\overline{\rho v''^k} \equiv 0$. We shall assume this condition to be met; then for an observer who is carried along with the mean motion and follows the evolution of the system on a microscopic scale, the flux of the diffusion of mass is expressed by means of $\rho v''^k \neq 0$ and the intensity of diffusion by means of $-\overline{\frac{\partial \rho v''^k}{\partial x^k}} \neq 0$ ($k = 1, 2, 3$). However, as we have said before, the absence of diffusion at the scale of mean conditions ($\overline{\rho v''^k} \equiv \overline{\frac{\partial \rho v''^k}{\partial x^k}} \equiv 0$) does not imply the absence at this scale of diffusion of any arbitrary property f (intensive quantity) of the fluid, within the flow of which the mean flux of f is given by $\overline{\rho f v''^k} \neq 0$. Thus it is possible to carry out a study of the turbulent diffusion (eddy diffusion) in the atmosphere of water vapor ($f \equiv$ the specific humidity ϵ) or of any other substance in suspension in the air, of sensible heat ($f \equiv c_{pa}T$), of latent heat ($f \equiv \epsilon L$), of internal energy ($f \equiv e$), of entropy ($f \equiv s$), of kinetic energy ($f \equiv k$), of momentum ($f \equiv$ v), etc.

One final remark: The necessary condition $\overline{\rho v''^k} \equiv 0$ shows that v''^k is the fluctuation relative to a weighted mean, and consequently the components of the mean flow are necessarily the weighted means $\widehat{v^k}$ of the components v^k of the velocity of the elementary eddies. The weighted mean has the unique property of decomposing, in an additive fashion, the average of the total kinetic energy k into kinetic energy of mean motion and mean kinetic energy of turbulence ($\widehat{k} = k_m + \widehat{k_t}$). No other mean possesses this same property. Therefore, it is impossible to avoid the necessity of introducing the weighted mean into the study of turbulence of fluids.

REFERENCES

1. Cowling, T. G., "The Stability of Gaseous Stars." *Mon. Not. R. astr. Soc.*, 96: 42–60 (1935).
2. de Donder, T., "*L'affinité*." Paris, Gauthier-Villars, 1927.
3. —— and Van Rysselberghe, P., *Thermodynamic Theory of Affinity*. Palo Alto, Calif., Stanford University Press, 1936.
4. Defay, R., "Introduction à la thermodynamique des systèmes ouverts." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 15: 678–688 (1929).
5. Gibbs, J. W., "Equilibrium of Heterogeneous Substances" in *Collected Works*, Vol. I: *Thermodynamics*. New York, Longmans, 1928.

6. Montgomery, R. B., "Vertical Eddy Flux of Heat in the Atmosphere." *J. Meteor.*, 5: 265–274 (1948).
7. Prigogine, I., *Étude thermodynamique des phénomènes irréversibles*. Liége, Desoer, 1947.
8. Starr, V. P., "On the Production of Kinetic Energy in the Atmosphere." *J. Meteor.*, 5: 193–196 (1948).
9. van Lerberghe, G., et Glansdorff, P., "Contribution à la thermodynamique des systèmes ouverts." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 22: 484–497 (1936).
10. Van Mieghem, J., "Thermodynamique des systèmes non uniformes en vue des applications à la météorologie." *Geofys. Publ.*, Vol. 10, No. 14, 18 pp. (1935).
11. —— "Les équations générales de la mécanique et de l'énergétique des milieux turbulents en vue des applications à la météorologie." *Inst. R. météor. Belg., Mém.*, 34: 1–60 (1949).
12. —— "Production et redistribution de la quantité de mouvement et de l'énergie cinétique dans l'atmosphère. Application à la circulation atmosphérique générale." *J. sci. Météor.*, 1: 53–67 (1949).
13. —— "Comment on the Global Energy Balance of the Atmosphere." *Cent. Proc. R. meteor. Soc.*, pp. 173–175 (1950).
14. —— et Dufour, L., *Thermodynamique de l'atmosphère*. Bruxelles, Office International de Librairie, 1949. (See pp. 107–133)

THE GENERAL CIRCULATION

The Physical Basis for the General Circulation *by Victor P. Starr* 541

Observational Studies of General Circulation Patterns *by Jerome Namias and Philip F. Clapp* 551

Applications of Energy Principles to the General Circulation *by Victor P. Starr* 568

THE PHYSICAL BASIS FOR THE GENERAL CIRCULATION

By VICTOR P. STARR

Massachusetts Institute of Technology

Since time immemorial man has inescapably observed the atmosphere in which he lives and has his being. It would therefore seem reasonable to expect that at the present date the science of meteorology should be one of the most advanced fields of human endeavor. Yet, if a distinction is made between the mere collection of descriptive facts of observation on the one hand and interpretative work which aims to give a rational intellectual understanding of phenomena on the other, it must be confessed that our knowledge concerning the large-scale motions of the atmosphere is restricted mostly to the former category of information. Thus, for example, no one has as yet given a satisfactory rational explanation for one of the most outstanding features of the general circulation, namely the large belts of westerly winds in the temperate latitudes of each hemisphere. However, it must be recognized that it is only in the last few decades that anything approaching sufficiently complete global observations *for the checking of hypotheses* regarding the general circulation has become available, so that progress at a more accelerated pace should now be forthcoming.

The physiognomy of present-day meteorology bears much of the imprint imparted to it by the (so to speak) accidental way in which synoptic reporting networks developed and grew. In fairly recent times it was quite an achievement for a meteorologist to have at his command a network of reports large enough to depict an entire cyclone. The immediate temptation was then to treat this feature as an independent entity and to separate it artificially from its meteorological environment in seeking a rational explanation for it. The cyclone thus became a phenomenon that existed independently. More extensive observations now available point to the inadequacy of this tacit assumption. The cyclone constitutes a cogwheel in a larger mechanism and probably can be understood only in relation to and not independent of its atmospheric context.

A criticism which is the same in principle can be leveled against many other efforts to explain synoptic structures. Indeed meteorology is replete with attempts to formulate *ad hoc* explanations for individual details of the general circulation without due cognizance of their role as functioning parts of a global scheme.

In the more recent literature there are signs that we are outgrowing this restricted point of view; there are indications which emphasize the essential oneness of the atmosphere which must be studied as an internally integrated and coordinated unit. The general circulation presents a puzzle to be solved. We must learn how the various parts fit together into a whole if we are to understand it. With the hemispherical data now becoming available and the clues already at our disposal the avenue to sound progress is wide open and, at least as far as the writer is concerned, quite inviting.

It is therefore in the spirit of appraising our knowledge from the larger point of view that this commentary is written. Admittedly and intentionally the treatment reflects the writer's viewpoint gained through a number of years of concentration on the subject under consideration. As a candid personal sidelight it should be mentioned, however, that much of the material is in a sense a relatively recent culmination of the writer's striving toward a more unified and coherent conception of the global circulation. Nor is there currently any sign that this process has reached a state of final crystallization. This paper is therefore in the nature of an individualistic progress report designed to portray and to share with others a certain outlook and approach in the hope that really substantial and enduring progress may eventually be attained thereby. Within the space of these pages we cannot hope to give anything approaching an exhaustive exposition of various topics. For this reason only certain basically important highlights will be elaborated, and we shall rely upon the reader's competence to interpolate various items of information which are generally available in the literature. Furthermore, the selection of the subjects touched upon is not one dictated by an aim at logical completeness, but rather is limited to those aspects which in the writer's mind are most likely to be conducive to further results at the present stage of development of the science.

As an outstanding problem of paramount importance for human activities, it is rather astonishing that the global circulation of the atmosphere has not up to the present time received more consideration from physical scientists generally. This situation is probably due in part to the lack of proper observational information so necessary for the successful prosecution of research concerning the subject. The gaps in at least our gross factual information are currently being removed rather rapidly, with the consequence that questions regarding the proper interpretation of the data begin to be the major issues. We might, with benefit in this connection, digress for a moment and compare the development of meteorology with that of another physical science, namely astronomy. No one can dispute the claims that Galileo and Newton created a new and orderly conception of the solar system. This achievement was necessarily preceded not only by the laborious accumulation of observational knowledge by Tycho Brahe and others but also by the proper interpretation of these data by Kepler who, it might be said, defined in precise terms the dynamic puzzle to be resolved. In meteorology we find ourselves in what might be called

the Keplerian era. It behooves us to marshal and correlate our observations into as precise and consistent a scheme as possible in order that we may know in sufficient detail what is to be explained.

From what has been said it might seem to the reader that meteorology must still undergo a rather protracted period of development before results can be expected at the final fruition of this process. Although there is reason to expect this pattern of events in the philosophical aspects of the subject, it should not be forgotten that the main practical use of meteorological knowledge, the preparation of weather forecasts, is at present largely an empirical procedure. As such, every improvement in our empirical information about the atmosphere enhances in some degree the possibility of improved forecasts, even though satisfactory understanding still remains to be achieved. Here only the intellectual aspects of problems are touched upon, leaving any possible practical applications for treatment elsewhere.

Following the general plan implied in what has been said, let us begin by examining how the general circulation, as we observe it, *achieves internal dynamic consistency in several important respects.* As will be seen, this is merely the extraction from data and interpretation of certain information, and does not in any sense constitute an *explanation* of why the facts are as they are found. In order to concentrate attention on the most basic processes, let us first consider the mean state, leaving the temporal fluctuations in the general circulation as a problem of much greater difficulty to be touched upon later.

If an observer equipped with suitable instruments were to measure the motions of the atmosphere from some extraterrestrial vantage point, in the same manner as we measure the motions in the sun, he would probably be impressed by the irregularities of the details, but at the same time he would discern that there is a pronounced general drift of the air from west to east relative to the earth in middle latitudes, extending from the surface to the stratosphere and even beyond. On the other hand, in the more equatorial regions (and at times near the poles, at least the North Pole) he would discern a drift from east to west from the surface to great heights. This situation immediately poses perhaps the most important problem concerning the general circulation. The specific question involved here is how the belts of westerlies can maintain their high rate of rotation in the face of the retarding effect of surface friction, flanked as they are by oppositely directed winds on at least their equatorial sides. No really satisfactory rational theory for this state of affairs has yet been given, although some deductions can easily be made concerning the nature of certain aspects of the mechanism which is necessary to maintain these existing motions.

The retarding effect of the surface frictional forces on the middle-latitude westerlies may be looked upon as a continuous abstraction of absolute angular momentum from the atmosphere in these regions. According to simple principles of Newtonian mechanics, this drain can be compensated only by an equivalent flow of angular momentum into the westerly belts. Likewise the surface frictional effect in the regions of the easterlies may be interpreted as a flow of angular momentum from the earth into the atmosphere. In order that the angular momentum so transferred into the easterly regions should not progressively accumulate and destroy these wind systems, it is necessary that an equivalent flow out of these regions should exist.

The inescapable conclusion is that the accounts are balanced by a flow of angular momentum from the easterlies in the more tropical regions poleward to the westerly belts (the polar easterlies are of relatively small importance in this connection). Such a flow of angular momentum, let us say northward at the northern border of the tropical easterlies, can be measured in terms of an equivalent tangential stress acting across a vertical surface parallel to the latitude circle. A crude estimate of the value of this stress may be made from existing information concerning the surface frictional forces on the easterlies to the south.[1] The result is of the order of 50 to 100 dynes cm^{-2}. Molecular and small-scale eddy viscosity cannot transmit stresses of this magnitude under the existing conditions, so that very large-scale nonzonal components of motion must furnish the necessary eddy-transfer of angular momentum. We thus come to the very pertinent observation due to Jeffreys [5], namely that the large nonzonal components of motion in the atmosphere are necessary for the maintenance of the average zonal components.

Most classical models for the general circulation as well as some more recent ones (see for example Rossby [9]) have followed along lines originally proposed by Hadley [4] in that they assume the existence of large, slow, convectively driven closed circulations in meridional planes. The development of the mean zonal motions is then ascribed to the effect of the earth's rotation on these primary circulations. According to this view the necessary transport of angular momentum could be achieved if, for example, the poleward branches of the meridional circulations carry more angular momentum than the returning ones at other levels.

For several reasons many modern meteorologists have come to view models of the Hadley type with skepticism. In the first place, the warmest regions of the atmosphere are not usually found in the tropics as most schemes of this kind visualize, but rather some distance away from the equator. Also, at best it is difficult to account for the great extent of the westerlies in the atmosphere on any such basis. Finally, there is a suggestion in the climatological distribution of precipitation and in the poleward flow of air in the friction layer for the existence of a slow meridional circulation with an upward branch in middle latitudes and a downward branch toward the subtropics, as originally suggested by Bergeron [2]. Such a "reverse" cell with equatorward flow aloft would, due to the action of Coriolis forces, tend to establish east winds in the

1. It is here assumed that the main transfer of momentum takes place in the troposphere.

region where the strongest westerlies are actually encountered. The writer rather inclines to the view that although very small mean meridional circulations do perhaps exist, their role in the horizontal transport of angular momentum, at least in the middle latitudes, is overshadowed by the characteristics of other horizontal motions. This is not in conflict with the views expressed by Jeffreys [5] and is reinforced by analysis of data to be quoted presently.

If emphasis is placed upon the horizontal circulations in transporting angular momentum poleward, by and large the zonal and meridional components of velocity should be correlated at each level where such transport occurs. Evidences of this correlation are then to be expected in the detailed structure of the instantaneous horizontal streamline patterns observed, as pointed out by the writer elsewhere [10]. Qualitatively these expectations are amply borne out even by casual inspection of weather maps. Except at high latitudes, closed horizontal circulations usually exhibit a northeast-southwest elongation (Northern Hemisphere) and the troughs, ridges, and shear lines show a tendency to tilt in this same sense. All these characteristics are recognized almost instinctively by the meteorologist as typical of atmospheric flow patterns. From our viewpoint they are telltale indications of a poleward flow of angular momentum. One may nevertheless ask whether an objective quantitative evaluation of the transport by this mechanism can be made from actual hemispheric data. For it is not sufficient to advance merely a qualitative substitute for the classical hypothesis without investigating the potency of the new alternative to produce the needed effect. A question involved here is whether the observations we possess are extensive enough and of sufficient accuracy. It is a simple matter to set up an integral expression for this (say) northward flow of absolute angular momentum across the vertical surface at a given latitude after the manner of Jeffreys, or to derive it from the atmospheric equations of motion as has been done by Widger [17]. The latter author proceeded to evaluate this flow of angular momentum by finite difference methods, as follows.

During the last several years sufficient observations of the free atmosphere have been made to allow the construction of daily isobaric charts through most of the troposphere. In addition it is becoming possible to obtain fairly complete direct radiowind observations extending to great heights on a circumpolar basis within restricted latitude belts. The global wind distributions may be approximated from isobaric charts according to the geostrophic wind formula or may be taken directly from actual wind observations. The use of the geostrophic estimates may introduce certain errors for the present purpose. On the other hand this use of the geostrophic winds automatically excludes the contributions to the angular momentum transport due to mean meridional circulations of the Hadley type, so that an advantage is gained if it is desired to study other modes of such transport. Several surveys of the angular momentum balance have been made in the past few years (Widger [17], Mintz,[2] Starr and White [12]) both from isobaric charts making use of the geostrophic approximation, and directly from a circumpolar network of actual wind observations. The survey of the angular momentum balance made by Widger covers the month of January 1946 and the results give the total geostrophic transfer by latitudes for various layers during this period up to the 7.5 km level. In this study estimates were made of the surface frictional torques. The investigation by Mintz covers the month of January 1949 and his results give the geostrophic flux of angular momentum at various levels up to 100 mb. Starr and White made use of a circumpolar network of actual wind observations at a mean latitude of 31°N for a period of six months from February 1949 to August 1949 up to an elevation of 50,000 ft. For various details of these investigations reference must be made to the original papers.

The computations give results which are entirely reasonable, being quite in accord with what would be expected on the basis of the foregoing discussion. Thus the total northward transport increases in magnitude from low latitudes to about 30°N as the surface easterlies are passed, then decreases progressively northward as angular momentum is removed by surface frictional torques acting in the westerly belt. Nearer to the pole the transport is reversed, indicating a flow southward from the polar easterly zone, although the magnitudes involved here are small as is to be expected from the fact that the torque arm associated with the surface frictional forces is small in the polar regions.

The main transfer across 30°N increases in intensity with elevation, reaching a pronounced maximum at about the level of the jet stream. On the basis of estimates of the surface torques during these periods it appears that sufficient angular momentum is transported into the belt of westerlies to maintain them against friction. Let it be noted, however, that the general results appear to be in harmony with the thesis that *practically all the necessary horizontal transfer of angular momentum could be accomplished without recourse to the agency of mean meridional circulations.* On the basis of what has been said, however, we cannot make any statement concerning other possible functions of meridional circulations such as, for example, the vertical transport of angular momentum.

At this point let us pause in order to take stock of what has been described and to see more clearly how it fits into the plan for research advanced in the introductory paragraphs. Have we by this study of angular-momentum considerations provided a theory for or achieved a rational solution for the problem of the distribution of zonal westerlies and easterlies in the atmosphere? Not by a long way. We did not solve the equations of motion[3] nor did we deal with radiative

2. "The Geostrophic Meridional Flux of Angular Momentum for the Month of January 1949." Presented at the 109th national meeting of the American Meteorological Society, Jan. 29–Feb. 1, 1951, New York.

3. Actually only one equation of motion, namely the one for the zonal direction, is needed to treat the balance of absolute

processes, which must ultimately be responsible for all air motions and are a determining factor for the form of the general circulation. What we have done is simply make a systematic analysis of data to portray as clearly as possible one important, but not patently apparent, attribute of the general circulation—namely, the angular-momentum balance. This study attempts to show how the actual atmosphere attains an internal consistency in one important respect.

What purpose can such information as this serve in the future of meteorological theory? The answer to this query is obvious. Once the announcement was made by Kepler that planets move in ellipses, any proposed dynamic theory of mechanics which would require them to move in significantly different paths was immediately pushed into the limbo. Likewise for the atmosphere any proposed rational theory for the general circulation which significantly contradicts the general outlines of the needed angular-momentum balance at once finds itself afflicted with a serious and perhaps even crucial drawback. On the other hand, the uses of such information need not be purely negative. The facts might well serve as one guide (among many others) for the formulation of satisfactory rational schemes.

What has been described is not a very profound concept, but merely a rather simple consequence of Newton's laws of motion applied to the atmosphere. The gist of the idea was presented many years ago, by Jeffreys in 1926. How is it then that it was permitted to lie fallow for so long a time? Aside from the lack of proper data, a contributing factor has doubtless been the preoccupation of research meteorologists with phenomena on a relatively local scale, this in turn being due to the limitation of our mental horizons to a scale commensurate with the daily weather map.

One might well venture the guess that at present we have not attained even a measure of the tasks which confront us. And this statement is in no way intended as a repetition of a common platitude. We have not fathomed the profundity of our subject, let alone solved the fundamental problems. In the years to come, when our knowledge will have increased, we may in retrospect wonder why we were so inappreciative of various gross and essential attributes of the general circulation. It would therefore be wrong and dangerous to post a sign along any plausible path of research which states "Thou shalt not enter here," seeking thereby to channel activity along certain predetermined lines. True science tolerates no prohibitions, and in principle no servile adherence to tradition. Rather we must open up and exploit all possible new channels that appear reasonable in order that we do not miss important facts.[4] It is in this spirit that the writer would recommend the further exploration of the essential properties of the global circulation *as we know it from observations* with regard to such dynamic quantities as angular momentum, vorticity, kinetic energy, heat energy, geopotential energy, and latent heat, so that we may find ourselves in a more advantageous position in formulating rational hypotheses.

For the purpose of exemplifying further the suggestions previously made, let us consider in a general way their application to questions concerning energy in the atmosphere. Much of what has been written concerning this subject has been deficient in two respects. In the first place investigators have been prone to lump together various diverse forms of energy indiscriminately, thereby losing the advantages to be gained from the fact that each form of energy is produced from and converted to other forms in its own characteristic fashion, permitting individual study. Likewise for each form there exist specific modes of transfer and redistribution. Unless these specific characteristics are subject to scrutiny in detail, only very broad generalizations can be reached, which lack a keen enough resolving power to yield much that is new concerning the mechanism of the atmosphere.

In the second place, as will be discussed presently, the modes of energy transfer within the atmosphere are so effective that no feature such as a cyclone can be treated independently without due allowance for exchanges of energy between it and the remaining atmosphere. It is therefore inappropriate to treat such a feature as in any sense a closed system. Modern trends in the literature are beginning to give proper cognizance to this circumstance, although much of the too-restricted point of view is still prevalent. Probably the only system which can be treated as a closed one (mechanically) is the atmosphere as a whole and even then it cannot be treated as closed from a thermodynamic viewpoint because of radiative exchanges of energy with the cosmic environment and otherwise.

Proceeding now to a discussion of the kinetic energy balance of the atmosphere, it is worthy of note in the first place to observe that whereas angular momentum is a quantity which is conserved in the sense that no real sources of it can exist, kinetic energy on the other hand may appear or disappear because of transformation processes involving other forms of energy. In other words, the flow and redistribution processes for kinetic energy are not source-free as is the case with absolute angular momentum. We may therefore say a priori that there is no reason to suppose that there may not exist in the atmosphere preferred source-regions for

angular momentum about the earth's polar axis. In effect this equation takes on the form of a continuity equation for absolute angular momentum.

4. When one compares the general nature of the results of research in meteorology as reported in our journals with the nature of research communications in let us say chemistry, there is a fundamental contrast to be noted. In our field there is a great diversity of views expressed on individual topics. Often these views are in direct conflict. Also one gains the impression that there is less of the orderly progressive accumulation of knowledge. This phenomenon is doubtless due to the fact that meteorology is in a far different stage of development as a science. In the better sense of the word we are alchemists still in the process of groping for a common denominator of unifying principles. We must not relax but rather continue to stumble and grope with more determination. The writer has the immutable belief that a proper foundation *does* exist and *will* be found.

atmospheric kinetic energy as well as preferred regions for its disappearance by conversion into other forms of energy through friction or otherwise.

With the aid of the examination and interpretation of the fundamental dynamical principles relating to kinetic energy given elsewhere in this volume[5] we shall now attempt to discuss this phase of the energy problem. In the reference mentioned there is presented a discussion of the process whereby kinetic energy is produced in the atmosphere. According to the views expressed there, kinetic energy of large-scale motions can be generated only through the work of expansion by pressure forces. Furthermore, kinetic energy associated with *horizontal* motions can be generated only through work done by *horizontal* pressure forces. It is thus pointed out that the rate at which horizontal kinetic energy is generated at a given point in the atmosphere is equal to the product of the pressure into the horizontal divergence of velocity. This carries the implication that regions of horizontal convergence are hydrodynamic sinks for kinetic energy which act independently of friction.

Speaking next of the transport of horizontal kinetic energy, it has been shown that this process is accomplished either through the work done by the pressure forces in virtue of the horizontal velocities across the boundary or by the advection of sensible kinetic energy. Since the latter (meridional) transport is small, it follows that the significant (meridional) transport of kinetic energy is accomplished through the work done by the pressure forces, which in turn is *proportional to the advection of internal heat energy* and occurs in addition to it. Applying these ideas to the transfer of kinetic energy across the vertical boundaries of an equatorial strip between two fixed middle latitudes $+\phi$ and $-\phi$, our general information would lead us to suppose that there exists a net transport of internal heat energy and therefore of kinetic energy poleward across these boundaries. Estimates from data tend to confirm this supposition, and indeed it also appears reasonable from common synoptic experience. We are therefore confronted by the very important conclusion that regardless of the nature of other details of the atmospheric circulations the more tropical regions appear to be preferred regions for kinetic-energy generation in that they not only produce sufficient kinetic energy to overcome frictional losses within these regions, but also provide an excess which is transferred to the polar caps.

Apparently we may also conclude that in the long run the more polar regions must serve as preferred regions for the disappearance of kinetic energy, since here the losses must be sufficiently great not only to offset whatever amounts of kinetic energy are generated locally but also to absorb the kinetic energy which is supplied from the more tropical regions. It is therefore reasonable to suppose that in the normal state of affairs in the atmosphere there are vast amounts of kinetic energy generated in the more tropical regions and vast amounts consumed in the polar caps. The existing kinetic energy is merely a small difference due to the fact that the positive generation process in the more tropical regions is slightly larger than the losses in regions of negative generation in the polar caps.

It is a matter of interest to examine the import of the views expressed above for the nature of the secondary circulations of the middle latitudes. It would thus be suggested that the kinetic energy supply for the main cyclone belt of each hemisphere is perhaps provided by the poleward flow of such energy from more tropical regions. Furthermore this flow is measured by (but exists in addition to) the poleward flow of internal heat energy. Although this flux of energy has some average mean value, there can be no doubt that in its details the process is a sporadic one with irregularities both in time and in longitude. Each cyclone may thus be considered as producing a spurt or poleward surge in this transport locally during its period of development.

According to the classical view, the kinetic energy for a developing cyclone is obtained from the sinking of dense masses of cold air in the immediate vicinity, so that by and large the decrease in potential and internal energy represents a corresponding increase in kinetic energy. From the present viewpoint the kinetic energy increase could equally well be derived from a pronounced local poleward flow from the more tropical regions, since a developing cyclone is accompanied by a marked increase in the net local poleward transfer of internal heat energy.

As a matter of fact it is not too difficult to visualize the general outlines of a certain type of instability associated with a developing cyclone when viewed in this way. For if it is granted that there exists a supply of warm and cold air in the vicinity, it may be that the pronounced local surge of kinetic energy once started serves to increase the intensity of the cyclone which in turn further increases the flow of kinetic energy and the intensity of the cyclone. The intensification finally ceases when complete occlusion takes place and the net heat transport disappears.

In a recent discussion the writer [11] has made an endeavor to elaborate further this picture of the kinetic energy balance of the atmosphere, making use of the assumption that mean meridional circulations do not play a significant role in these processes. The plausibility of this supposition is supported by the study of the angular momentum balance outlined earlier. Under such circumstances, since data show that near the surface poleward-moving individual air masses possess a higher specific volume than the equatorward-moving ones, it follows that there should exist a *mean horizontal velocity divergence* in the more tropical regions and a *mean horizontal velocity convergence* in the polar caps. This situation should thus contribute to a net production of kinetic energy, in view of the normal mean meridional distribution of pressure along horizontal

5. Consult "Applications of Energy Principles to the General Circulation" by Victor P. Starr, pp. 568–574.

surfaces.[6] Such estimates as can be made indicate that this gross average action may be of very great effectiveness.

In view of the fact that the correlation of density with meridional motion is most marked near the surface, it follows that the mean meridional distribution of this divergence and convergence shows greatest contrasts at low levels. This at once suggests that the mean meridional distribution of *net external heating and cooling* of the atmosphere is linked with the process, since the heating of the atmosphere is most intense near the surface and much of the cooling is accomplished through surface effects also. Thus it would be logical to suppose that the net external heating in the more equatorial latitudes causes a net horizontal expansion, while the net cooling in the polar caps brings about a net horizontal contraction, accounts then being balanced, as far as mass continuity is concerned, by differential advection across middle latitudes.

If one agrees to argue on this basis, a far-reaching observation may be made, subject of course to the correctness of the premises involved. Earlier in the discourse it was pointed out that one of the most fundamental questions in meteorology relates to the presence of broad belts of westerlies in the middle latitudes of each hemisphere. Speaking now only of the lower levels we see that the presence of the westerlies coincides with the requirement that the regions of net external heating and divergence should coincide with regions of higher pressure along horizontal surfaces. Unless this situation exists a net production of kinetic energy to overcome friction would be impossible. Thus, if for a moment we were to visualize a hypothetical situation with mean easterlies in middle latitudes, the net heating and consequent divergence would take place at a relatively low pressure in the more tropical regions, while the convergence and cooling in the polar caps would take place at a relatively high pressure. Such a hypothetical situation would then be rapidly destroyed, since more kinetic energy would necessarily disappear in the polar caps through convergence than could be generated in the more tropical regions by heating and divergence.

Although the net heating and cooling of the earth is ultimately caused by radiational exchanges with space, when one isolates the atmosphere and considers the circumstances in the winter season especially, due regard has to be given to the strong heating of the atmosphere by the oceans when cold air masses flow out over relatively warm water surfaces. As has been pointed out by Sverdrup and others [13], the oceans serve in a manner of speaking as a hot-water heating system. It should thus be expected that during winter the regions of net heating for the atmosphere extend much farther into middle latitudes than might otherwise be supposed.

6. Since variations of pressure along horizontal surfaces are necessary for the generation of a net amount of kinetic energy to overcome friction, it is indeed reasonable on general grounds to suppose that the atmosphere behaves in such a way as to take advantage of the large systematic pressure differences between the polar and tropical regions.

Since we have been speaking of the meridional transfer of internal heat energy, it is desirable to mention also the effects of the meridional transfer of latent heat. Crude estimates made by the writer using such data as are available indicate that very considerable amounts of energy are transferred poleward by this means. However, whereas the transport of internal heat energy cannot take place without a proportional transport of kinetic energy, the meridional transfer of latent heat energy is not related to the kinetic energy in the same way. It is thus possible to transfer large amounts of energy in the form of latent heat without a proportional kinetic energy flow being present. Whereas the flow of kinetic energy and therefore of internal heat energy may be regulated by the distribution of pressure and of divergence and convergence, this particular limitation does not enter as far as latent heat is concerned. One could therefore find arguments to support the view that the transport of latent heat furnishes a means for balancing the radiation requirements of the general circulation which bypasses the kinetic energy balance. For if the portion of the total meridional energy flow represented by the transfer of latent heat energy had instead to be transported as additional internal heat energy, a far more intense flow of kinetic energy would also be necessary. Whether this would imply a more "vigorous" general circulation is, however, a question which cannot be readily dealt with.

The meridional transport of geopotential energy is of course immediately ruled out in the average condition discussed here, because of the assumption that mean meridional circulations are not significant.

In the discussion mentioned above [11], an endeavor was made to study the fluctuations in the kinetic energy balance of the Northern Hemisphere with the aid of 5-day mean data such as are used by the U. S. Weather Bureau in extended-period forecasting. Only data for the colder half of each of seven successive years were used, so that essentially winter conditions were studied. By approximate methods a measure was secured of the intensity of differential advection of air with varying density across 45°N. This quantity τ is therefore a measure of the net volume transport northward across 45°N and hence also a measure of the mean convergence in the polar cap and to some degree probably a measure of the mean divergence in the more tropical zones of the Northern Hemisphere. A mean for the layer between sea level and 10,000 ft was used.

It became immediately apparent that the quantity τ undergoes large fluctuations, being high during those periods which are commonly designated as "low-index" periods and low during "high-index" conditions. The practical question involved here concerns itself with the cause of these vagaries in the behavior of the general circulation, for if this were known a better insight into the long-range forecasting problem might result.

In view of the fact that τ is also a rough measure of the rate of kinetic energy transport into the polar cap, it is interesting to note its relationship to the general pressure distribution there. If these pressures are low compared to the pressures farther south, a relatively

large fraction of the kinetic energy so transferred should remain without being destroyed by the mean convergence, while if these pressures are high practically all the kinetic energy fed into the cap should disappear because of the mean convergence. For this purpose it was decided to examine τ with reference to the *area-mean pressure deficit* north of 45°N as compared to the pressure at 45°N. This pressure deficit was denoted by π and was computed from sea-level pressures. It became apparent that abnormally large values of τ are, generally speaking, encountered only when the pressure deficit is small or negative, in other words when the capacity of the polar cap to destroy kinetic energy is abnormally great.

From these general observations it would appear that large meridional transports of internal heat energy and of kinetic energy take place during those periods when the mean meridional pressure distribution at lower levels is relatively flat north of the subtropics with no marked polar low present. How can a situation such as this maintain itself over appreciable intervals of time, as is known to be the case? Is it a quasi-steady state or is it a dislocation which becomes progressively more difficult to maintain? We can only speculate about the answers, but it is rather interesting to do so.

First let us consider the fact that it is during these low-index conditions that very cold air masses are injected into southerly latitudes. This in turn means a very strong heating of the atmosphere in middle latitudes and the subtropics, especially over ocean areas, so that a rapid replenishment of heat and also a rapid generation of kinetic energy would be a normal accompaniment of this state of affairs. This consideration would argue for the possibility of quasi-steady state.

On the other hand in order to maintain a quasi-steady condition, means would have to be present to dispose of large quantities of heat in the polar cap. It is unlikely that the net heat loss through radiation could be stepped up enough to meet this requirement. Progressive melting of ice in the polar regions could absorb large quantities of heat, but probably exactly the opposite is actually the case, since very low temperatures are apt to prevail at low levels in the arctic and subarctic regions under these circumstances.

Observationally there is some qualitative evidence that during prolonged low-index conditions there is apt to be present in polar regions an accumulation of abnormally warm air a little distance above the surface in the troposphere and in the stratosphere. Also there is some evidence that there is actually a progressive depletion of heat energy from middle latitudes (in spite of the strong surface heating) with a consequent progressive southward shift of the latitude of the maximum westerlies at higher levels. What ultimately puts a stop to the trend of developments is not obvious.

Contrasted with the low-index regime, during periods of a strong polar vortex at low levels the heat and kinetic energy transport is weak. Relatively warm air streams across the continents and oceans in a more zonal manner so that no very vigorous heating takes place. So far no difficulties seem to be present as far as the more southerly regions are concerned. It would appear, however, that progressive cooling should now take place in the arctic regions since the abnormally low heat input would appear to be insufficient to maintain the *status quo*.

On the whole there may be good and sufficient physical reasons, such as those mentioned above, why the general circulation does not persist indefinitely in one or the other abnormal conditions but rather tends to shift about a general average state. There is nevertheless the important question why the average state does not persist without such pronounced departures as those just described. The writer's conjecture on the matter is that the average state can too easily be disrupted by the effects of the nonzonal continentality found especially in the Northern Hemisphere. Such effects may be partly due to the fact that the continents provide avenues for the southward penetration of intensely cold air masses which can thus arrive at relatively low latitudes without undue modification. The occasional outpourings of such cold masses over adjacent water areas at fairly low latitudes could easily result in abnormal heating of the atmosphere sufficient to disrupt the average regime.

Also it should not be forgotten that large mountain ranges extending over an appreciable spread of latitude can exert powerful mechanical effects upon the atmosphere, as pointed out by White [16]. By way of example large pressure differences across the Rockies in North America could perhaps disrupt the angular momentum balance in the belt of the prevailing westerlies.

Finally the suggestion made by Willett [18] that the seat of the variation could perhaps be found in the variations in solar output, and hence the heating of the atmosphere, deserves study. For it may be that the final answer is bound up in a combination of effects, especially if long-term aberrations of the general circulation during historic and geological periods are also considered.

From what has been said in the several preceding paragraphs it might appear that too little emphasis has been placed upon the influence of the conditions in the upper troposphere and in the stratosphere. In the final analysis it is obvious that the entire atmosphere must form an integrated system. It would seem, however, that the lower layers furnish the seat of thermodynamic activity and provide the motive force for the general circulation. The writer's colleague, Dr. H. L. Kuo, is currently engaged in a theoretical study of the effects which might result at the level of the jet stream from forced disturbances originating in the lower levels. From his preliminary work it would appear that the generation and maintenance of the jet stream itself are a necessary consequence of such impulses received from below. Likewise the characteristics of the angular momentum balance as outlined previously, including such features as the tilted (and properly curved) trough and ridge lines, as well as the development of regions of sharp shear to the north and to the south of the jet follow as consequences of the analysis.

Having made an excursion into some of the ramifi-

cations involved in the study of the balance of angular momentum and of kinetic energy in the atmosphere, let us now consider the general circulation from the standpoint of a *rational theory* for atmospheric motions. By this term we mean essentially a *solution of the equations of motion* which purports to give a picture of the circulation or some portion of it. This is sharply contrasted with the study of the balance of various quantities, because the latter involves hydrodynamical principles only to the extent of formulating several integral requirements for the atmosphere. The investigation of how the atmosphere actually fulfills these requirements is then necessarily an empirical and observational problem.

Logically, then, the empirical picture obtained from the study of the balance of various quantities and from other observational studies should give us the fundamental physical characterization of the atmosphere which is then to be explained in terms of some rational theory. But, in a larger sense, the proper physical characterization of the general circulation is a subject which up to the present has hardly been entered upon. We need quantitative estimates of the flow of heat, of kinetic energy, of momentum, etc., for the mean state and also when consideration is given to synoptic and seasonal variations. The labor involved in order to supply this information is bound to be tremendous, involving the cooperation of meteorological services over the entire globe. Nevertheless progress is being made and there is reason for optimism.

In view of this state of affairs it is not surprising that the development of a rational theory for the general circulation has not made very much headway up to the present time. Thus the scope of the more successful efforts toward the solution of the hydrodynamic equations has necessarily been severely limited to certain portions of the atmospheric circulation where the introduction of drastic simplifications still leads to results of interest. As an example we might cite the several solutions of the two-dimensional barotropic vorticity equation given by Rossby [8] and recently elaborated by Charney and Eliassen [3] and others. For short-period extrapolation it appears that these results may prove to be of more and more value in forecasting, even though these solutions leave unanswered some of the fundamental questions concerning the energy balance when long-term effects are contemplated.

As a general comment it can be said that really adequate solutions which link the mechanics of the atmospheric circulation to the source of energy from solar radiation are a desideratum for future research, although the situation does not warrant undue pessimism. In this connection reference may be made to a recent publication by Lorenz [6] which is concerned with the solution of the two-dimensional equations of motion for the region near the pole, taking into account friction and external heating and cooling.

Leaving now the discussion of illustrative topics, what are the avenues for future progress which are indicated as of today? This is necessarily a subjective matter in which various investigators must follow their own inclinations and tastes, for there is no shortage of diverse problems which are worth while. The inclinations of the writer, as one of these investigators, are probably quite apparent from what has already been said. The underlying and oft-recurring motif of this essay is the need for a more complete and systematic empirical description of the basic physical processes involved in the general circulation. We are now beginning to have sufficient observational material at least in the Northern Hemisphere.

Before making concrete proposals, let us take a glance at what is being done which may serve as a beginning toward this aim. In the enumeration of activities which follows, any omission of important work is inadvertent and is due simply to the writer's lack of information. At the Massachusetts Institute of Technology Willett [18] has for a period of years gathered statistical information concerning the general circulation in the Northern Hemisphere in the form of 5-day averages. More recently he has begun gathering similar data for the Southern Hemisphere. The writer has found this material invaluable for the study of the general circulation. The Department of Meteorology of the University of California at Los Angeles has undertaken a study of the angular momentum balance of the atmosphere by means of finite difference integration procedures. This project will probably also include a similar study of the energy balance later.

Priestley [7] of Australia has proposed a program for the observational study of the momentum and energy balance of the atmosphere, utilizing radiosonde and radio-wind observations directly. Samples of his analyses already prepared by him have proven to be of great interest, but require elaboration.

Van Mieghem [15] of Belgium has proposed the study of the balance of various forms of energy, entropy, and momentum, emphasizing the importance of such studies for the progress of research concerning the general circulation.

The Extended Forecast Division of the U. S. Weather Bureau has for some years been amassing observational material in the form of 5-day averages for the Northern Hemisphere. The Weather Bureau [14] has recently also been preparing in published form very complete daily Northern Hemisphere maps for the surface and 500 millibars. These charts are most excellent and should prove to be invaluable for research purposes.

Finally mention may again be made of the studies of the angular momentum and energy balance initiated by the writer at the Massachusetts Institute of Technology.

The items listed above do not purport to cover all important research on the general circulation conducted currently, but rather only such activities as are apt to furnish statistical information concerning the gross character of the general circulation. Thus various other synoptic and theoretical activities are not included even though they may in the end furnish indispensable insight as to the interpretation of the empirical picture given by the statistics.

The general trend to be discerned is that the em-

pirical study of the general circulation is becoming a matter involving vast amounts of data properly and purposefully organized in order to be useful for the systematic evaluation of various specific and well-defined processes over long periods of time. It is being undertaken piecemeal by various institutions to an extent determined by their financial resources and facilities. In the initial stages of such exploratory activities it is probably most desirable that there should exist diversified direction of individual small-scale projects of this nature. However, once the best general patterns for such research have become apparent, a number of considerations point to the desirability of more consolidated effort. Thus some central organization could perhaps offer the advantages of a long-term program, more economical operation, better facilities such as high-speed computing equipment and better availability of computed results. Such an institute could even be of an international character under the auspices of the United Nations, since the purpose would indeed be of global importance in a very literal sense.

Beyond the broad recommendations just set forth, the writer wishes to state his belief that periodic reviews of progress such as the ones contained in this volume should be planned by appropriate organizations in an effort to stimulate comparison of differing opinions and to bring forth suggestions. Another matter of somewhat similar nature is the periodic publication of selected reprinted papers, adjudged to be of fundamental importance, in the form of collections similar to those edited by Cleveland Abbé [1] several decades ago. A practical problem here would be the selection of a capable (and so to speak disinterested) editor.

Finally a word about certain philosophical and theoretical aspects. In a science such as meteorology in which the rational explanation of the most gross features is still to be found, much attention must be given to questions regarding the general techniques for *interpretive* research in order to ascertain if possible the ones which are suitable. In any well-developed science, as for instance physics, it is often possible to arrive at new results by purely deductive means. However, it must be granted that the more fundamental scientific achievements of an interpretative nature have usually stemmed from *the deliberate introduction of new physical hypotheses*. Such hypotheses are usually the direct product of creative minds who through sufficient insight are able to recast a given problem into a new form. Although such formulation of novel hypotheses is perhaps the most difficult task confronting a scientist, we cannot afford to dispense with it in any attempt at understanding the motions of the atmosphere at the present time.

Without the use of creative imagination to give it form, our science could easily become a repetitious accumulation of bits of petty dogma, a jargon of catchwords and sophistries, a species of scholasticism lacking a firm basis or unifying principles. We need men with vision and courage who would not be content in occupying themselves with odd bits of research, but would tackle the fundamental problems. Men who have something of the spirit of daring which prompted Newton to propose the theory of universal gravitation and Einstein to suggest anew the quantum character of radiation.

Much of what has been said in the last two paragraphs has a direct bearing upon the use of mathematics in meteorological research. To regard the fundamental problems of meteorology as purely mathematical ones is hardly warranted. It is true that many of the hydrodynamical and physical principles involved are capable of mathematical statement, but it is only through the leaven of some purely physical hypothesis that we are guided to the appropriate mathematical use of these principles.

As stated previously, present-day meteorology may be thought of as being in the Keplerian era. It finds itself, however, side by side with very much more advanced physical sciences in which research is carried on with the aid of very sophisticated mathematical and theoretical tools designed for the purpose through the laborious efforts of past generations. Although much may be gained by borrowing certain of these techniques where proper analogies have been established, still there are certain philosophical pitfalls which arise when this is done merely in slavish effort at imitation without proper caution. Such superficial efforts are sometimes entered upon in order to gloss over the prerequisite of clear physical thinking. We cannot telescope the evolution of a new science by omitting essential phases of its natural development. We cannot hope for a magic carpet which would carry us directly from Kepler to Einstein, eliminating the growing pains of Newtonian mechanics.

REFERENCES

1. Abbé, C., "The Mechanics of the Earth's Atmosphere. A Collection of Translations by Cleveland Abbé." Third Collection. *Smithson. misc. Coll.*, Vol. 51, No. 4 (1910).
2. Bergeron, T., "Über die dreidimensional verknüpfende Wetteranalyse. I." *Geofys. Publ.*, Vol. 5, No. 6 (1928).
3. Charney, J., and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
4. Hadley, G., "Concerning the Cause of the General Trade Winds." *Phil. Trans. roy. Soc. London*, 39: 58 (1735–36). (Reprinted in [1, pp. 5–7].)
5. Jeffreys, H., "On the Dynamics of Geostrophic Winds." *Quart. J. R. meteor. Soc.*, 52: 85–104 (1926).
6. Lorenz, E. N., "Dynamic Models Illustrating the Energy Balance of the Atmosphere." *J. Meteor.*, 7: 30–38 (1950).
7. Priestley, C. H. B., "Heat Transport and Zonal Stress between Latitudes." *Quart. J. R. meteor. Soc.*, 75: 28–40 (1949).
8. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2: 38–55 (1939).
9. Rossby, C.-G., "The Scientific Basis of Modern Meteorology" in *Climate and Man: Yearbook of Agriculture*. Washington, D. C., U. S. Govt. Printing Office, 1941. (See pp. 599–655.)

10. Starr, V. P., "An Essay on the General Circulation of the Earth's Atmosphere." *J. Meteor.*, 5: 39–43 (1948).
11. —— *A Physical Characterization of the General Circulation.* Dept. of Meteorology, Mass. Inst. Tech., Rep. No. 1, General Circulation Project No. AF 19–122–153, Geophys. Res. Lab., Cambridge, Mass., 1949.
12. —— and White, R. M., "A Hemispherical Study of the Atmospheric Angular-Momentum Balance." *Quart. J. R. meteor. Soc.*, 77: 215–225 (1951).
13. Sverdrup, H. U., Johnson, M. W., and Fleming, R. H., *The Oceans.* New York, Prentice-Hall, Inc. 1942.
14. U. S. Weather Bureau, *Daily Synoptic Series Weather Maps, Northern Hemisphere, Sea Level and 500 Millibar Charts with Synoptic Data Tabulations.* Washington, D. C., U. S. Department of Commerce.
15. Van Mieghem, J., "Production et redistribution de la quantité de mouvement et de l'energie cinétique dans l'atmosphère. Application à la circulation atmosphérique générale." *J. sci. Météor.*, 1: 53–67 (1949).
16. White, R. M., "The Role of Mountains in the Angular Momentum Balance of the Atmosphere." *J. Meteor.*, 6: 353–355 (1949).
17. Widger, W. K., "A Study of the Flow of Angular Momentum in the Atmosphere." *J. Meteor.*, 6: 291–299 (1949).
18. Willett, H. C., "Patterns of World Weather Changes." *Trans. Amer. geophys. Un.*, 29: 803–809 (1948).

OBSERVATIONAL STUDIES OF GENERAL CIRCULATION PATTERNS

By JEROME NAMIAS and PHILIP F. CLAPP

U. S. Weather Bureau, Washington, D. C.

Any reasonably brief treatment of a subject so vast as "Observational Studies of General Circulation Patterns" is bound to be biased in its selection of material. To be sure, any and all real information derived from the atmosphere constitutes an integral part of the material upon which students of the general circulation must draw. Inasmuch as our knowledge of the general circulation is very far from complete, it is difficult at this time to decide what statistically or observationally determined facts are especially pertinent. In selecting these data the authors confess to a bias in the direction of hemispheric or world-wide phenomena and also to those facts which have emerged from a fifteen-year-old experiment in the anatomy and physiology of the general circulation with which the authors have had the pleasure to be affiliated [33].

NORMAL STATE OF THE GENERAL CIRCULATION OVER THE EARTH

Sources of Data—Historical and Current. Ideally, the data upon which studies of general circulation patterns are based should consist of a series of complete and well-analyzed sea-level and upper-level charts covering the globe or at least a hemisphere for a period of decades. At the present writing this ideal state of affairs is far from being realized. For the Northern Hemisphere the longest series of analyzed charts at sea level embrace a period of about fifty years and at upper levels about fifteen years. Such a short period is woefully inadequate for studies of large-scale weather phenomena, particularly when one considers time intervals of a month or more. Furthermore, many of these charts are incomplete or inaccurate, and the upper-air charts are mainly for one level only (either 700 or 500 mb). In the Southern Hemisphere the situation is materially worse, although it will be partly remedied in a few years after several units, including the Department of Meteorology at the Massachusetts Institute of Technology, have completed files of Southern Hemisphere sea-level charts. The M.I.T. series began with January 1949.

Figures 1 and 2 give some idea of the amount and extent of data currently available for the construction of Northern Hemisphere charts at the Central Office of the U. S. Weather Bureau. This represents a vast improvement over conditions existing ten years ago, largely as a result of the increasing need for data during World War II.

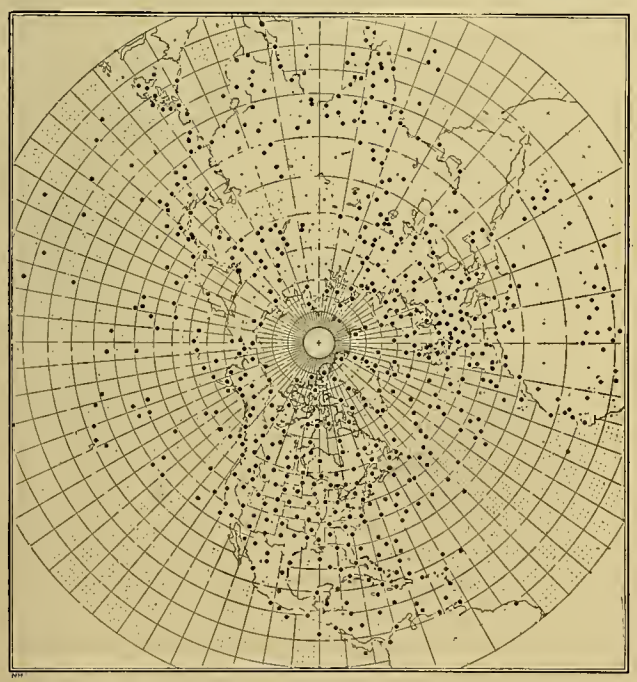


FIG. 1.—Typical data coverage at sea level for the Northern Hemisphere. Solid circles represent data actually received at 1200Z June 25, 1949. Some data have been omitted over the United States and parts of Europe because of space limitations.

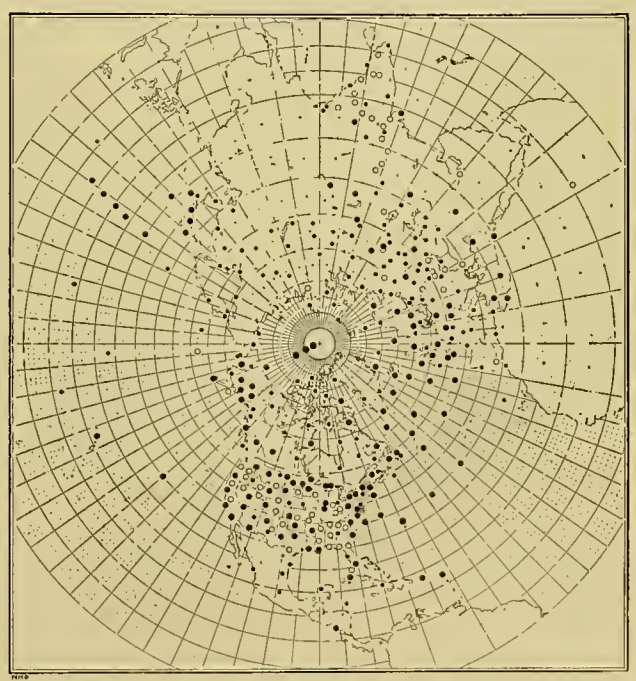


FIG. 2.—Typical data coverage at 500 mb. Small solid circles represent radiosonde data; large open circles, wind data; and large solid circles, both radiosonde and wind data actually received at 0300Z June 24, 1949. Winds are by pilot balloon, radar, or direction finder. The three observations near the pole and the five in the southwest Pacific centered at 20°N, 147°E are aircraft reconnaissance data.

While the coverage shown in these figures is reasonably adequate in heavily populated land areas and in the much-traveled North Atlantic, there are still many important gaps, notably in the southern portions of the North Atlantic and North Pacific Oceans and in many parts of Asia. This situation is due in part to communication difficulties and to the fact that it is more expensive and difficult (or even dangerous) to establish stations in unfrequented and less habitable areas. Nevertheless, political and economic forces are also partly responsible for the tendency of the density of observations to be proportional to the number of inhabitants. Since there is mounting evidence (some of which will be presented in this paper) that weather in any given area of the globe may be importantly influenced by conditions in other areas, no matter how remote, it would appear preferable to reduce the amount of data in populated areas and use the money thus saved to increase the number of reports, particularly upper-air reports, from remote areas. Certainly it is a mistake to reduce the amount of hemispheric data on the grounds that it is no longer necessary for military purposes. For the past two or three years there has been a strong tendency in this direction. The International Meteorological Organization, the International Civil Aviation Organization, the Pacific Science Association, and similar organizations should be commended for and encouraged in their present efforts to improve the data coverage of the world.

A brief list of published hemisphere charts is given in the references [1, 9, 17, 28, 29, 30, 46, 50, 51]. The Air Weather Service and U. S. Weather Bureau Northern Hemisphere sea-level and 500-mb analyses contain, in addition to analyzed charts, a valuable listing of all data received. To date the Air Weather Service charts have been published only from October 1945 to September 1946. The U. S. Weather Bureau commenced analysis and publication of this series beginning with January 1949. Charts for January through November 1949 are currently available. Efforts should be made to close the gap in the historical map series from 1939 to 1945. This includes the years 1944 and 1945 when hemispheric data, particularly in the South Pacific, were more complete than at any other time.

Many of the investigations to be discussed in this article were made using unpublished operational analyses. These include twice-daily sea-level and 10,000-ft (or 700-mb) charts over the Northern Hemisphere from 1940 to date as analyzed at the Extended Forecast Section of the U. S. Weather Bureau. Not until mid-1943 do these charts cover half a hemisphere or more. From them, twice-weekly five-day mean maps and twice-monthly thirty-day charts have been constructed. Five-day mean twice-weekly sea-level and 3-km charts for the cold season October through March 1933 to 1940 have been prepared by the Department of Meteorology at M.I.T. The 40-yr daily historical map series [28] has led to the construction of monthly mean sea-level charts. Corresponding 10,000-ft monthly mean charts have been constructed partly by extrapolation from sea level for the period 1932 to 1939.[1]

Reasons for the Use of Mean Charts. Mean charts, showing the average state of the general circulation for specified periods of time, have been used extensively by many investigators, including the authors, for studying the observed behavior of the general circulation. Here are some of the advantages of mean charts over daily synoptic charts:

Because of the unreliable nature of portions of hemisphere-wide analysis, the use of means leads to a smoothing which tends to eliminate experimental errors. Furthermore, such smoothing also eliminates irregularities in the flow pattern, such as small eddies or minor atmospheric waves. These irregularities, while real enough, tend to confuse the large-scale features of the circulation which are of most importance in long-range forecasting. To a lesser degree this second type of smoothing may be attained by the use of synoptic charts from higher tropospheric layers or by mathematical devices, such as Charney's geostrophic wind approximation [10].

Perhaps the most important justification for the use of means is that they reveal large-scale features having a long period. Indeed the speed with which a given pattern changes appears to vary inversely as the number of days over which the averaging is done. The authors feel that slow evolution of this character holds out the hope for successful long-range weather forecasting.

Granting the correctness of the arguments given above, the question has often been raised as to how mean charts are to be interpreted from a physical point of view. In particular, it has been argued that the equations of motion cannot be applied to a mean flow pattern, for, since these equations are nonlinear, the average value of acceleration will not equal the sum of the corresponding average values of the other terms. The answer to this question cannot yet be given. However, this problem seems to be closely connected with the statistical theory of turbulence [3]. Briefly, this theory states that if a given fluid consists of an average flow upon which are superimposed random eddies, then the equations of motion for laminar flow can be applied, provided certain frictional stress terms are added which express the cross-flow eddy transfer of momentum. There seems to be no obvious reason why it should be necessary to place any arbitrary time limit to the average flow, so that the equations of motion ought to be just as applicable to mean charts of a month's duration as they are to mean charts of a few minutes (synoptic charts)—provided the proper frictional terms are added [20]. This type of reasoning has been used by Defant [16], Lettau [32], and more recently by Elliott and Smith [21] to measure the frictional stress exerted by traveling cyclones and anticyclones. The omission of frictional terms in applying the equations of motion to

1. All of these charts have been microfilmed and copies may be purchased by writing to the Librarian, U. S. Weather Bureau, Washington, D. C.

mean charts may be no more serious than on daily charts. In the following pages it will be seen that mean circulations behave as though they were true physical entities.

Normal Circulation of the Troposphere and Lower Stratosphere. The prevailing motion of the greatest mass of the earth's atmosphere is in general eastward. The most recent normal maps prepared by the U. S. Weather Bureau [30, 50] and by the British Meteorological Office [9] effectively bring to light the existence of a vast circumpolar whirl which, while somewhat obscured at the earth's surface, becomes stronger with elevation until near the base of the stratosphere it assumes the characteristics of a sharp maximum in speed. This narrow band of high speed, called the jet stream, and illustrated by Willett [56] and Hess [24] in north-south mean wintertime sections (not reproduced), seems to be found at most meridians around the Northern Hemisphere. What meager data are available in the Southern Hemisphere (worked up by Flohn [22]) also suggest a jet stream there. A series of unpublished vertical cross sections for each 20° of longitude around the Northern Hemisphere have been prepared at the U. S. Weather Bureau as part of the work in constructing normal monthly maps at various levels up to 19 km [30]. From these cross sections, showing the geostrophic zonal wind component as a function of elevation, the latitude, strength, and horizontal shear across the jet stream have been entered and analyzed around the Northern Hemisphere as shown in Figs. 3 and 4 [38].

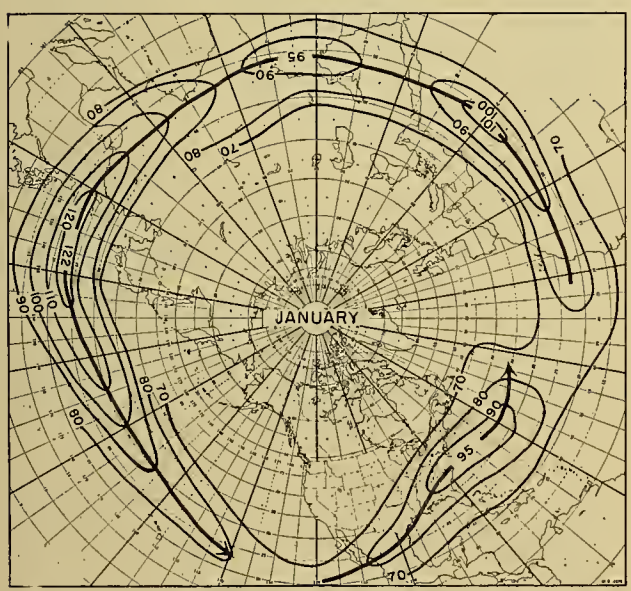


FIG. 3.—Average position and strength of the jet stream for January, prepared from eighteen hemispheric meridional cross sections. Heavy solid lines indicate geostrophic wind speed in miles per hour at the level of maximum speed. Heavy arrows represent axis of the jet. (*After Namias and Clapp* [38].)

The elevation of the jet, not shown on these figures, is everywhere in the range between 11 and 14 km. The material presented in these figures must, of course, be taken with some reservation, and one cannot scale off from them numerical values upon which to draw refined conclusions. Besides, they represent only the zonal component of flow; but this is by far the largest component, and an analysis of the total flow (not shown) gives a quite similar picture of the normal character of the jet stream.

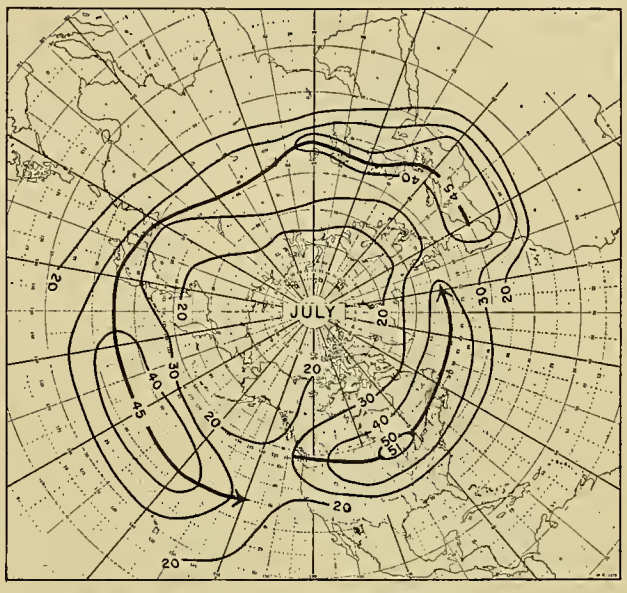


FIG. 4.—Average position and strength of the jet stream for July. (See Fig. 3 for legend.)

The characteristic features of the normal jet stream seem to be:

1. Its nature is essentially circumpolar in both summer and winter.

2. It is considerably farther north and weaker in summer than in winter—from a surprisingly low latitude of about 20°N to 35°N in winter, to perhaps 35°N to 45°N in summer, and with roughly double the speed in winter.

3. The location of the jet stream appears to coincide with the strongest meridional temperature gradient just below it in the mid-troposphere.

4. The jet stream is found where the tropopause has its greatest change in elevation.

5. The ideal circumpolar appearance of the jet stream is marred by appreciable longitudinal differences in strength. Thus in winter (Fig. 3) the jet stream reaches its maximum strength along and just off the east coasts of Asia and North America and decays in the eastern parts of the oceans. Another maximum appears over Africa.

6. The distribution of wind speed across the jet stream shows very strong lateral shear both to the north and to the south; the vertical component of the total absolute vorticity north of the jet stream (not illustrated here) suggests approximate constancy with latitude.

To a considerable extent charts constructed in mid-troposphere (Figs. 5 and 6) bear a striking resemblance to those at the level of the jet stream. This might be

inferred from the classical work of Dines [18] in which was pointed out the good correlations between temperature and pressure in mid-troposphere and at the tropopause. The strength of the circumpolar whirl is diminished with decreasing elevation. Below the level of 700 mb the circumpolar vortex begins to be somewhat obscured by the appearance of a more cellular character of the general circulation, and by the time we reach maps at sea level (Figs. 5 and 6) the "centers of action" take on considerable cellular character. Noteworthy is the emergence of easterly winds north of the subpolar lows and in the subtropics. The easterlies of the polar regions are for the most part shallow phenomena generally below 3 km, but in low latitudes, particularly in summer, easterly winds are very deep and extend well into the stratosphere.

The circumpolar vortex in the upper troposphere and lower stratosphere, while primarily zonal in character, undergoes gentle north-south undulations so that the isobars have a roughly sinusoidal character. The ridges and troughs aloft have their reflections at the surface

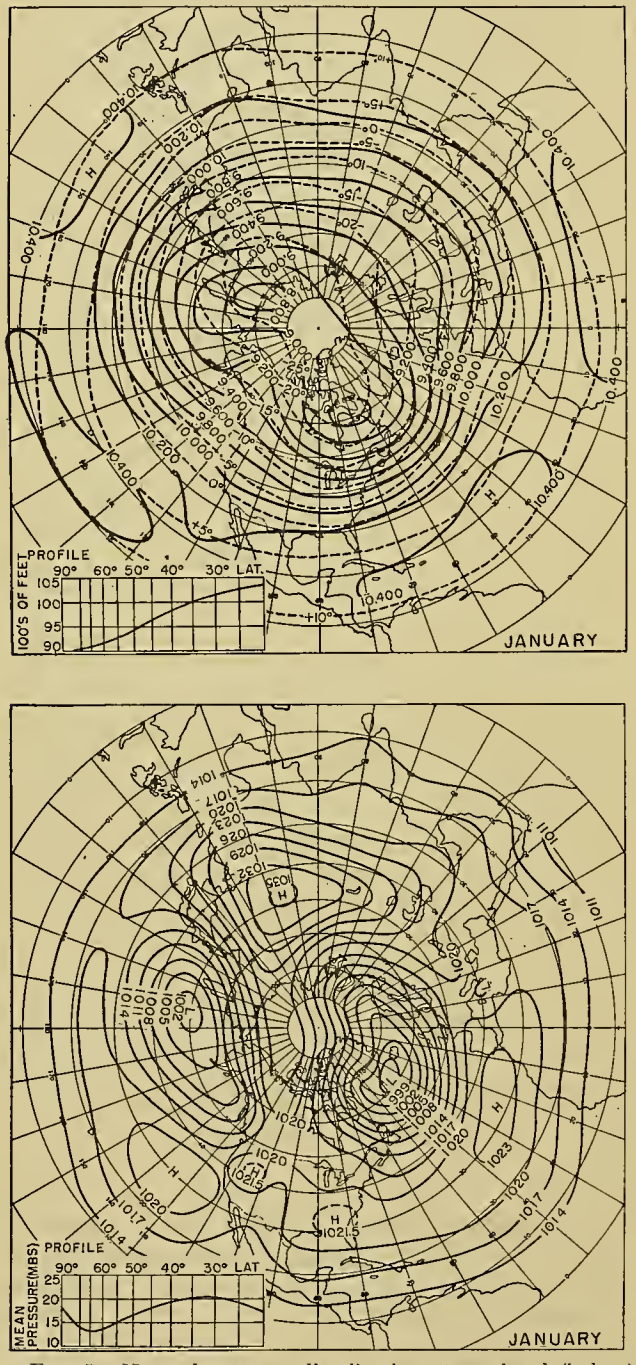


Fig. 5.—Normal pressure distribution at sea level (below) and normal contours at 700 mb (above) for January. Profiles inserted. Dashed lines at 700 mb are isotherms for every five centigrade degrees.

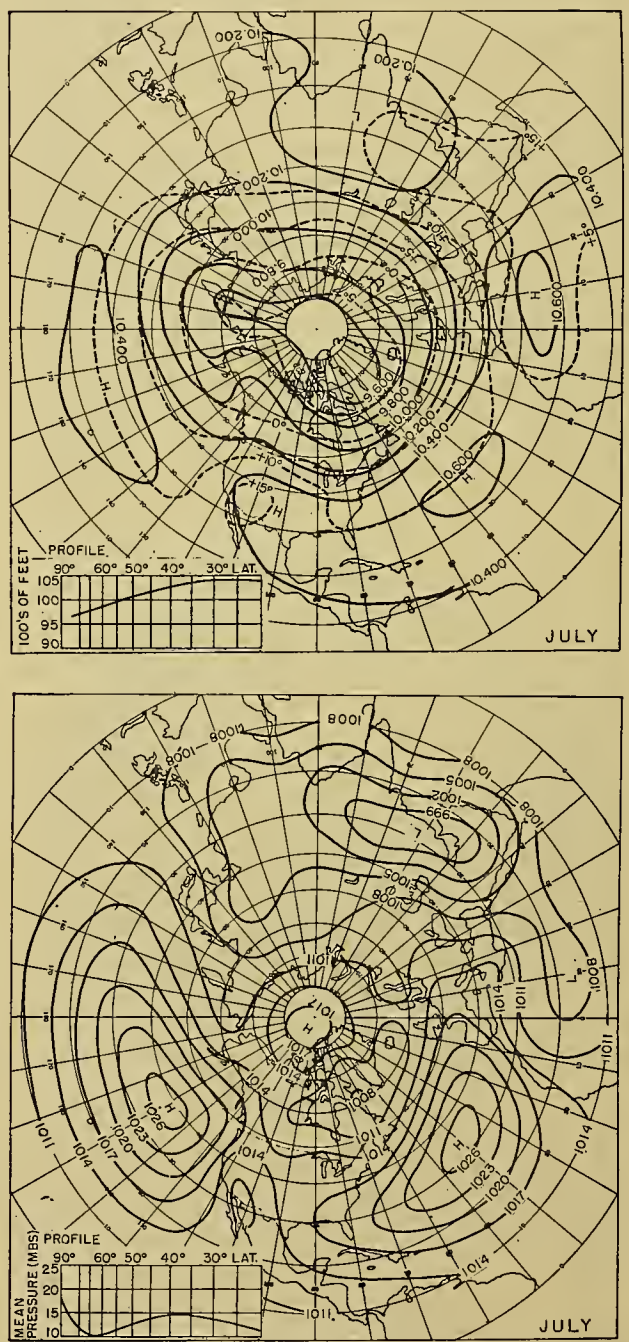


Fig. 6.—Normal pressure distribution at sea level (below) and normal contours at 700 mb (above) for July. Profiles inserted.

in the centers of action. Even at 300 mb the Icelandic and Aleutian lows of winter have not lost their identity although they are displaced considerably northwestward. The Siberian wintertime anticyclone apparently gives way to westerlies aloft.

The contrast between winter and summer shows up chiefly in diminished gradients and northward migration. With the approach of summer the sea-level subpolar lows lose much of their strength, the subtropical oceanic highs become much stronger, and over Eurasia the well-known monsoonal reversal of pressure sets in.

The comparison of circulations between the Northern and Southern Hemispheres is much more difficult in view of the relatively scant supply of data in the Southern Hemisphere—particularly at upper levels.[2] Certain attempts have recently been made to improve on Shaw's Southern Hemisphere charts [47] notably by the British Meteorological Office [9] for upper levels and by Willett [60] for sea level using Serra's world charts [46]. From these data a few seemingly reliable statements appear to be possible:

1. The zonal circulation at sea level appears to be stronger in the Southern Hemisphere than in the Northern Hemisphere. This is strongly indicated in a table of average circumpolar zonal wind speeds computed by Willett [60] who made use of Serra's monthly mean maps [46] for the months of January, April, July, and October from 1879 through 1934. (See Table I.)

Table I. Selected Normal Circulation Indices at Sea Level*

	Maximum zonal westerlies		Maximum subtropical easterlies	
	Wind speed (m sec^{-1})	Lat.	Wind speed (m sec^{-1})	Lat.
Winter				
Northern Hemisphere (Jan.)	2.25	54°N	5.60	20°N
Southern Hemisphere (July)	7.10	50°S	8.15	13°S
Summer				
Northern Hemisphere (July)	1.10	53°N	1.25	25°N
Southern Hemisphere (Jan.)	5.60	50°S	3.10	22°S

* Determined by choosing the 20° latitude band of greatest meridional pressure change, converting to geostrophic speeds, and noting the central latitude of the 20° band.

Clearly, the Southern Hemisphere zonal wind system at sea level is relatively stronger, displaced farther equatorward, and shows a smaller percentual seasonal variation than that of the Northern Hemisphere.

Willett also points out the surprisingly large seasonal fluctuation of the subtropical easterlies and their great strength as compared even to the zonal westerlies. This is indeed a fact which must be of importance in treating the general circulation, for the area covered by this wind system is far greater than that embraced by other branches of the zonal circulation.

A comparable table for upper-level circulations for both hemispheres is impossible to present at this time. From the British upper-level normals [9], however, it appears that the contrast of zonal wind speeds, so evident at sea level, is much less pronounced or perhaps even lacking at mid- and high-tropospheric levels. For example, the geostrophic zonal wind speed computed between latitudes 20° and 40° from the British 300-mb charts for both hemispheres during winter (December–February in the Northern Hemisphere, and June–August in the Southern Hemisphere) gives 29.5 m sec^{-1} for the Northern Hemisphere as compared with 29.7 m sec^{-1} for the Southern Hemisphere.

2. The undulations in the circumpolar vortex of the Southern Hemisphere appear to be much less pronounced (*i.e.*, have much less amplitude) than in the Northern Hemisphere. In fact, according to the British charts at 300 mb, it seems questionable even to talk of troughs and ridges in the Southern Hemisphere.

3. The cellular character at sea level is correspondingly less pronounced in the Southern Hemisphere than in the Northern Hemisphere, although "centers of action" are plainly present.

Various attempts have been made to characterize the state of the general circulation by numerical values. Thus certain circulation indices have come into use. Among the most prominently used are the geostrophic zonal wind speed, the "maximum" index, the meridional index, and zonal wind-speed profile. Definitions follow:

1. The geostrophic zonal wind speed within different latitudinal bands is computed either regionally or for the entire hemisphere. Such quantities as the "zonal index" express numerically the zonal wind speed averaged between two latitudes (generally 35°N and 55°N) and characterize the temperate-latitude westerlies. The polar easterlies (55°N to 70°N) and subtropical easterlies (35°N to 20°N) are similarly computed. These indices may also be evaluated for upper levels of the atmosphere.

2. The "maximum index" expresses the peak strength of some particular branch of the zonal circulation, as well as the latitude at which it appears.

3. Meridional indices express numerically the total north-south flow across a particular latitude circle.

4. In the last few years another effective form of representing important characteristics of the general circulation has come into use. This is the zonal wind-speed profile, on which the mean zonal speed within 5° latitude zones is plotted against latitude.

The seasonal behavior of Northern Hemisphere zonal indices computed for monthly periods is shown in Fig. 7. Recent work [35] with hemisphere-wide upper-air maps suggests that, at least at certain times of the year, normals computed for periods as long as a month contain smoothing sufficient to obscure certain important shorter period variations. Most important of these is the primary decline and subsequent rise of the mid-troposphere zonal westerlies which usually takes place

2. Subsequent to the preparation of this article two papers dealing with conditions at upper levels in the Southern Hemisphere have been published: "A Meridional Aerological Cross Section in the Southwest Pacific" by F. Loewe and U. Radok, *J. Meteor.*, 7:58–65 (1950); and "A Meridional Atmospheric Cross Section for an Oceanic Region" by J. W. Hutchings, *J. Meteor.*, 7:94–100 (1950).—*Ed.*

in February and March, the zonal index reaching a minimum around the beginning of March. Clearly such a minimum could not show up in monthly mean values such as are represented in Fig. 7. But averages com-

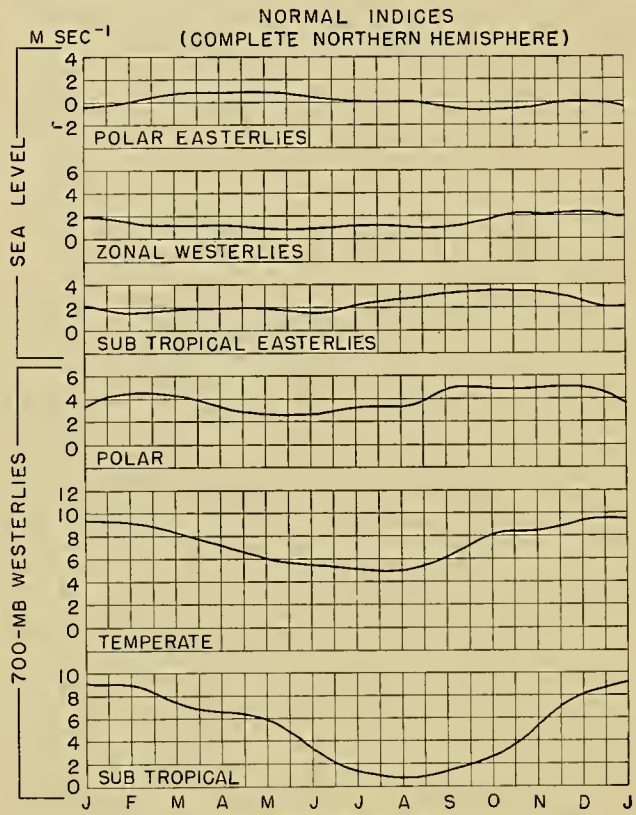


Fig. 7.—Normal Northern Hemisphere monthly indices.

puted from five-day mean charts worked up for ten-day intervals during the six years 1944–49 (when reliable upper-air data were available) strongly suggest the reality of this "global singularity." Another period of similar data (1932–39) worked up in part with the help of extrapolation techniques [59] also suggests the reality of this sharp dip in the normal zonal index. More concerning this topic will be said in a following section.

As pointed out earlier, the state of our knowledge of the observed normal state of the general circulation is still quite inadequate. Our deficiencies lie particularly in the absence of a long period of record of upper-air data over large areas of the Northern Hemisphere and over most of the Southern Hemisphere. Not much can be done for many years to remedy the "long-period" part of this deficiency. But it would appear well worth while to obtain, if only for one year, a reasonably complete series of daily soundings well into the stratosphere covering the entire world. From this material many of our blind spots would be cleared up. It might even be possible to link up interactions between the circulations of the hemispheres. Also, secrets of behavior of atmospheric circulation might conceivably be easier to find in the Southern Hemisphere where topographically produced distortions are comparatively lacking. In terms of modern expenditures for scientific work the cost of such a truly international observation year would not appear to be prohibitive. A more modest though much less effective proposal would be to establish at least one meridional cross section from the Arctic to the Antarctic.

There is also certain additional work which can be accomplished with our present material which could assist in a better definition of the character of the general circulation. It must be remembered that the normal state of the general circulation is not something which can be determined once and for all. It is something toward which successive approximations must be made. In the Northern Hemisphere the rate of increase in aerological data over the past ten years has been large. A stock-taking of our present rate of accumulation of this material indicates that new monthly normal charts for upper levels should again be prepared. This might be especially worth while in polar and subtropical regions. It would appear that the present rate of accumulation of aerological information would make necessary the routine preparation of upper-level normals at least every five years.

Physical Climatology of the General Circulation. The large-scale pressure patterns and other features revealed by a study of the normal or average conditions of the general circulation are generally assumed to have definite physical significance in the sense that they represent a stable stationary state of the circulation rather than a haphazard combination of daily highs and lows. To the extent that this is true it is desirable to find logical physical-mathematical definitions and explanations of these normal features, since they represent relatively simple states of the circulation which up to the present have been easier to depict than individual daily circulations.

A striking feature of the normal charts for the Northern Hemisphere [30] are the very long waves of finite amplitude at upper levels which are superimposed on the circumpolar westerly current. While their amplitude increases with latitude and decreases with altitude in the troposphere, their wave lengths do not change markedly with either of these coordinates.

These waves have dimensions greater than those of any other atmospheric wave systems. The smallest of them has a length of about 4000 mi and an amplitude of about 550 mi, compared to a dimension of about 1000 mi for waves associated with extratropical cyclones. Other interesting features are the consistently larger wave lengths and amplitudes found over Asia as compared to those found over North America, and the marked seasonal variations.

Some of these features appear to be explained at least qualitatively by the simple theory of conservation of vorticity as applied to planetary waves. Thus Rossby in 1939 [44] was the first to explain on a theoretical basis how a system of waves could be superimposed on the circumpolar band of westerlies, with troughs and ridges located in fixed geographical regions. He explained the anchoring of certain waves on the basis of distortions of the westerly flow due to north-south ori-

ented solenoidal fields at coast lines. More recently a similar role has been ascribed to mountain barriers.

By the use of Rossby's now classical formula,[3] plus the normal strength of the westerly flow, theoretical stationary wave lengths may be computed for each season for North America and Asia. The smallest of these has a length of 2500 mi, considerably larger than the dimensions of the average cyclone. Also, the seasonal variation of theoretical and observed stationary wave lengths is similar over North America. However, the observed waves are considerably longer than the theoretical waves. In view of the fact that average zonal wind speeds over the Eastern Hemisphere are somewhat lower than those over the Western, it becomes difficult to explain the larger observed wave lengths over Asia.

Among the many factors responsible for these discrepancies perhaps the most obvious is that the continents, in contrast to the oceans, interpose not merely one narrow obstacle in the path of the westerly flow but a whole series of more or less continuous obstacles of varying importance all the way around the globe. Thus, at best, we must regard the observed flow pattern as a heterogeneous combination of waves set up by each of the numerous obstructions. Precisely this line of attack has been taken by Charney and Eliassen [12].

Perhaps the most important reason for the discrepancies is that the atmosphere is baroclinic so that the westerly winds increase with height. As shown by Bjerknes and Holmboe [4], the only way to maintain a system of free stationary waves of constant wave length in such a current is to have a distribution of divergence such that in the lower layers there is mass convergence east of the trough lines and mass divergence to the west. At high levels they postulate the reverse distribution of divergence. Now the average level of nondivergence has been estimated to be at roughly 16,000 ft. Thus, the assumption of nondivergence at the 10,000-ft level, made in deriving the classical wave formula, is not valid. The distribution of divergence postulated above has the effect of increasing the stationary wave length.

Charney and Eliassen [11, 12] have shown that the effects of topography, surface friction, and baroclinity can all be expressed in terms of divergence at upper levels. It is therefore desirable to obtain an estimate of the normal fields of divergence. An attempt along this line has been made for the 10,000-ft level by the authors [37] and a sample of the results for February is shown in Fig. 8.

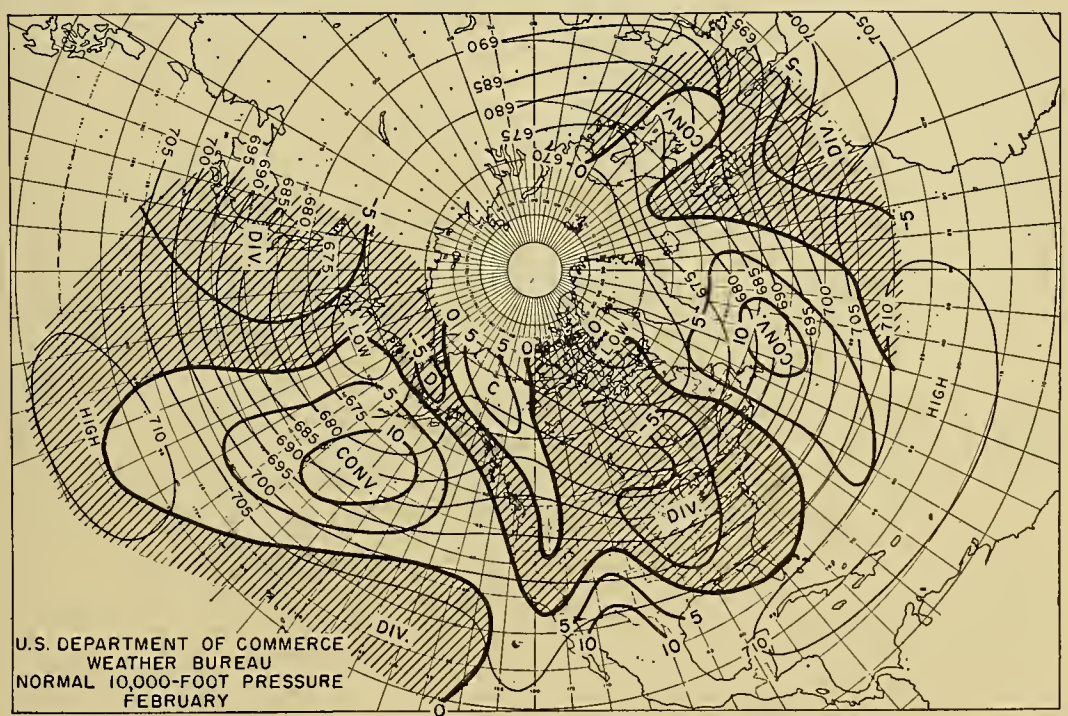


FIG. 8.—Normal field of divergence at 10,000 ft for February. Thin curved lines are isobars for 5-mb intervals; heavy curves, divergence for intervals of 5×10^{-7} sec^{-1}. (*From Namias and Clapp* [37].)

The large-scale features within the belt of strongest westerlies indicated in Fig. 8 show regions of maximum divergence to the west and convergence to the east of the trough lines, in partial confirmation of the Bjerknes-Holmboe theory [4] mentioned above. Another noteworthy finding (not illustrated here) is the increasing strength of the divergence fields from winter to summer accompanying subtropical highs while the fields associated with the circumpolar westerlies are weakening.

The effect of topography is also apparent from Fig. 8. As air is lifted over the Rocky Mountain chain from Alaska to Mexico, it is subjected to vertical shrinking (horizontal divergence), and as it sinks down the eastern

3. $L_s = 2\pi\sqrt{U/\beta}$, where U is the zonal wind speed (index), L_s the stationary wave length, and β the northward rate of change of the Coriolis parameter.

slopes it is subjected to vertical stretching (horizontal convergence). A similar phenomenon is observed in August but is interrupted at lower latitudes by the upper-level anticyclone over the United States. Similar charts, prepared for other levels, might throw considerable light on many problems of the dynamics of the general circulation.

If the normal fields of divergence at 10,000 ft are fairly representative of the average divergence in lower atmospheric levels where there are no mountains, it is possible to obtain some idea of the normal fields of vertical motion. Thus there is probably a region of gradual subsidence over the eastern United States and a region of gradual ascent over most of the Atlantic. Similar regions are indicated over the Pacific. The charts also indicate subsidence in eastern portions of the subtropical highs and ascent in western portions, conclusions which are in good agreement with the Bjerknes subtropical models [5] and with observed precipitation regimes.

It is recognized that in order to obtain a complete explanation of the normal state of the general circulation it is necessary to locate and determine the magnitude of the sources of heat and moisture in the entire world's atmosphere, for it is these sources which supply the necessary energy to drive the atmospheric heat engine. Some of the most recent work on this topic has been done by Wexler [54] and by Jacobs [25, 26].

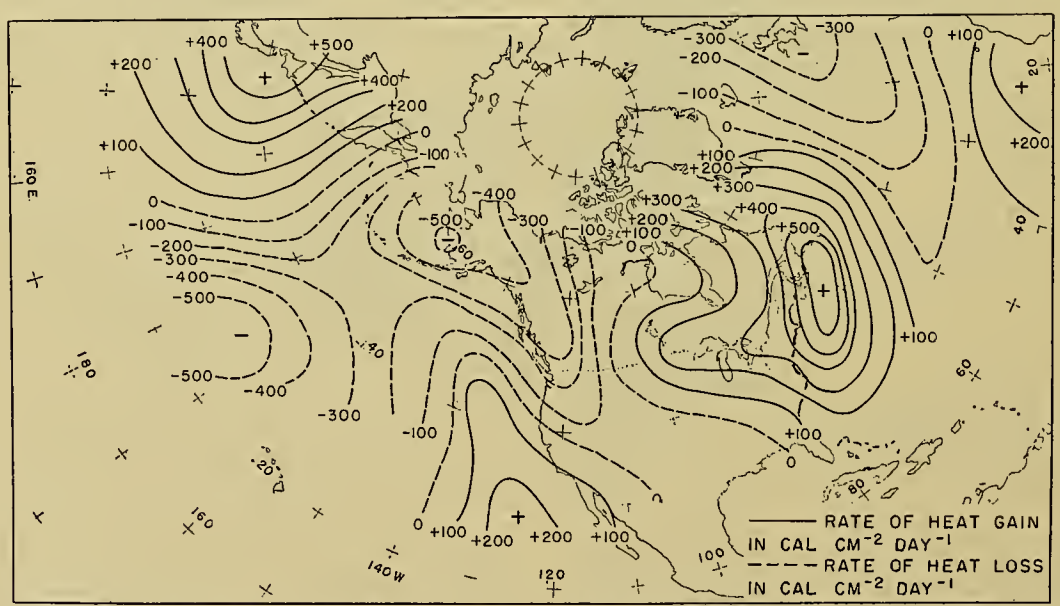


FIG. 9.—Regions of normal heat gain and loss for the layer from sea level to 10,000 ft for February. (*After Wexler* [54].)

Using data for the Northern Hemisphere, Wexler estimated the normal regions of heating and cooling of the air in the layer between sea level and 10,000 ft for the months of February (Fig. 9) and August. As Wexler points out, more exact values of the total heat gain or loss would be possible if the effects of vertical motion were considered. But the quantitative evaluation of this term is difficult. In an attempt to extend Wexler's study, Winston and Aubert,[4] estimating the effects of vertical motion, find that Wexler's method appears to give good estimates of total heating in regions where this heating or cooling is large. Thus, while part of the heat gain shown over eastern North America in Fig. 9 may be attributed to adiabatic warming through subsidence, it is not possible to explain in this manner the centers of strong heating appearing off the coasts of the continents. Apparently these represent heat sources instrumental in driving the circulation of the Northern Hemisphere. Similar reasoning applied to other areas suggests that the regions of cooling over the northeastern parts of oceans are heat sinks, while subsidence accounts for the apparent heat sources over the southeastern Atlantic and Pacific.

The studies discussed above do not consider individual heating processes affecting the circulation, for example, radiation, condensation, sensible heat exchange with the earth's surface, and mixing. Contributions to these aspects of the problem have been made by Jacobs [25, 26] who calculated for the air overlying oceanic regions the normal seasonal heating due to condensation and to sensible heat exchange with the surface. Jacobs found that the air was being heated from below in both the western and the northeastern portions of the oceans. The amounts of heating he found are in general agreement with the total heating as given by Wexler only in the western portions, but in the northeastern portions Wexler's values (showing net cooling) must be accounted for in some other manner, perhaps by radiation or mixing. Indeed the importance of horizontal mixing in cooling northward-moving air

4. Part of the work of this continuing project at the Extended Forecast Section of the U. S. Weather Bureau will be published shortly. A further study is being made of normal and anomalous sources and sinks for other months and other levels.

masses in the northeastern oceans is strongly suggested by the work of Elliott and Smith [21].

In studying the sources and transformations of atmospheric moisture Jacobs showed that the greatest evaporation in winter takes place over western oceans. According to him this latent heat appears to be released to the atmosphere largely in the eastern portions of the oceans where condensation reaches its maximum.

Another method of investigating the availability of moisture for atmospheric processes resulted from the technique of isentropic analysis. By this means Namias and Wexler [55] were able to construct monthly mean isentropic charts for summer over the United States and thereby indicate the normal upper-level sources of moisture. Unfortunately, due more to the discontinuance of the isentropic technique than to the lack of upper-air data, such charts for other areas and seasons have not been constructed.

The transport of various meteorological quantities by methods other than isentropic analysis has received considerable attention in recent years. Thus, Priestley [41] has suggested a systematic study of the transport of heat, moisture, and momentum through the use of radiosonde and radiowind measurements at a large network of individual stations. The importance of the transport and convergence of momentum and heat in producing local wind accelerations has been stressed by Namias and Clapp [38] and by Starr [48]. More study can profitably be made of the significance of thermal convergence in producing local transformations of energy.

NONSEASONAL (IRREGULAR) VARIATIONS IN THE GENERAL CIRCULATION

Qualitative Synoptic Studies. The "normal" state of the general circulation, as shown by normal hemisphere maps, is one which is never strictly encountered in any given daily map nor, for that matter, in any weekly mean or monthly mean. Indeed, the normal map is composed of the averages of circulations which appear remarkably different from week to week and even between corresponding months of different years. In any given region, however, there appears to be a preferred circulation type for a given month or season, so that the normal map computed from a long series of data achieves a certain character.

The variations of the circulation about its normal, considering weekly or monthly mean charts, is far greater than that which one would obtain if the circulation types followed each other in purely random sequence. This fact is reflected in the amazingly large variability of means for a month, year, or decade and probably for still longer periods up to the ice ages, as stressed by Willett [58] and Tannehill [49].

The observed sequence of types in any particular month or season more often than not takes the form of an alternation between two or more mid-tropospheric flow patterns, one of which persistently recurs and dominates the mean over longer periods. It is presumably this persistent recurrence which led the Multanovsky School of long-range forecasting [39] to draw conclusions with regard to "natural synoptic periods," to the opposition of types between natural periods, and to the relative constancy of length of the natural period during one "synoptic season."[5]

Some idea of the weekly variations of the strength of the zonal circulation at sea level may be obtained from Fig. 10. The absolute variations at higher elevations appear to increase up to the level of the jet stream, but the percentage variation of normal strength is probably quite similar at all levels. If the average zonal wind speed for the hemisphere is considered as a whole and for weekly periods, this variation is of the order of 50–150 per cent of the normal speed. Regional and shorter-period variations may, of course, exceed these values, and localized jet streams reaching three times their normal value have been recorded. There is some suggestion in the distribution of zonal indices of mid-troposphere that the frequency curve of index values is skewed in a direction such that high values (relative to normal) are less frequent than low. Perhaps this is suggestive of an atmospheric braking mechanism whereby some form of stability permits very low zonal wind speeds to persist but kills off unusually fast currents.

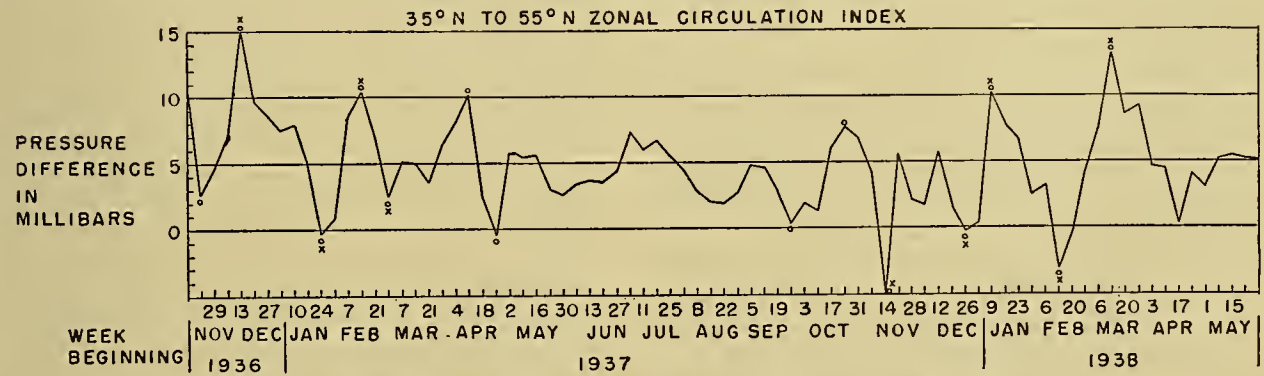


Fig. 10.—Weekly averages of zonal westerlies index, November 1936—May 1938.

5. Consult "General Aspects of Extended-Range Forecasting" by J. Namias, pp. 802–813 in this Compendium.

The variation in speed of the zonal circulation may also be treated from the standpoint of the maximum index, or the values of zonal wind speed taken at whatever latitude the maximum speed is reached. While the nature of the speed variations thus treated is similar to those taken between fixed latitudes, this form of treatment brings out some important additional information indicating that some of the highest speeds are reached when the jet is found at low latitudes. It also brings to light latitudinal variations in the position of the jet stream about the normal which are amazingly large. Even if averaged over the hemisphere for periods as long as a month, this variation can be as large as 20° of latitude during one season. For shorter periods, and particularly for selected regions, the latitudinal variation of the jet stream can be much greater. Regional jet streams have been apparently observed in areas ranging from the pole almost to the equator.

To the extent that we may speak of the maximum high-level zonal current averaged over the hemisphere as a jet stream, we may summarize the foregoing remarks by saying that there are large day-to-day, week-to-week, and month-to-month variations in the strength and latitude of the jet stream. The latitudinal variations may be looked upon fundamentally as an expansion and contraction of the circumpolar vortex toward or away from the pole.

The regional variations in the latitude and perhaps in the strength of the jet stream are associated with the wave patterns in the circumpolar vortex which were, to some extent, treated in a foregoing section. Ideally these long waves are generally looked upon as a simple wave pattern in the upper troposphere and lower stratosphere with from four to six meridionally extensive ridges and troughs around the hemisphere. The observed wave patterns are rarely if ever so ideal, and a more accurate description would be that there are generally two (sometimes more) different families of waves in different latitude bands. These two wave trains may, and usually do, possess different wave lengths (in degrees of longitude) and hence around the hemisphere there may be a different number of waves in higher latitude bands than in lower latitude bands. For this reason, too, the waves in the two bands are frequently out of phase. These remarks pertain equally well to daily, weekly, or monthly mean maps.

When the jet stream of the circumpolar vortex is well to the north of its normal position, the ridges and troughs are generally strongly tilted in a northeast-southwest direction or even fractured as indicated in Fig. 11, Stage 1. When the waves get into phase (*i.e.*, when the troughs and ridges are more or less meridionally extensive) the latitude of the jet stream usually lowers but does not reach its lowest value. This position is reached only after the entire character of the hemispheric upper-tropospheric circulation is radically altered to a cellular rather than a wavelike structure. The explanation of this fundamental change in character, in spite of its enormous importance to meteorology, is not yet known. However, in following the development of such transitions it often appears that the changes go on somewhat as follows: In the initial stage (illustrated schematically in Fig. 11) there are extensive zones of confluence where deep cold polar and warm

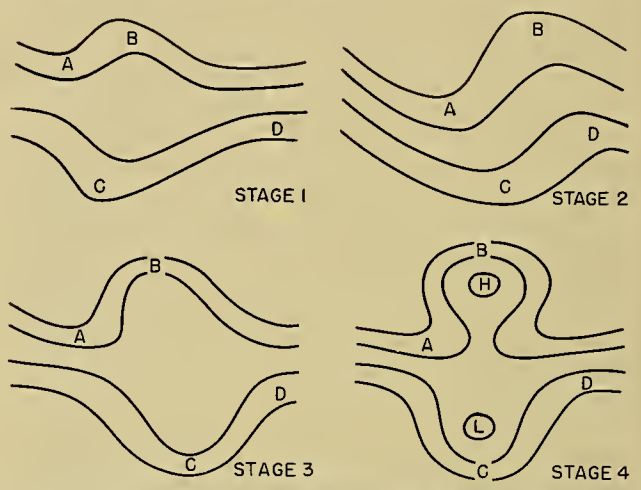


Fig. 11.—Schematic representation of successive circulation patterns of the index cycle. The time interval from Stage 1 to Stage 4 is about two weeks.

tropical air masses are forced to flow side by side, associated with filaments of regional jet streams to the east of the zones of confluence. In Stage 2 the waves, perhaps through some form of relative motion associated with a nonstationary wave length, have combined into deep full latitude troughs and ridges. This combination permits vast meridional transports of deep polar and tropical air. Stage 3 results from a refracture of the troughs at high and low latitudes, again perhaps in an attempt to readjust wave lengths to equilibrium values. The cold polar air masses brought equatorward and the warm currents which were delivered to polar regions by the in-phase troughs and ridges now have no return passage to their source regions—that is, the evolution of the flow patterns is now in the direction of trapping both the equatorial air in the north and polar air in the south. Finally, in Stage 4 the cold pools (*Kaltlufttropfen*) in the subtropics and the warm pools in the subpolar regions take on characteristic cyclonic and anticyclonic cellular circulations. In these schematic models it will be noted that certain asymmetry and tilt of the troughs and ridges are introduced where necessary to bring about the advection of momentum required to maintain the westerlies in the manner postulated. Jeffreys [27] and lately Starr [48] have emphasized the importance of such asymmetry as related to the transfer of momentum.

This idealized sequence of events generally does not take place simultaneously in all parts of the hemisphere. However, when such a development on a large scale takes place in one region, it is often followed by similar reactions in other regions. This transition of the circulation wherein deep cold cyclones are ultimately formed at low latitudes and deep warm anticyclones are formed at high latitudes is referred to as *blocking*. The most commonly observed behavior of blocking action is a progressively westward development of re-

gionally low-index circulation patterns at a variable speed which synoptic experience suggests may be of the order of 60° of longitude per week. At the surface in the area of blocking, pronounced anticyclogenesis (or cyclolysis) is observed in higher latitudes, while wave cyclones are forced to skirt the perimeter of the developing high-latitude anticyclone, either northward west of the developing upper-level ridge at high latitudes, or southeastward into the cold developing vortex of southern latitudes. A good description of the evolution of the day-to-day westerly waves in the mid-troposphere westerlies [2] during blocking is that they appear to become progressively shorter and increase in amplitude as the scene of the blocking action is approached. In this respect they may be likened to waves advancing on to a beach, or to the pleats in an accordion.

Westward progression of characteristic circulation features, while by no means rare, does not appear to be as common as eastward progression. Thus a sudden increase in amplitude of a wave system in North America may be shortly followed by a similar increase over the Atlantic and Europe. The speed at which this readjustment occurs often exceeds that of the zonal speed of the air particles—a fact which of late has received considerable attention from theoreticians [43, 61] following its discovery in synoptic practice [36].

The transition in temperate latitudes from strong to weak zonal circulation and back again has been referred to as the index cycle. On the basis of a careful synoptic and statistical study of seven years of Northern Hemisphere data, Willett [45] has given a description of the index cycle which can hardly be improved upon. His statements follow:

> . . .four principal states of the index cycle are recognized, each of which can be briefly characterized essentially as follows:
>
> (1) Initial high index (strong sea-level zonal westerlies), characterized by (*a*) sea-level westerlies strong and north of their normal position, long wavelength pattern aloft; (*b*) pressure systems oriented east-west, with strong cyclonic activity only in higher latitudes; (*c*) maximum latitudinal temperature gradient in the higher middle latitudes, little air mass exchange; and (*d*) the circumpolar vortex and jet stream expanding and increasing in strength, but still north of the normal seasonal latitude.
>
> (2) Initial lowering of sea-level high-index pattern, characterized by (*a*) diminishing sea-level westerlies moving to lower latitudes, shortening wave-length pattern aloft; (*b*) appearance of cold continental polar anticyclones in high latitudes, strong and frequent cyclonic activity in middle latitudes; (*c*) maximum latitudinal temperature gradient becoming concentrated in the lower middle latitudes, strong air mass exchange in the lower troposphere in middle latitudes; and (*d*) maximum strength of the circumpolar vortex and jet stream reached near or south of the normal seasonal latitude.
>
> (3) Lowest sea-level index pattern, characterized by (*a*) complete breakup of the sea-level zonal westerlies in the low latitudes into closed cellular centers, with corresponding breakdown of the wave pattern aloft; (*b*) maximum dynamic anticyclogenesis of polar anticyclones and deep occlusion of stationary cyclones in middle latitudes, and north-south orientation of pressure cells and frontal systems; (*c*) maximum east-west rather than north-south air mass and temperature contrasts; and (*d*) development of strong troughs and ridges in the circumpolar vortex and jet stream, with cutting off of warm highs in the higher latitudes and cold cyclones in the lower latitudes.
>
> (4) Initial increase of sea-level index pattern, characterized by (*a*) a gradual increase of the sea-level zonal westerlies with an open wave pattern aloft in the higher latitudes; (*b*) a gradual dissipation of the low-latitude cyclones, and a merging of the higher-latitude anticyclones into the subtropical high-pressure belt; (*c*) a gradual cooling in the polar regions and heating of the cold air masses at low latitudes to re-establish a normal poleward temperature gradient in the higher latitudes; and (*d*) dissipation of the high-level cyclonic and anticyclonic cells, with a gradual re-establishment of the circumpolar vortex jet stream in the higher latitudes.

While the index cycle described above appears to be recognized in its essential character during most winters, its precise form of operation, time of onset, and even its length and intensity are subject to appreciable variations. Extensive studies directed along these specific lines appear to be long overdue. A beginning on this problem [35] was made with the help of hemisphere-wide upper-air coverage obtained during the six-year period 1944–49. These data suggested the following conclusions:

1. There is a strong preference for the winter's primary index cycle to occur in February and March, the minimum index occurring around the beginning of March, and the beginning and end points being removed by about three weeks from the minimum.

2. The total momentum of the mid-troposphere westerlies around the hemisphere tends to reach a certain value characteristic of the season, and it is only the distribution of this momentum with latitude that varies. Hence the low index of temperate latitudes really consists of a displacement of the "high-index" strong westerlies (farther southward).

3. Each primary index cycle is usually associated at its onset with a strong wave of Atlantic blocking, with a tendency to form warm anticyclones in high latitudes and cold cyclones in low latitudes. However, the blocking appears to be only a necessary and not a sufficient condition for a primary index cycle.

4. The intensity of long index cycles appears to be largely determined by the reservoir of cold air in polar regions preceding their onset.

These four statements appear to be sufficiently backed by empirical evidence to be considered in any theoretical treatment of the problem of the evolution of the radically different states of the general circulation. A qualitative treatment along these lines has been given [35]. The reader is referred to this paper since the presentation of theories lies outside the scope of this article. Other earlier attempts to explain the index cycle have been given by Rossby and Willett [45].

The index cycle, consuming about six weeks, is naturally not the only important period of evolution of radically different circulation patterns. Indeed for many years the problem of monthly mean, seasonal, and even secular large-scale circulation anomalies has been of paramount concern to meteorologists. While such evolutions undoubtedly lie in the province of "Observed

Studies of General Circulation Patterns," it would appear that the principal treatment of this vast subject belongs elsewhere in this Compendium.[6]

Before we leave the subject, however, a few pertinent remarks are in order. First, it appears from actual studies that the choice of the length of period does not essentially change the character of the problem of fundamental oscillations of circulations from high- to low-index states. In fact, there is evidence to suggest (Willett [57], Tannehill [49]) that the longer the period, the correspondingly greater the interperiod variability compared with what might be expected if the variations were random in character. Secondly, experience with monthly mean charts over the past eight years suggests an evolution which is not entirely chaotic and which indeed appears capable of kinematic and possibly physical rationalization [31, 34]. The kinematic aspects were pointed out as early as 1926 by Brooks [8].

In short, the problem of the index cycle and its evolution would appear to be an integral part of any theory of secular, climatic, and geological (*i.e.*, ice-age) variations of world weather—a fact repeatedly stressed by Willett [58].

Quantitative Empirical Studies. The qualitative behavior of large-scale circulation patterns described above is obviously of sufficient interest to justify more objective quantitative studies. Such studies have two basic purposes: to provide empirical formulas for predicting the motion and development of large-scale circulation features, and to furnish the theoretical meteorologist with quantitative evidence for supporting or extending his theories. Since the forecasting aspects of these studies are treated elsewhere,[5] we shall treat here only the theoretical implications.

One investigation of this sort, involving the motion of long waves at different latitudes, was made by the authors with the aid of two-and-a-half years of five-day mean 700-mb charts [36]. This study was an attempt to verify and make use of the simple theory of planetary wave motion based on the principle of conservation of absolute vorticity [44].

From this material it was found that while there is positive (but far from perfect) correlation between observed and theoretically computed displacements of selected trough systems, observed waves usually travel much faster than the speed given by the vorticity theory, thereby making necessary large empirical corrections. The reason for this discrepancy seems to lie in part in the fields of divergence accompanying observed waves, whereas the theoretical formula is based on the assumption of no divergence. This problem, as it affects the stationary wave length, was discussed in the section of this article dealing with physical climatology.

Of considerable importance is the empirical finding that when wave length and zonal wind speed are considered separately there appears no significant relationship between wind speed and displacement. This was certainly not expected from theory. While no satisfactory explanation for this discrepancy has yet been found, it may be due in part to a possible interdependence of wave length and zonal index [13] or, as Cressman [14] indicates, to the small variability of the zonal index which cannot produce changes in displacement larger than the errors of measurement.

Because of this finding, empirical formulas were derived by considering that displacement is a simple linear function of wave length. This is equivalent to substituting the average or normal value for zonal wind speed in the theoretical displacement formula and then assuming that the range in wave length is small in comparison to its average magnitude.

The latest unpublished studies of this kind show that the stationary wave lengths (the wave lengths observed when trough displacement is zero) for North America are roughly the same as the normal wave lengths obtained from normal upper-level charts. This means that when the wave lengths found on individual five-day mean charts exceed the normal value, the waves will tend to retrograde (move westward) while they will be progressive (move eastward) if the observed wave lengths are smaller.

It was previously suggested that seasonal variations of the normal wave lengths could be accounted for by seasonal changes in the zonal index. But there are also pronounced geographical variations. Thus the stationary wave lengths for the Atlantic are found to be shorter than those for North America, suggesting that for the same wave length, displacements in an easterly direction are slower over the Atlantic. This better agreement with the theoretical wave formula indicates that waves over flat ocean areas more closely approximate free perturbations.

The empirical studies also show a systematic decrease of wave speed with time. This may be a purely statistical result, but a possible physical explanation may lie in the Rossby wave formula in which a decrease in wave speed may result from a systematically increasing wave length. It has been suggested [36] that such an increase in wave length as troughs move eastward may be due to the tendency for certain troughs and ridges to be fixed because of topographical or solenoidal effects. For example, the length between a trough moving eastward over the Mississippi Valley and a ridge fixed to the Rocky Mountain area must continually increase, resulting in a deceleration.

Perhaps a more universal explanation for systematic changes in wave length has been offered by Cressman [15], who has adapted the theory of group velocity to changes in wave length of individual systems. Synoptic evidence, on both daily and five-day mean charts, appears to support his conclusion that if the wave length increases in an upstream direction the easternmost wave will slow down, while the opposite distribution of wave length will lead to acceleration.

In investigating other parameters besides wave length which might be related to displacement, it was found that the shape of the wave was often quite important [13]. Thus, if the wave amplitude to the west of a given

6. See, for example, the discussion in "Solar Energy Variations as a Possible Cause of Anomalous Weather Changes" by R. A. Craig and H. C. Willett, pp. 379–390.

trough is greater than that to the east, the trough will tend to move faster than expected from the empirical wave formula. The opposite result is often obtained if the amplitude to the east is greater.

Studies of trough motion similar to those above have been made with the aid of constant absolute vorticity trajectories [42]. The theory, which gives an estimate of the paths of individual air parcels, is based on the conservation of absolute vorticity just as in the wave theory, but avoids many of the latter's assumptions. For example, local wind speeds are used in place of a fictitious uniform zonal wind, and the assumption of small perturbations is avoided, so that estimates of wave amplitude may be obtained. However, a new assumption that the speeds of individual air parcels remain constant is introduced.

Using this theory, Bortman [6] obtained empirical corrections (for each season and several geographical areas) to the "first characteristic points." These are the first trough or ridge points following the inflection point of the vorticity path at which the computation is made. The results were quite similar to those using the wave formula in that large eastward empirical corrections were found for the theoretical trough points over eastern North America, indicating that the theoretical trough speeds are too low. Also, there is a positive correlation between theoretical and observed wave speeds. Corrections are generally less over ocean than over land areas. Of special interest are the corrections in the western Rocky Mountain area, where both trough and ridge points give evidence of strong divergence as a result of lifting. The study thereby emphasizes the importance of geographical and climatological features on the behavior of long waves.

Another unpublished empirical study of wave motion on five-day mean 700-mb charts [7] over North America attempts to make use both of the theory of planetary waves and that of constant vorticity trajectories. Thus, following a suggestion by Cressman [14], the wave length and wind speed of observed waves were measured along the streamlines (or isobars) rather than along fixed latitude circles. Amplitudes were measured in the same manner. These three parameters were then related by graphical correlation methods to the observed trough displacements at 45°N. Such a procedure is an empirical attempt to take cognizance of the jet stream, which tends to follow the streamline flow. The results show not only the expected relationship between wave length and displacement, but indicate that amplitude and to a certain extent wind speed are also significant parameters. The most interesting result was the significant positive correlations found among all three independent parameters. The relationships among wave length, amplitude, and wind speed for small and intermediate values of wave length were found to be qualitatively the same as those to be expected from the theory of constant vorticity trajectories, while for larger wave lengths the agreement was poor (Fig. 12). The explanation for this discrepancy has not yet been found but when found it should aid in seeking an understanding of the observed behavior of planetary waves. Studies of a similar nature should be made for other levels and areas besides North America, to which the above study applies.

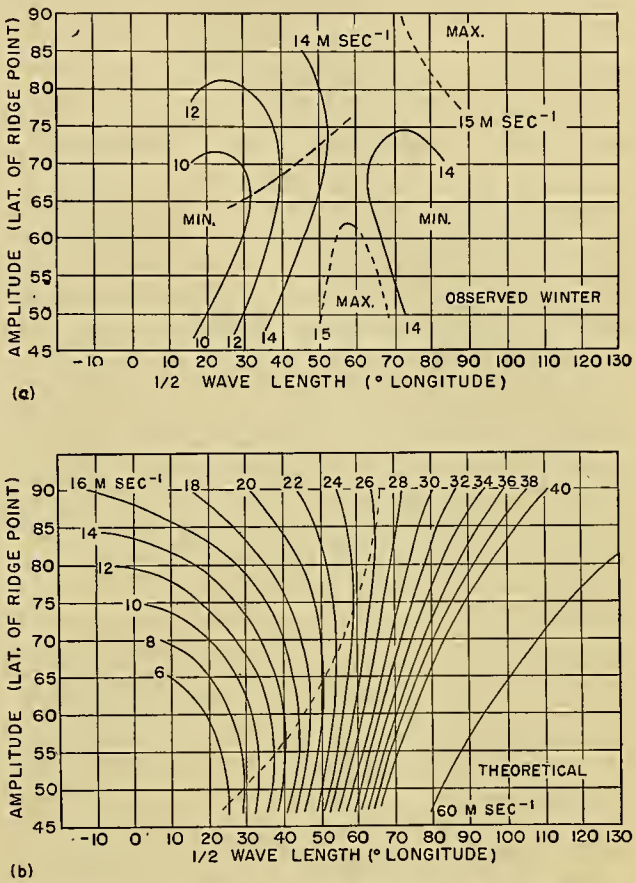


FIG. 12.—(*a*) Observed relationship between wave length, amplitude, and wind speed for waves whose minimum latitude is 45°N, from five-day mean 700-mb charts for the years 1941–47. (*b*) Theoretical relationship between wave length, amplitude, and wind speed for waves whose minimum latitude is 45°N, using the Bellamy vorticity slide rule. (*From Bortman* [7].)

One of the basic difficulties of all studies treated so far is that they require the identification from map to map of various characteristic points or lines (highs, lows, troughs, or ridges). Because of the complexity of atmospheric behavior this is not always easy. This suggests that a more profitable empirical (or theoretical) approach is to study local changes in various meteorological quantities, since this does not involve the troublesome problem of continuity. To a certain extent the vorticity paths are designed to do this, and such use has been made of them by Dorsey and Brier [19] and by Fultz [23].

Dorsey and Brier, using selected trajectories made on daily synoptic 10,000-ft charts, compared the theoretically computed wind speeds and directions with those observed at the same points at intervals of 24 hr up to seven days in advance. Among their interesting results is the finding that trajectories are likely to be more successful when made in a well-defined fast flow. It was also found that if the trajectory verified well for the

first 24 hr it was more likely to be good for subsequent time intervals. In spite of the obvious advantage of avoiding the necessity for determining continuity, it is felt that this method, unlike that used by Bortman [6] places too heavy requirements on the theory, and makes the determination of empirical corrections difficult.

A similar study has been made by Fultz [23] who also used trajectories computed on daily 10,000-ft charts, and in addition estimated the true trajectory of the air parcels. Such a method could obviously lead to quantitative empirical corrections to the vorticity paths in many different geographical areas. However, the number of cases studied by Fultz was small (largely because of the length of time necessary to compute true paths) and the corrections were apparently so complex that the author confined his practical results to a number of valuable qualitative rules.

Among recent studies designed to avoid the problems of determining continuity perhaps the most important is the work of Charney and Eliassen [11, 12] mentioned previously. This represents a considerable refinement of the simple planetary wave formula particularly as it applies to the determination of local pressure (or height) changes. Several empirical studies are currently underway to test and expand this theory.

Before concluding this brief review of quantitative studies of circulation patterns it is necessary to say a few words about the accuracy that has been attained. It must be confessed that in spite of the fact that they seem to fit into a logical pattern which can be explained on physical grounds, these empirical results represent only the average behavior of circulation features. For example, the largest correlation coefficient found between theoretical and observed wave displacement represents a degree of success accounting for perhaps only 50 per cent of the variability of observed motions. These meager results are probably due in part to the great complexity of the atmosphere which makes it impossible for any single factor, such as a wave length, a vorticity path, or a region of confluence to be responsible for more than a small part of the subsequent circulation changes. Only by integrating in some way *all* the many processes taking place throughout the *whole* atmosphere can one hope to approximate a complete solution. Such a desirable goal can be obtained only through continued close collaboration between theoretical and synoptic meteorologists.

Another contributing factor lies in the limits of observational accuracy. Because of instrumental errors, sparseness of upper-air reports, and analytical difficulties it is not generally possible to define a trough or ridge at 45°N any closer than about four degrees of longitude. Because of this one source of error (even supposing that a perfect linear relationship exists between wave length and trough speed), the highest correlation that could be attained is −0.85. This result suggests that in meteorology, as in other sciences, it is highly desirable to include in any quantitative comparison of theoretical and empirical findings a careful analysis of errors.

Statistical Studies of the General Circulation. Meteorology has always been a fertile field for statistics. It contains numbers of observations perhaps equal to or in excess of those in any science. The number of combinations of elements making use of different places and different time intervals is well nigh infinite. Partly for this reason, meteorology has encouraged the use of the correlation technique for discovering significant relationships between meteorological elements or between one and the same element in space or in time. While this statistical procedure has its secure place in meteorological research, it has probably been the cause of more heated controversy among meteorologists than any other research tool. Owing to the amazingly large mass of correlations to be found in the literature and the lack of general agreement as to tests of significance to apply to such statistics, it becomes exceedingly difficult at this time to portray or evaluate the role of statistical studies of the general circulation. Perhaps the most indefatigable worker along these lines has been Sir Gilbert Walker, whose statistical studies of contemporaneous and lag correlations form the basis of many interesting and informative papers [52]. For the most part these papers have been reviewed at length elsewhere [40] and thus the present brief survey will not attempt to treat this work but instead will concentrate on the work of Willett [57, 59] who, perhaps next to Walker, has given the atmosphere its most vigorous statistical workout. Besides, Willett had the distinct advantage of having hemisphere-wide maps at the surface and aloft and could thereby obtain integrated indices of the general circulation which Walker had to estimate by values of elements chosen at selected points. Moreover, Willett's work is of particular interest inasmuch as it ties in with, and indeed, forms some of the basis of, the material discussed in earlier portions of this article.

The principal meteorological material upon which this work was based consists of daily analyzed charts at sea level and at 10,000 ft over the entire Northern Hemisphere for the cold months of the year from October through March for the seven years 1932–39. From these charts, five-day mean maps were prepared twice a week (so that there was one or two days overlap between five-day periods) as well as a host of statistical derivatives. A list of the complete number of these "indices" of the general circulation would in itself occupy considerable space, even without the addition of the vast number of correlations computed between pairs of indices. The principal indices considered on a hemisphere-wide scale can be relegated to three types: zonal, meridional, and solenoidal. The zonal indices have been defined on page 555.

The meridional indices used by Willett express the mean (over space and time) of the north-south component of geostrophic flow averaged for each of the latitudes 30°N, 45°N, and 60°N. This quantity was necessarily obtained from values taken from daily synoptic charts and, as in the case of the zonal indices, was obtained for both sea-level and 10,000 ft.

The solenoidal index measures the poleward gradient of mean virtual temperature between sea level and

10,000 ft in the 20° latitude band of greatest northward rate of change.

With these indices (and many more) Willett proceeded to look for contemporary and lag correlations, after removing the seasonal trend. Ordinary linear (Pearsonian) coefficients of correlation were computed from fifty-two overlapping five-day mean periods for each six-month period (October–March) during each of the seven years and for the seven years as a whole. The terms "significant" and "highly significant" were applied to those correlations which had less than five per cent and less than one per cent probability, respectively, of occurring by chance.

Of the contemporary correlations one of the most significant was the negative one (−0.54) found with fair year-to-year consistency between the average pressure along latitudes 70°N and 35°N. This negative correlation is, on a global scale, the same phenomenon that Sir Gilbert Walker stressed in his North Atlantic and North Pacific oscillations, when he pointed out the opposing reactions of Aleutian and Icelandic lows to North Pacific and Azores subtropical anticyclones, respectively. These results stress the important fact that the polar-cap anticyclone grows at the expense of the subtropical highs, the shifts in mass being associated with an expanded circumpolar vortex of low index described above on page 561.

Largely as a result of this negative correlation the sea-level polar easterlies correlate insignificantly with sea-level zonal westerlies in spite of the fact that these two indices possess the latitude 55°N in common. But the sea-level polar easterlies correlate significantly and consistently from year to year in a positive sense with the zonal westerlies at the 10,000-ft level, indicating that when the sea-level polar easterlies are strong, the poleward temperature gradient in the 10,000-ft layer above the surface is also strong. This is also borne out by significant negative correlation between solenoid index and zonal westerlies and positive correlation between solenoid index and 10,000-ft westerlies. The correlation between the strength of the maximum 3-km westerlies and their latitude, while not significant, is consistently negative from year to year, suggesting the observed fact that the jet stream at times intensifies as it expands equatorward.

The meridional index computed for 45°N shows a tendency (though not quite significant) for strong meridional circulations to be associated with weak zonal components. The meridional exchange at 45°N also appears to be related to the latitude of the maximum 3-km westerlies, since a significant negative correlation exists between these two quantities. Again this reflects the meridional (*e.g.*, cellular) character of the low-index pattern as compared with the high-index pattern.

From these correlations Willett [57] concludes that

> . . .the low index circulation pattern, in contrast to the high index pattern, is characterized by:
>
> 1. A relatively strong poleward temperature gradient, at least between sea-level and the 3-km level.
>
> 2. Relatively weak zonal westerlies at sea-level which increase to relatively strong aloft, with a tendency to be displaced equatorward, that is, an intensified and expanded circumpolar vortex in the upper troposphere.
>
> 3. Strong polar easterlies as a result of a relatively strong sea-level polar anticyclone, which in turn is produced primarily from a weakening of the sub-tropical high pressure belt.
>
> 4. In middle latitudes a relatively strong meridional circulation at sea-level which tends to weaken with height as the zonal westerlies become relatively stronger.

The step from contemporaneous to lag correlations involving features of the general circulation appears to be of a high order of difficulty. In earlier decades, particularly the 1920's, a popular form of research aimed at long-range forecasting consisted of correlating, with different lags, weather or circulation elements at various points over one or both hemispheres. One such ambitious program, undertaken at the U. S. Weather Bureau under Weightman [53], involved correlations between pressure, temperature, and rainfall for lags of three, six, and nine months. It was hoped that these supposedly fact-finding and purely statistical studies would lead to a better understanding of the general circulation and provide a basis for long-range weather forecasting. But the technique was apparently too crude to bring to light any very revealing secrets of the behavior and evolution of atmospheric circulation and concomitant weather. Similarly, in their vast undertaking Willett and his associates, working with data described previously, were unable to detect any reliable predictors of the general circulation in its zonal or meridional branches or in regional divisions, aside from a small degree of persistence. From these negative results Willett [57] concluded that the general circulation possessed no internal mechanism of operation and that there was apparently no one way that a subsequent state could dynamically and thermodynamically evolve from an initial state. From this conclusion it was apparently not a difficult step for Willett to join that group of meteorologists who maintain that controls of nonseasonal variations in the general circulation lie in the sun. One can, of course, question the fundamental premise upon which this latter conclusion was based. Indeed, Willett himself is quite aware of the possibility that the correlation technique, applied as he has applied it, is too crude to unveil the secrets of a highly complex organism like the atmosphere. The restrictions placed upon correlation techniques by fixed time and space intervals, in the opinion of the present authors, are much too severe to fasten upon an evasive atmosphere notoriously resistant to strait jackets.

In this semiphilosophical question seems to lie one of the principal deficiencies of research in the problem of general circulation. Admittedly, all meteorologists aim for objectivity. But in the quest for objectivity it is sometimes forgotten that adequate statistical tools may not yet be developed to treat many complex problems. The problems of time series, so common yet so annoying to meteorology, serve as an example. It thus behooves the meteorologist to encourage the development of statistical tools which are by definition objective

and are at the same time not overly restrictive. Perhaps the way out of this dilemma is offered by new electronic high-speed computing equipment.

CONCLUSIONS

In the brief summary of recent work on the general circulation presented above, three basic aspects have been stressed. The first deals with what is known of the observed fields of atmospheric mass and motion, both in their normal state and in their week-to-week or month-to-month variations. While there is a fair delineation of the pressure field in the Northern Hemisphere and in the lower troposphere for a short period of years, there are considerably less systematic analyses of wind and moisture on a hemispheric scale. The second aspect concerns certain quantities such as divergence, vertical motion, heat sources and sinks, transport of mass and momentum, and other factors which can be derived from the fundamental fields of wind and pressure. Finally, there has been some discussion of the physical interpretation of this great mass of observations. It is this interpretation which provides a basis for long-range forecasting.

Much work remains to be done in all three aspects of the problem, especially in the last. At present there is no completely quantitative utilization of theoretical findings in forecasting. Perhaps the closest approach to this is confined to certain relatively simple applications of the classical hydrodynamical concept of conservation of vorticity. This meager result is due largely to the fact that, just as in other sciences, advances in theoretical meteorology must go hand-in-hand with advances in observation. For example, there appears to be much promise in the application of electronic machines and numerical integration methods to the problem of short- and long-range forecasting. However, even this will be of little avail unless it is accompanied by a better understanding of atmospheric processes through more intensive studies of existing data and improvement in observational techniques.

In various places throughout the text the authors have suggested several ways in which further progress in observational studies can be made. Perhaps the most important of these are the completion of the historical Northern Hemisphere map project for the war years 1939 through 1945. With a complete and improved supply of historical data, further attempts can be made to obtain a more accurate picture than we now have of the normal state of the circulation at many levels, and to study such things as the characteristics of atmospheric waves and circulation indices. Clearly many other projects in addition to those mentioned can be suggested which will lead to better utilization of our existing supply of data.

In addition to utilizing existing data more effectively, meteorologists have the responsibility of seeing that future supplies of hemispheric data are improved. This improvement appears to be more necessary in quality rather than in quantity. Thus, observational networks should be chosen so that there is a more even distribution over the globe. Observations, including winds and moisture, should be taken from high atmospheric levels. The idea of the "polar year" or "aerological days" should be extended so that a global supply of data at all levels will be available for at least one year.

REFERENCES

1. Air Weather Service, *Northern Hemisphere Historical Weather Maps, Sea Level and 500 Millibar.* October 1945 to September 1946 (to be completed through December 1948). Washington, D. C.
2. Berggren, R., Bolin, B., and Rossby, C.-G., "An Aerological Study of Zonal Motion, Its Perturbations and Break-Down." *Tellus*, Vol. 1, No. 2, pp. 14–37 (1949).
3. Berry, F. A., Jr., Bollay, E., and Beers, N. R., ed., *Handbook of Meteorology.* New York, McGraw, 1945. (See pp. 451–453)
4. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1:1–22 (1944).
5. Bjerknes, V., and others, *Physikalische Hydrodynamik.* Berlin, J. Springer, 1933. (See pp. 668–702)
6. Bortman, R. S., "Empirical Corrections to Constant Absolute Vorticity Trajectories." *Trans. Amer. geophys. Un.*, 29:324–330 (1948).
7. —— *The Interrelationship of Wavelength, Amplitude, and Wind Speed in Upper-Air Flow Patterns.* Unpubl. Rep., Extended Forecast Sect., U. S. Weather Bureau, Washington, D. C., 1949.
8. Brooks, C. E. P., "The Variation of Pressure from Month to Month in the Region of the British Isles." *Quart. J. R. meteor. Soc.*, 52:263–276 (1926).
9. —— and others, "Upper Winds over the World." *Geophys. Mem.*, No. 85, 149 pp. (1950).
10. Charney, J. G., "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2 (1948).
11. —— "On a Physical Basis for Numerical Prediction of Large-Scale Motions in the Atmosphere." *J. Meteor.*, 6:371–385 (1949).
12. —— and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
13. Clapp, P. F., *Further Empirical Studies of Wave Motion in the Westerlies.* Unpubl. Rep., Extended Forecast Sect., U. S. Weather Bureau, Washington, D. C., 1948.
14. Cressman, G. P., "On the Forecasting of Long Waves in the Upper Westerlies." *J. Meteor.*, 5:44–57 (1948).
15. —— "Some Effects of Wave-Length Variations of the Long Waves in the Upper Westerlies." *J. Meteor.*, 6:56–60 (1949).
16. Defant, A., "Die Zirkulation der Atmosphäre in den gemässigten Breiteu der Erde." *Geogr. Ann., Stockh.*, 3:209–266 (1921).
17. Deutscher Wetterdienst, *Current Daily Sea-Level and 500-Mb Analyses.* Bad Kissingen, 1949 *et seq.*
18. Dines, W. H., "The Characteristics of the Free Atmosphere." *Geophys. Mem.*, 2:45–76 (1919).
19. Dorsey, H. G., Jr., and Brier, G. W., "An Investigation of a Trajectory Method for Forecasting Flow Patterns at the Ten-Thousand-Foot Level." Res. Paper FR-08 in *A Collection of Reports on Extended Forecasting Research.* U. S. Weather Bureau, Washington, D. C., 1944.
20. Dryden, H. L., "A Review of the Statistical Theory of Turbulence." *Quart. appl. Math.*, 1:7–42 (1943). (See p. 9)
21. Elliott, R. D., and Smith, T. B., "A Study of the Effects of Large Blocking Highs on the General Circulation of the Northern-Hemisphere Westerlies." *J. Meteor.*, 6:67–85 (1949).

22. Flohn, H., "Grundzüge der allgemeinen atmosphärischen Zirkulation auf der Südhalbkugel." *Arch. Meteor. Geophys. Biokl.*, (A) 2:17–64 (1950).
23. Fultz, D., "Upper-Air Trajectories and Weather Forecasting." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 19 (1945).
24. Hess, S. L., "Some New Mean Meridional Cross Sections through the Atmosphere." *J. Meteor.*, 5:293–300 (1948).
25. Jacobs, W. C., "Sources of Atmospheric Heat and Moisture over the North Pacific and North Atlantic Oceans." *Ann. N. Y. Acad. Sci.*, 44:19–40 (1943).
26. —— "The Energy Acquired by the Atmosphere over the Oceans through Condensation and through Heating from the Sea Surface." *J. Meteor.*, 6:266–272 (1949).
27. Jeffreys, H., "On the Dynamics of Geostrophic Winds." *Quart. J. R. meteor. Soc.*, 52:85–104 (1926).
28. Joint Meteorological Committee, *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere, Sea Level, 1899–1939*. Washington, D. C.
29. —— *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere, 3,000 Dynamic Meters, 1933–1940*. Washington, D. C.
30. —— *Normal Weather Maps, Northern Hemisphere, Upper Level*. Washington, D. C., 1944.
31. Klein, W. H., "The Unusual Weather and Circulation of the 1948–49 Winter." *Mon. Wea. Rev. Wash.*, 7:99–113 (1949).
32. Lettau, H., "Luftmassen und Energieaustausch zwischen niederen und höhen Breiten der Norhalbkugel während des Polarjahres 1932–1933." *Beitr. Phys. frei. Atmos.*, 23:45–84 (1936).
33. Namias, J., *Extended Forecasting by Mean Circulation Methods*. U. S. Weather Bureau, Washington, D. C., 1947.
34. —— "Evolution of Monthly Mean Circulation and Weather Patterns." *Trans. Amer. geophys. Un.*, 29:777–788 (1948).
35. —— "The Index Cycle and Its Role in the General Circulation." *J. Meteor.*, 7:130–139 (1950).
36. —— and Clapp, P. F., "Studies of the Motion and Development of Long Waves in the Westerlies." *J. Meteor.*, 1:57–77 (1944).
37. —— "Normal Fields of Convergence and Divergence at the 10,000-Foot Level." *J. Meteor.*, 3:14–22 (1946).
38. —— "Confluence Theory of the High Tropospheric Jet Stream." *J. Meteor.*, 6:330–336 (1949).
39. Pagava, S. T., *Basic Principles of the Synoptic Method of Long-Range Weather Forecasting*, 3 Vols. Leningrad, Hydrometeor. Publ. House, 1940. (Translation prepared by Weather Information Service, Headquarters, Army Air Forces.)
40. Page, L. F., and others, "Reports on Critical Studies of Methods of Long-Range Weather Forecasting." *Mon. Wea. Rev. Wash.*, Supp. 39 (1939).
41. Priestley, C. H. B., "Heat Transport and Zonal Stress between Latitudes." *Quart. J. R. meteor. Soc.*, 75:28–40 (1949).
42. Rossby, C.-G., "Planetary Flow Patterns in the Atmosphere." *Quart. J. R. meteor. Soc.*, 66 (Supp.):68–87 (1940).
43. —— "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2:187–204 (1945).
44. —— and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
45. Rossby, C.-G., and Willett, H. C., "The Circulation of the Upper Troposphere and Lower Stratosphere." *Science*, 108:643–652 (1948).
46. Serra, A. B., *Atlas de Meteorologia, 1873–1909*. Rio de Janeiro, Serviço de Meteorologia e Conselho Nacional de Geografia, 1948.
47. Shaw, N., *Manual of Meteorology. Vol. II—Comparative Meteorology*, 2nd ed. Cambridge, University Press, 1936.
48. Starr, V. P., "An Essay on the General Circulation of the Earth's Atmosphere." *J. Meteor.*, 5:39–43 (1948).
49. Tannehill, I. R., *Draft Notes on Weather of Future Years*. Mimeogr. Rep. U. S. Dept. of Commerce, Weather Bureau, Washington, D. C., Part I, September 1944; Part II, May 1945.
50. U. S. Weather Bureau, *Normal Weather Maps, Northern Hemisphere, Sea-Level Pressure*. Washington, D. C., 1946.
51. —— *Daily Series, Synoptic Weather Maps, Northern Hemisphere, Sea-Level and 500-Millibar Charts with Synoptic Data Tabulations*. Washington, D. C., January 1949 *et seq.*
52. Walker, G. T., and Bliss, E. W., "World Weather, IV—Some Applications to Seasonal Foreshadowing." *Mem. R. meteor. Soc.*, 3:81–95 (1930).
53. Weightman, R. H., "Preliminary Studies in Seasonal Weather Forecasting." *Mon. Wea. Rev. Wash.*, Supp. 45 (1941).
54. Wexler, H., "Determination of the Normal Regions of Heating and Cooling in the Atmosphere by Means of Aerological Data." *J. Meteor.*, 1:23–28 (1944).
55. —— and Namias, J., "Mean Monthly Isentropic Charts and Their Relation to Departures of Summer Rainfall." *Trans. Amer. geophys. Un.*, 19:164–170 (1938).
56. Willett, H. C., *Descriptive Meteorology*. New York, Academic Press, 1944. (See pp. 131–135)
57. —— "Patterns of World Weather Changes." *Trans. Amer. geophys. Un.*, 29:803–809 (1948).
58. —— "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6:34–50 (1949).
59. —— and others, *Final Report of the Weather Bureau—M. I. T. Extended Forecasting Project for the Fiscal Year July 1, 1946–July 1, 1947*. Mimeogr., Cambridge, Mass., 1947.
60. —— *Final Report of the Weather Bureau—M. I. T. Extended Forecasting Project for the Fiscal Year July 1, 1948–June 30, 1949*. Mimeogr., Cambridge, Mass., 1949.
61. Yeh, T.-c., "On Energy Dispersion in the Atmosphere." *J. Meteor.*, 6:1–16 (1949).

APPLICATIONS OF ENERGY PRINCIPLES TO THE GENERAL CIRCULATION

By VICTOR P. STARR

Massachusetts Institute of Technology

INTRODUCTION

Theoretical hydrodynamics and thermodynamics furnish the basic equations of energy which in the end must describe the energy transformations which take place in the atmosphere. These equations in themselves are not capable of furnishing a sufficient rational explanation of the causes of atmospheric processes, but nevertheless provide a guide to systematic exploration for purposes of finding empirically important facts concerning the behavior of the atmosphere. Thus their utility is much enhanced if consideration is given to observational data.

The most important problem which confronts us in such an effort is therefore not one of merely stating the several pertinent equations in a formally complete manner, but rather one of discussing atmospheric processes as given by observations in terms of these relationships. To this end the principles involved must be moulded and recast in such a form as to permit the desired applications to be made. In this procedure the failure to give proper cognizance to the special circumstances characteristic of the atmospheric processes dealt with can only lead to endless complications and needless confusion.

Much of what has been written concerning this subject has been deficient in two respects. In the first place, investigators have been prone to lump together various diverse forms of energy, thereby losing the advantages to be gained from the fact that each form of energy is produced from and converted to other forms in its own characteristic fashion, permitting individual study. Likewise for each form there exist specific modes of transfer and redistribution. Unless these specific characteristics are subject to scrutiny in detail, only very broad generalizations can be reached. Even here, however, the implications of the balance of total energy for the globe have not yet been studied in sufficient detail as will be discussed later.

In the second place, the modes of energy transfer within the atmosphere are so effective that no feature such as a cyclone can be treated independently without due allowance for exchanges of energy between it and the remaining atmosphere. It is therefore inappropriate to treat such a feature as in any sense a closed system. Modern trends are beginning to give proper cognizance to this circumstance, although much of the too restricted point of view permeates meteorological thought.

In the present discourse only certain phases of the subject are discussed by way of illustrating a general approach. Thus the discussion which follows treats only the global balance of kinetic energy as an example of the study of one individual form of energy. In the last section the total energy balance is re-examined. It is of course true, as has already been stated, that it is also possible to study the global balance of other individual forms of energy such as geopotential and internal heat energy. A beginning in this direction has been made by Van Mieghem [10].

GLOBAL BALANCE OF KINETIC ENERGY

General Considerations. One of the basic problems in the science of meteorology relates to the manner in which thermal energy received by the atmosphere through short-wave solar radiation becomes in part transformed into kinetic energy of motion relative to the rotating earth. Plausible estimates show that the fraction of the total energy so transformed is very small, but must nevertheless be sufficient to account for all air motions, in the absence of any other significant energy sources. Since the kinetic energy of organized motions is continually degraded and ultimately dissipated by turbulence and viscosity, the process of kinetic energy production must be a continuous one with, probably, certain fluctuations about a mean rate when the whole atmosphere is considered. The purpose of this discussion is to examine this production process from a hydrodynamical point of view.

Changes in the kinetic energy of a particle or system of particles can result only from the action of mechanical forces, and hence the rate of kinetic-energy production can be discussed in terms of the joint action of such forces and the kinematics of existing motions. In this light it is not essential to inquire how systems of such forces and such motions in the atmosphere are related to the thermodynamical processes which are ultimately responsible for their existence. In order to demonstrate the particular point in question as simply as possible we shall first consider an example of fluid motion under circumstances which are somewhat artificial, but which still have theoretical interest. In view of the fact that the kinetic energy of vertical motions in the atmosphere is very small compared with the kinetic energy of the large-scale horizontal motions we shall consider only the latter.

The approach used is one suggested by the beautiful classic paper of Osborne Reynolds [6] entitled "On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion." Since Reynolds was concerned only with the dissipation of kinetic energy, his treatment must be modified in order to envisage also the process which creates kinetic energy. For this reason his assumption of incompressibility will be abandoned. Also, our restriction to the study of the

kinetic energy of horizontal motions introduces certain changes, although these changes are not actually in the nature of approximations.

Study of a Simple System. Let it be supposed that a mass of gas is confined in a chamber with a plane bottom and vertical walls, under the action of gravity which we assume to be acting vertically downward. If the chamber is of sufficiently great height, it is not necessary that it have a top. Likewise, the gas need not be an ideal one, since for the time being no use will be made of an equation of state. Coriolis forces will, for the present, be omitted. Let it be supposed further that the gas is in some state of motion induced by differential heating and cooling.

If we take x, y, and z to be a Cartesian coordinate system with the positive z-axis vertical, we may write the equations of motion for the horizontal directions in the form

$$\begin{aligned} \frac{du}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + F_x, \\ \frac{dv}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} + F_y. \end{aligned} \tag{1}$$

Here u and v are the velocity components in the directions x and y; ρ is the density; p the pressure; t time; and F_x and F_y are the components of the viscous forces in the x and y directions. Generally speaking, the motions in the chamber might be turbulent. If we wish to regard the dependent variables in equations (1) as representing mean values free of the turbulence components, we shall assume that the only change necessary is to include eddy-stress effects in the quantities F_x, F_y after the manner of Reynolds. More will be said concerning this point later.

The kinetic-energy equation corresponding to the system (1) is

$$\begin{aligned} \rho\frac{\partial}{\partial t}\frac{V_h^2}{2} + \rho u\frac{\partial}{\partial x}\frac{V_h^2}{2} + \rho v\frac{\partial}{\partial y}\frac{V_h^2}{2} \\ + \rho w\frac{\partial}{\partial z}\frac{V_h^2}{2} = -\left(u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y}\right) - d. \end{aligned} \tag{2}$$

We use the symbol d to represent the rate at which the turbulence and viscosity are decreasing the kinetic energy per unit volume, and $V_h^2 = u^2 + v^2$. It is possible to rewrite (2) in the following form:

$$\begin{aligned} \frac{\partial E}{\partial t} + \frac{\partial Eu}{\partial x} + \frac{\partial Ev}{\partial y} + \frac{\partial Ew}{\partial z} \\ = -\left(\frac{\partial pu}{\partial x} + \frac{\partial pv}{\partial y}\right) + p\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) - d, \end{aligned} \tag{3}$$

where use has been made of the continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} + \frac{\partial \rho w}{\partial z} = 0, \tag{4}$$

which in any case must be true, and where $E = \frac{1}{2}\rho V_h^2$ is the horizontal kinetic energy per unit volume. The quantity represented by the last three terms on the left-hand side of (3) is the divergence of the (three-dimensional) kinetic energy transport vector $E\mathbf{V}$. The quantity in the first parenthesis on the right is the divergence of the horizontal vector $p\mathbf{V}_h$. If equation (3) is integrated over an arbitrary volume, both of these quantities may be represented as surface integrals with the aid of the divergence theorem. Thus, if the limits are fixed, we may write

$$\begin{aligned} \frac{\partial}{\partial t}\iiint E\,dx\,dy\,dz \\ = \int EV_n\,dS - \iint p(v\,dx - u\,dy)\,dz \\ + \iiint p\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)dx\,dy\,dz \\ - \iiint d\,dx\,dy\,dz, \end{aligned} \tag{5}$$

where V_n is taken to be the inward component of velocity at the boundary, and dS is a surface element.

Equation (5) may now be given the following interpretation. The total horizontal kinetic energy (T) in a fixed region may be changing in consequence of:

1. An advection of new fluid having kinetic energy across the boundary. This is represented by the term $A = \int EV_n\,dS$. This is then one mode of *redistribution* of kinetic energy.

2. The performance of work by pressure forces at the boundary in virtue of the displacements due to the horizontal velocity components. This is represented by the term $W = -\iint p(v\,dx - u\,dy)\,dz$. This is a second mode of *redistribution* of kinetic energy.

3. A production of kinetic energy within the volume itself. This is represented by the term

$$S = \iiint p\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)dx\,dy\,dz,$$

which contains the *primary source* of kinetic energy.

4. The action of frictional forces. This effect would ordinarily consist of a dissipation and is represented by the term $D = \iiint d\,dx\,dy\,dz$.

If the limits of integration include all of the fluid in the fixed chamber, it is clear that the surface integrals must vanish, so that in a *mechanically closed* system (5) reduces to

$$\frac{\partial K}{\partial t} = S - D. \tag{6}$$

Since for such a system the frictional effect would ordinarily lead to dissipation, it follows that S must be positive if the total horizontal kinetic energy K is to remain constant or increase. If a more or less constant amount of kinetic energy is to be present, the dissipation must be balanced by a corresponding positive average rate of production.

The production S may be looked upon as the integral of the contributions from the various horizontal layers of fluid present and written as

$$S = \int \left[\iint p\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) dx\, dy \right] dz. \tag{7}$$

In view of the fact that the surface integral

$$\iint \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) dx\, dy$$

must vanish if the horizontal velocity is zero across the fixed walls, it follows that a given horizontal stratum of fluid cannot give a positive contribution to S unless larger values of the pressure p are associated with areas of horizontal divergence than are associated with areas of convergence. Thus areas of *horizontal* divergence represent primary kinetic energy sources, while areas of convergence represent sinks for kinetic energy. Furthermore, in a mechanically closed system of the kind here considered it is impossible to have source regions for kinetic energy without at the same time having sinks of a hydrodynamic nature, entirely independent of frictional effects.

Equations for the Atmosphere. Before embarking upon a discussion of the meteorological implications of the material presented above, it is desirable to develop the concepts involved in more general terms, so as to render it possible to perform integrations over the entire mass of the atmosphere.

To a sufficiently close degree of approximation the shape of the geopotential surfaces may be considered as spherical so that we may make use of spherical polar coordinates in which r is the radius, ϕ is latitude, and λ is longitude. By analogy with the Cartesian case we may then write the equations of motion for the horizontal directions (see Brunt [2]) in the form

$$\left.\begin{aligned} \frac{du}{dt} - \frac{uw}{r}\tan\phi + \frac{uw}{r} + 2\Omega(w\cos\phi - v\sin\phi) \\ = -\frac{1}{\rho}\frac{\partial p}{\partial x} + F_x, \\ \frac{dv}{dt} + \frac{u^2}{r}\tan\phi + \frac{vw}{r} + 2\Omega u\sin\phi \\ = -\frac{1}{\rho}\frac{\partial p}{\partial y} + F_y, \end{aligned}\right\} \tag{8}$$

where u, v, and w are the linear velocity components in the eastward, northward, and upward directions, respectively; x and y are measures of linear distance eastward and northward, respectively; and Ω is the angular velocity of the earth. The analogous energy equation in this case may be written as

$$\begin{aligned} \rho\frac{d}{dt}\frac{V_h^2}{2} + \rho\frac{V_h^2}{r}w + 2\rho\Omega uw\cos\phi \\ = -\left(\frac{\partial pu}{\partial x} + \frac{\partial pv}{\partial y} - \frac{pv}{r}\tan\phi\right) \\ + p\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} - \frac{v}{r}\tan\phi\right) - d. \end{aligned} \tag{9}$$

If we make use of the observational fact that the last two terms on the left-hand side of (9) are of a very small order of magnitude, these terms will be dropped.[1] The manipulation of the remaining term on the left side may now be carried out with the aid of the continuity equation much as before, since this operation is independent of the specific coordinate system used, so that we may write

$$\frac{\partial E}{\partial t} + \operatorname{div}_3 E\mathbf{V} = -\operatorname{div}_2 p\mathbf{V}_h + p\operatorname{div}_2 \mathbf{V}_h - d. \tag{10}$$

A volume integral of (10) may now be taken and written in the form

$$\begin{aligned} \frac{\partial}{\partial t}\int E\, d\tau = \int EV_n\, ds - \iint p(v\, dx - u\, dy)\, dr \\ + \int p\operatorname{div}_2 \mathbf{V}_h\, d\tau - \int d\, d\tau, \end{aligned} \tag{11}$$

where $d\tau$ is a volume and ds a surface element. Equation (11) is physically identical with (5) and has, therefore, the same interpretation. In symbolic form we may write

$$\frac{\partial K}{\partial t} = A + W + S - D, \tag{12}$$

which states that the rate of increase of horizontal kinetic energy for a fixed volume is equal to the net rate of advection of such kinetic energy into the region, plus the rate at which work is being done by the surroundings on the fluid in the region through horizontal motions, plus the production of kinetic energy in the volume, minus the frictional dissipation. For a system which is mechanically closed, A and W again vanish. This is therefore true when the entire atmosphere is considered. In this case the surface integral of the horizontal divergence over each closed geopotential surface must vanish as in the case of the chamber previously considered.

Although it is possible to form other energy integrals for fluid motion, as pointed out in standard texts on hydrodynamics (*e.g.*, [1]), the particular merit of the procedure followed above is that the expression for production of kinetic energy assumes a form which is of interest in meteorological problems. The implications of equation (12) may be stated in brief as follows:

1. The intensity of the primary source of horizontal kinetic energy at a given point in the atmosphere is given by the product of the pressure and the divergence of the horizontal velocity.

2. Positive primary sources must always occur in combination with negative sources or sinks independ-

1. In reality these terms represent a conversion of kinetic energy of horizontal motions into kinetic energy of vertical motions, and as such do not involve a production of kinetic energy. Indeed, by methods similar to those used in this paper one can investigate separately the kinetic energy of motions in each of the three directions, namely, zonal, meridional, and vertical. In that case other conversion terms of a similar nature arise.

ent of frictional effects, when the entire atmosphere is considered.

3. In addition to the action of the sources and frictional effects, the horizontal kinetic energy in a fixed region not embracing the entire atmosphere may change due to advection of kinetic energy across the boundary and due to the redistribution of kinetic energy through the boundary by work done by pressure forces and horizontal velocity components at the boundary.

From the standpoint of the general circulation it would appear that the sources of kinetic energy are to be found in the regions of horizontal divergence. The net contribution from a given level results from the fact that areas of divergence generally occur at a different pressure than do the areas of convergence. Thus at lower levels it is common for horizontal divergence to be present in anticyclonic areas while convergence takes place in cyclonic areas, the net result being positive. We do not as yet have sufficient observational material concerning the distribution of divergence at higher levels, but the fact that the pressure decreases with elevation would seem to indicate that the importance of the higher levels rapidly diminishes. Generally speaking, it would thus appear that the energy sources for the general circulation are to be found principally in the subtropical high-pressure cells, the migratory polar anticyclones, and the subsiding cap of cold air over the polar regions. From these primary centers the kinetic energy is continually transferred to the cyclonic areas with convergence which act as sinks in addition to the action of friction.

One might ask why it is that if the diverging anticyclones act as primary sources of kinetic energy, they are not the scenes of major activity. Actually, however, the generation process cannot be present in such systems without the simultaneous operation of the transfer processes. If divergence exists in an anticyclone, the peripheral outward motion results in a rapid outward flow of kinetic energy through work done by pressure forces and through advection.

It was pointed out that a transfer of kinetic energy of horizontal motions across the boundary of a region which is not mechanically closed may be brought about by advection of existing kinetic energy and through the work done by pressure forces in virtue of the components of horizontal velocity across the boundary. Thus, if one considers a symmetrical polar cap extending from the north pole to some middle latitude ϕ and embracing the entire vertical extent of the atmosphere, the expression for the transfer of kinetic energy across the vertical southern boundary at the latitude ϕ is

$$\int (\tfrac{1}{2}\rho V_h^2 v + pv)\, ds \approx \int pv\, ds = \frac{R}{m} \int \rho T v\, ds, \quad (13)$$

where ds is an element of area, R/m is the gas constant for air, and T is the absolute temperature, while the other symbols have the same significance as previously. The approximate equality of the first two integrals is based upon the fact that the advection of kinetic energy in the atmosphere is of a smaller order of magnitude than the contribution of the term pv except possibly at very high levels. The final form depends also upon the feasibility of applying the ideal equation of state to the atmosphere. If these simplifications are accepted, the following observations may be made.

The last integral is proportional to the advection of internal heat energy northward, and hence is in all probability positive. This would indicate that there is normally a poleward flow of kinetic energy across middle latitudes from the tropics and subtropics which apparently serve as important source regions for such energy. Since this flow must cease as the polar regions are approached, it follows that the cyclone belts in middle and polar latitudes serve as dissipative mechanisms for this kinetic energy through friction and through the horizontal convergence present in them.

It is a matter of common synoptic experience that an extratropical cyclone is more apt to intensify if there is a relatively large contrast in the heat advection on its eastern and western sides. According to the present discussion, it is not essential that the increase of kinetic energy in such cases be produced *in situ* through conversion from other forms of energy. The intensification may be brought about through the increased local poleward transport of kinetic energy from the general source region in lower latitudes, as measured by the large net local heat transport poleward.

We have made the tacit assumption in the development given above that the "frictional" term D leads to a dissipation of kinetic energy. If only molecular viscosity and small-scale turbulent viscosity are included in this term, the assumption is undoubtedly valid. However, if relatively large-scale eddies and other large features of the atmospheric motions are included in the form of a gross turbulence as distinguished from the remaining mean motion, it is apparent that the quantity D may then embrace energy-producing systems and it is possible that it may change sign. Thus, for example, if only the average zonal circulation of the atmosphere be considered as the true mean motion so that the cyclones, anticyclones, and other nonzonal motions appear as turbulence, there is no clear a priori reason for assuming that the term D represents a dissipation.

Finally, it is interesting to compare the results obtained here with those of Margules [5] in his classic paper, "On the Energy of Storms." Very broadly speaking, the two approaches deal with essentially the same process. We have simply enlarged the "chamber" containing the gas used by Margules so as to include the whole atmosphere. Furthermore, whereas Margules considered a discrete process, we have replaced it by a continuous one and restricted our attention to the production, redistribution, and dissipation of kinetic energy of horizontal motions only. Also, we have recognized that under these circumstances the pressure multiplied by the *horizontal* divergence is the measure of the rate at which other forms of energy such as potential and internal energy are being converted into kinetic

energy.[2] When the divergence is negative the sense of this conversion process is reversed. It should be noted that this particular result is independent of the physical nature of the "working substance," which might indeed be partly liquid (or even solid), with the gaseous and liquid components undergoing changes of phase. The result therefore automatically embraces the consequence of all condensation phenomena insofar as they contribute to the horizontal kinetic energy.

GLOBAL BALANCE OF TOTAL ENERGY

Basic Equations. Thus far we have found it convenient to deal with the kinetic energy problem alone, since this quantity can be changed only by mechanical forces, and hence may be studied separately in terms of the systems of such forces considered as given by observational data. However the problems connected with the total global energy balance must in the end be of significance in the further understanding of atmospheric and oceanic circulations. For this reason we shall now attempt to formulate certain relationships involved in this more general subject.

Proceeding along more classical lines, let us consider the statement of the general physical energy equation written in the form

$$\rho \frac{dq}{dt} + \psi = \rho \frac{dU}{dt} + p\rho \frac{d\alpha}{dt} = \rho \frac{dU}{dt} - \frac{p}{\rho} \frac{d\rho}{dt}. \quad (14)$$

Here $\rho\, dq/dt$ is the rate of external heat addition per unit volume, U is the total internal energy per unit mass, $\alpha \equiv 1/\rho$ is the specific volume, ψ is the rate of generation of heat by friction per unit volume, while the other symbols have already been defined. With the aid of the continuity equation

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{c} = 0, \quad (15)$$

where $\mathbf{c}$ is the total vector particle velocity, we may write that

$$\rho \frac{dq}{dt} + \psi = \rho \frac{dU}{dt} + p \nabla \cdot \mathbf{c}. \quad (16)$$

We next proceed to evaluate the last term in (16) from the dynamical equation of motion written in vectorial form as follows:

$$\rho \frac{d\mathbf{c}}{dt} = -\nabla p - \rho \nabla \Phi - 2\rho \boldsymbol{\Omega} \times \mathbf{c} - \mathbf{F}, \quad (17)$$

where Φ is geopotential energy per unit mass, $\boldsymbol{\Omega}$ is the constant angular velocity of the earth's rotation, and $\mathbf{F}$ is the vectorial retarding force per unit volume due to friction.

The scalar product of (17) with the particle velocity yields the corresponding equation of energy which may be written after slight rearrangement as

$$\rho \frac{d}{dt} \frac{c^2}{2} = p \nabla \cdot \mathbf{c} - \nabla \cdot p\mathbf{c} - \nabla \cdot \rho \Phi \mathbf{c} + \Phi \nabla \cdot \rho \mathbf{c} - d. \quad (18)$$

In (18), c is the magnitude of $\mathbf{c}$, and $d \equiv \mathbf{c} \cdot \mathbf{F}$ is the rate at which work is done by the fluid against frictional forces per unit volume. Assuming that the geopotential Φ is constant with time at a fixed point with respect to the earth (this is true except for such things as the small tide-producing disturbances), we may write, with the aid of the continuity equation in the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{c} = 0, \quad (19)$$

that

$$\Phi \nabla \cdot \rho \mathbf{c} = -\frac{\partial}{\partial t} (\rho \Phi). \quad (20)$$

Using (18) and (20), we can now rewrite (16) in the form

$$\rho \frac{dq}{dt} + \psi - d = \rho \frac{d}{dt} \left(U + \frac{c^2}{2} \right) + \frac{\partial}{\partial t} (\rho \Phi) + \nabla \cdot (p + \rho \Phi) \mathbf{c}. \quad (21)$$

Since with the aid of (19) it follows that

$$\rho \frac{d}{dt} (\quad) = \frac{\partial \rho (\quad)}{\partial t} + \nabla \cdot \rho (\quad) \mathbf{c},$$

we finally have the equation

$$\rho \frac{dq}{dt} + \psi - d = \frac{\partial}{\partial t} \left(\rho U + \rho \frac{c^2}{2} + \rho \Phi \right) + \nabla \cdot \left(\rho U + \rho \frac{c^2}{2} + \rho \Phi + p \right) \mathbf{c}. \quad (22)$$

The various considerations which have entered into the formulation of equation (22) are true for any fluid medium without significant approximation. We may therefore apply the equation to the entire fluid envelope of the earth or portion of it, making no distinction between the atmosphere and the hydrosphere. We can thus integrate it over an equatorial belt between latitudes $-\phi$ and $+\phi$ and include all bodies of water such as the oceans, rivers, lakes, etc. Considering again the average conditions so that local time variations disappear we have, with the aid of the divergence theorem, that

$$H \equiv \int \left(\rho \frac{dq}{dt} + \psi - d \right) d\tau = \int \left(\rho U + \rho \frac{c^2}{2} + \rho \Phi + p \right) v_n \, ds, \quad (23)$$

where $d\tau$ is a volume element. Equation (23) has of course a very simple interpretation and could have indeed been written directly from general considerations. If we include under friction only the effects of

2. It is worthy of note that the present treatment gives no information as to whether the bulk of the kinetic energy generated in the atmosphere represents a conversion from geopotential energy or whether it represents a conversion directly from internal heat energy.

molecular viscosity or of small-scale disturbances which can produce no significant tangential stresses at the boundaries $-\phi$ and $+\phi$, the contribution of the term d in the integral on the left-hand side may be assumed to represent the mean rate of dissipation of kinetic energy into heat within the equatorial belt and therefore cancels the contribution of the term ψ. It follows therefore that H is the total net rate of heating of the air in the belt. Equation (23) simply states that this net incoming energy is transferred meridionally in the form of (1) internal energy ρU per unit volume, (2) kinetic energy of existing motions $\rho c^2/2$, and (3) potential energy $\rho\Phi$, as well as (4) through work done by pressure forces p. We shall refer to these four items as advective modes of energy transfer.

It is here assumed that there is no advection of energy through the surface of the lithosphere. This is essentially correct except for processes such as volcanism and seismological phenomena, but these are deemed to be too unimportant for the present considerations. Also it is assumed that there is no advection of energy through the top of the atmosphere, which therefore neglects the effect of interchange of molecules with astronomical space and of the mass accretions of meteoric origin. The quantity H, representing the net heat gain within the equatorial belt by processes other than advection, may be very closely identified with the net heat received through exchange of radiation with the extraterrestrial environment. This identification neglects such processes as conduction of heat from the interior of the earth, which is of appreciable importance only locally in connection with volcanism, and it also neglects heat liberated (or consumed) by net progressive chemical changes such as oxidation or photosynthesis processes. Net heat gain through exchange of radiation with other portions of the earth is likewise neglected. All the various corrections mentioned are however in all probability insignificant.

If we take H to be the net gain of heat through exchange of radiation with space, various necessarily crude estimates of this quantity have been prepared. A convenient arrangement of one set of such estimates has been presented by Bjerknes [1]. As is well known, the estimates give positive values of H for all choices of $\pm\phi$ between the equator and the poles with a maximum for about $\phi = \pm 45°$ latitude. It therefore follows that there must be an advective transport of energy poleward by the combination of terms indicated in (23), with a maximum at $\phi = \pm 45°$ latitude in the mean.

It is apparent that the important problem posed by the global energy balance concerns itself with the partition of the poleward energy transport among the several terms in the integrand of the right-hand member of equation (23).

Discussion. Unfortunately our observational information concerning the problem posed by the global energy balance is very sketchy and incomplete. We shall nevertheless endeavor to discuss such aspects of it as are possible with existing knowledge. In the first place, the contribution of the hydrosphere to the transfer integral is probably small (see Sverdrup [9]), but directed toward the poles. A reasonable estimate of this contribution would appear to be about ten per cent of H. Denoting this fraction by h, let us next turn our attention to the state of affairs within the atmosphere.

It has been previously pointed out that the advection of existing kinetic energy $\rho c^2/2$ meridionally is very small, relatively speaking. We are therefore again justified in omitting it. The internal energy U may be considered as being the sum of the internal heat energy and the latent heat of water vapor. Since there is assumed to be practically no net meridional mass transport in the atmosphere, it will suffice to assume that the internal heat energy is given by $c_v T$, c_v being the mean specific heat at constant volume. It thus follows that $U \approx c_v T + \epsilon L$, where ϵ is the specific humidity and L is the latent heat of condensation, assumed to be constant.[3] The term involving the work done by pressure forces may again be transformed according to the ideal equation of state, and finally combined with the internal heat-energy term using the relation between the specific heats of a gas. In the end (23) may be written in the form

$$H = h + \int (c_p T + \epsilon L + \Phi)\rho v_n \, ds, \tag{24}$$

where c_p, the specific heat at constant pressure, is assumed to have a constant mean value.

The contribution of the term involving the latent heat may be estimated from the mean excess of evaporation over precipitation in the equatorial belt. Using data of this kind given by Conrad [3], the writer has estimated that the magnitude of this effect is about one-half of H for an equatorial belt extending to $\pm 40°$ latitude. Let us denote this quantity by l.

The remaining terms may be examined as follows. If we write

$$\rho v_n = \overline{\rho v_n} + \{\rho v_n\}, \tag{25}$$

where $\overline{\rho v_n}$ is the average of ρv_n along the entire length of a closed latitude circle and $\{\rho v_n\}$ is the deviation from this average, it is clear that the identical vanishing of $\overline{\rho v_n}$ implies absence of closed mean meridional circulations, while its presence is required for the existence of such circulations. Here we neglect all topographic inequalities of the earth's surface. In view of the fact that Φ is constant along a latitude circle at any given elevation and that $\{\overline{\rho v_n}\}$ is zero, it follows that (24) may be rewritten in the form

$$H = h + l + \int c_p T\{\rho v_n\} \, ds + \int (c_p T + \Phi)\overline{\rho v_n} \, ds, \tag{26}$$

3. The latent heat as ordinarily discussed is the sum of the change in specific internal energy plus the work done in the expansion during evaporation. Strictly speaking, we are here concerned only with the first quantity, although the difference is not great enough to be of much significance.

where the last integral now represents the contribution of mean meridional circulations of the Hadley type to the poleward energy flux.

It should be remarked that, in a stable atmosphere, a meridional cell of the so-called direct type produces a poleward flow of energy, while one of the indirect type produces a flow in the opposite sense. This can easily be shown from the form of the last integral in (26).

Very preliminary estimates by the writer of the value of the third term on the right-hand side, made from geostrophic wind data for individual Northern Hemisphere maps for various levels seem to suggest that this contribution is somewhere in the vicinity of one-half of H at 40°N latitude. All in all it would thus appear that the contribution of the mean meridional circulations is small or even negative in middle latitudes, although very little reliance can be placed on the figures given or on the value of H obtained from the data given by Bjerknes. Suffice it to say that in all cases reasonable orders of magnitude are obtained, which in itself is somewhat encouraging.

Concluding Remarks. Most classical models for the general circulation of the atmosphere have followed along the lines originally proposed by Hadley [4] in that they assume the existence of large convectively driven closed circulations in meridional planes, at least in the average conditions. The development of the mean zonal motions is then ascribed to the effect of the earth's rotation on these primary circulations. In such a scheme the meridional circulations are a necessary mechanism in the production of kinetic energy. Also, according to this model the necessary meridional transport of angular momentum could be achieved if the poleward branches of the circulations carry more angular momentum than the returning ones at other levels.

For a number of reasons modern meteorologists have come to view models of the Hadley type with skepticism. A discussion of the basis for this current skepticism has been recently given by Rossby [7]. The writer rather inclines to the view that, although some mean meridional circulations in all probability do exist, their role in the energy balance and in the horizontal transport of angular momentum, at least in middle latitudes, may be overshadowed by the characteristics of other types of motion. Thus, following an original suggestion by Jeffreys, the writer has pointed out elsewhere [8] that the transport of angular momentum could be achieved through the observed properties of horizontal motions. This contention has since received a certain amount of corroboration by the observational studies of Widger [11].

The general views expressed in the present paper indicate that atmospheric meridional circulations likewise may not be essential for the global energy balance. Much more could be said if a more satisfactory appraisal were available for the magnitudes of the terms appearing in equation (26), since the values given are useful only for purposes of orientation. Further work in this direction is currently in progress at the Massachusetts Institute of Technology.

REFERENCES

1. Bjerknes, V., and others, *Physikalische Hydrodynamik*. Berlin, J. Springer, 1933.
2. Brunt, D., *Physical and Dynamical Meteorology*, 2nd ed. Cambridge, University Press, 1939.
3. Conrad, V., "Die klimatologischen Elemente und ihre Abhängigkeit von terrestrischen Einflüssen." *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. I, Teil B. Berlin, Gebr. Borntträger, 1936. (See pp. 360–362)
4. Hadley, G., "Concerning the Cause of the General Trade-Winds." *Phil. Trans. roy. Soc. London*, 39:58 (1735–36). Reprinted in "The Mechanics of the Earth's Atmosphere. A Collection of Translations by Cleveland Abbé," 3rd Collection. *Smithson. misc. Coll.*, Vol. 51, No. 4 (1910).
5. Margules, M., "Über die Energie der Stürme." *Jb. ZentAnst. Meteor. Wien*, Anh. (1903). Reprinted in "The Mechanics of the Earth's Atmosphere. A Collection of Translations by Cleveland Abbé," 3rd Collection. *Smithson. misc. Coll.*, Vol. 51, No. 4 (1910).
6. Reynolds, O., "On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion." *Phil. Trans. roy. Soc. London*, (A) 186:123–164 (1895). Reprinted in O. Reynolds, *Papers on Mechanical and Physical Subjects*, Vol. II. Cambridge, University Press, 1901. (See pp. 535–577)
7. Rossby, C.-G., "On the Nature of the General Circulation of the Lower Atmosphere" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed., pp. 16–48. Chicago, University of Chicago Press, 1949.
8. Starr, V. P., "An Essay on the General Circulation of the Earth's Atmosphere." *J. Meteor.*, 5:39–43 (1948).
9. Sverdrup, H. U., Johnson, M. W., and Fleming, R. H., *The Oceans*. New York, Prentice-Hall, Inc., 1942.
10. Van Mieghem, J., "Production et redistribution de la quantité de mouvement et de l'énergie cinétique dans l'atmosphère. Application à la circulation atmosphérique générale." *J. sci. Météor.*, 1:53–67 (1949).
11. Widger, W. K., Jr., "A Study of the Flow of Angular Momentum in the Atmosphere." *J. Meteor.*, 6:291–299 (1949)

MECHANICS OF PRESSURE SYSTEMS

Extratropical Cyclones *by J. Bjerknes* 577

The Aerology of Extratropical Disturbances *by E. Palmén* 599

Anticyclones *by H. Wexler* 621

Mechanism of Pressure Change *by James M. Austin* 630

Large-Scale Vertical Velocity and Divergence *by H. A. Panofsky* 639

The Instability Line *by J. R. Fulks* 647

EXTRATROPICAL CYCLONES

By J. BJERKNES

University of California, Los Angeles

With the consent of the editor the present article has been written as a summary of the research on extratropical cyclones in which the author himself has been directly involved. References to the work of others are therefore few, and the reader will not get a complete survey of the title subject.

This article is composed of two parts. In the first, extratropical cyclones are treated as simplified models for the sake of clarifying the theoretical principles. In the second, these principles are applied to a real storm over North America whose development may be considered as the prototype of a simple life history of extratropical cyclones. The modifications of that life history, caused by the varying initial conditions and the influence of neighboring systems in the general circulation, are treated by Dr. E. Palmén in his contribution to the Compendium.

DYNAMICS OF SIMPLIFIED CYCLONE MODELS

Theory of Pressure Changes and Thermal Structure of Extratropical Cyclones. The fully developed extratropical cyclone consists of a counterclockwise[1] vortex which extends upward into a wave trough in the upper westerlies. The dynamics of the extratropical cyclone is therefore a composite one, combining the dynamic phenomena of the vortex and the wave. We will here state separately the essential features of the atmospheric vortex and the atmospheric wave and then proceed to describe the composite dynamics of the extratropical cyclone.

In analyzing the displacement, intensification, and weakening of vortices and waves it is useful to consider the accompanying pressure changes, which obey the "tendency equation,"

$$\left(\frac{\partial p}{\partial t}\right)_h = -\int_h^\infty g \operatorname{div}_H (\rho \mathbf{v}) dz + (g\rho v_z)_h. \qquad (1)$$

Expressed in words, the rate of pressure change with time at a fixed point at the level h is determined partly by the net horizontal inflow into the vertical unit air column from h to the top of the atmosphere and partly by the vertical inflow of air through the base of that column.

A circular cyclonic vortex with vertical axis centered at the pole of a planet without mountains represents the simplest case of atmospheric vortex dynamics. In the case of frictionless motion in such a polar vortex the particles could be kept in steady-state zonal motion from west to east. The horizontal divergence is then everywhere zero and no vertical motion occurs, so that the tendency equation must indicate zero local pressure change at all points. With friction against the ground, the flow in the lowest part of the atmosphere would be given a component of indraft towards the vortex center, and this horizontal convergence of mass would make the pressure rise in the central portion of the pressure minimum, thus decreasing the zonal air motion. No steady state would be reached until the flow at the ground and the horizontal pressure gradient at the ground have reached zero. If the central core of the vortex is colder than its environment, there would still be a pressure gradient towards the pole in the free atmosphere, and there the air may continue its west to east circulation without horizontal divergence. This picture corresponds rather well to reality as represented by the time-averaged motion in the arctic region: almost zero meridional pressure gradient and zero zonal motion at the ground, and increasing poleward pressure gradient with height, accompanied by increasing westerlies with height. The initial assumption of a cyclone at the pole, and the additional assumption of friction at the ground, thus lead to the dynamical prediction that the cyclone at the ground should eventually disappear, while in the free atmosphere it should be conserved. This behavior of the circular cyclonic vortex can be generalized to apply also at other latitudes; the simple circular vortex is liable to die out gradually at the ground because of friction.

The circular cyclonic vortex centered in middle latitudes does not represent a steady-state dynamic system even in the absence of friction at the ground. Although the horizontal pressure gradient may everywhere be directed towards the center and its intensity may be a function only of the distance from the center, the motion around the center cannot be a simple circular one, because the Coriolis parameter varies from the northern to the southern part of the vortex. As a result of the variation of the Coriolis parameter the wind will be stronger in the southern than in the northern part of the vortex, and the net air transport across a north-south median wall will be from the western to the eastern half of the vortex. Consequently the pressure will rise in the eastern half because of horizontal convergence and fall in the western half of the low pressure system because of horizontal divergence, so that the pressure minimum and the accompanying vortex will drift westward. Eccentricity of the pressure field of such a sense as to involve a stronger pressure gradient in the southern than in the northern part of the vortex may reduce, neutralize, or even reverse that drift. The dynamic theory for the eccentric vortex has been developed in approximate form

1. All references to the sense of rotation in this article apply to the Northern Hemisphere.

by Holmboe [3]. He defined a "critical eccentricity" which would balance the exchange of air between the two halves of the vortex. Values of the quantity $|v| - |v'| - 2c = 4\Omega a\sigma_p^2 \cos\phi$, evaluated in a narrow isobaric channel of critical eccentricity, are given in Table I (v and v' are the wind velocities at southernmost and northernmost points, respectively, of the isobaric channel, c is the eastward speed of displacement of the vortex, Ω the angular speed of the earth, a the earth's radius, σ_p the angular radius of the isobaric channel, and ϕ the geographical latitude). In the case of the stationary vortex, the exchange of air between the eastern and western halves of the vortex is balanced if the wind velocity in the southernmost point of the isobaric ring exceeds that in the northernmost point by the tabulated amount. Near the center the flow can be almost constant all around the isobaric ring, but the greater the radius the more will the west wind in the south have to exceed the east wind in the north, particularly in low latitudes.

TABLE I. VALUES OF $|v| - |v'| - 2c = 4\Omega a\sigma_p^2 \cos\phi$ FOR CRITICAL ECCENTRICITY (in m sec^{-1})

ϕ	Angular radius of isobaric channel			
	1°	5°	10°	20°
90°	0	0	0	0
80°	0.1	2.5	9.9	—
70°	0.2	4.9	19.4	78
60°	0.3	7.1	28.4	114
50°	0.4	9.1	36.6	146
40°	0.4	10.9	43.6	175
30°	0.5	12.3	49.4	198

In applying the table to a moving vortex, twice the speed of displacement of the vortex must be added to the tabulated speed to give $|v| - |v'|$. For the eastward moving vortex the critical eccentricity is thus stronger than for the stationary one. Even slightly greater eccentricity would be needed to bring about accumulation of air in the western half of the pressure minimum and depletion of air in the eastern half, a condition which would seem necessary to make the system move eastward. A check with measured eccentricities shows, however, only "subcritical" cases; in other words all observed cyclonic vortices accumulate air in their front parts at the expense of the rear parts. The pressure change in such a moving vortex can therefore be explained only if other flow patterns prevail above the vortex. This conclusion is corroborated by the experience of synoptic aerology, and the typical upper flow pattern above moving vortices is that of the atmospheric wave. In the composite extratropical cyclone, the vortex part resists the eastward motion by piling up air in the front half; this resistance increases the faster the vortex is forced to move.

The atmospheric wave[2] superimposes a quasi-horizontal oscillation upon the fundamental current of straight westerlies. This is accompanied by a periodic distribution of horizontal mass divergence, which in first approximation depends on the relative strength of the "curvature and latitude effects" upon the wind speed. The curvature effect makes the air move supergeostrophically while overtaking the wave crest and subgeostrophically while overtaking the wave troughs. The latitude effect upon the wind speed comes from the fact that each flow channel is in a higher latitude at the anticyclonic bend than it is at the cyclonic bend, so that, the horizontal pressure gradients being equal, the geostrophic wind would be stronger at the cyclonic than at the anticyclonic bend. The result of

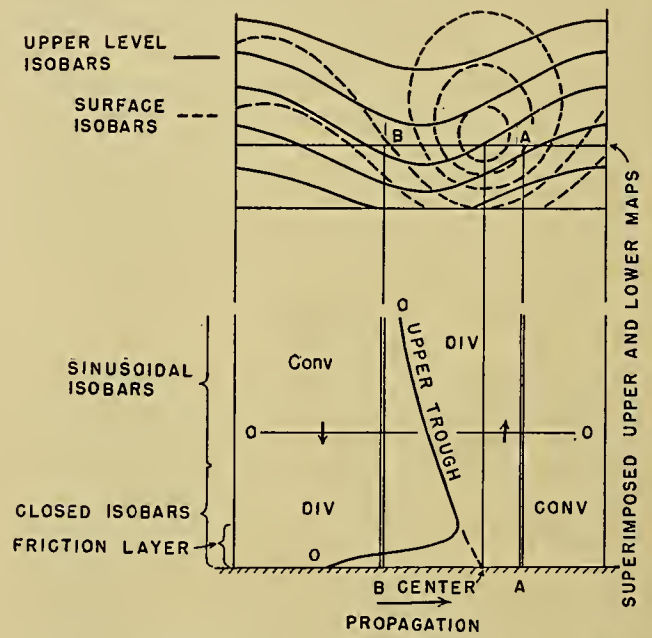


FIG. 1.—Schematic model of cyclone with lower-vortex and upper-wave part. West-to-east vertical profile shows location of horizontal divergence and convergence of mass.

these two opposite effects on horizontal divergence is in favor of the curvature effect when wave lengths are short and the fundamental current is strong. In this case the wave crests are preceded by horizontal convergence and the wave troughs by horizontal divergence. In the case of long waves and/or a weak fundamental current, the opposite distribution of horizontal divergence is established. The same conditions are found in air layers which move eastward more slowly than the wave.

In the usual baroclinic westerly current of middle latitudes a wave extending from the slow-moving lower layers to the fast-moving upper layers would have opposite patterns of horizontal divergence in the upper and lower part (see Fig. 1). At the level of transition between the upper and lower pattern a wave motion with zero horizontal divergence will exist. This "level of nondivergence" will be found at the height where the speed of the undisturbed current, v_x^*, is given by

2. The fundamental properties of the atmospheric waves of synoptic meteorology were first sketched by J. Bjerknes [2] and later developed mathematically by Rossby [26, 27] and Haurwitz [10]. In this article we follow the more recent treatment by J. Bjerknes and Holmboe [3].

$$v_x^* = c + \frac{2\Omega a \cos^3 \phi}{n^2}. \quad (2)$$

In this equation c is the eastward speed of propagation of the wave, Ω is the angular velocity of the earth, a is the earth's radius, ϕ is the geographical latitude, and n is the wave number per circumference of the earth. Numerical values of $(2\Omega a \cos^3 \phi)/n^2$ are given in Table II. This table, in conjunction with equation (2), shows that in short waves, such as are found to accompany the individual traveling cyclone, the level of nondivergence lies at an elevation where the undisturbed current moves only a little faster than the wave. In long waves the air at the level of nondivergence moves eastward much faster than the wave itself, particularly in low latitudes.

TABLE II. VALUES OF $(2\Omega a \cos^3 \phi)/n^2$ (in m sec^{-1})

ϕ	Wave length (deg. long.)				
	180°	120°	60°	36°	18°
70°	9.3	4.1	1.0	0.4	0.1
60°	29.0	12.9	3.2	1.2	0.3
50°	61.6	27.4	6.8	2.5	0.6
40°	104.3	46.3	11.6	4.2	1.0
30°	150.7	67.0	16.8	6.0	1.5

The level of nondivergence may be determined from sets of aerological maps, with the aid of Table II. Its height differs from case to case, but according to Charney [4] and Cressman [5] it averages around 600 mb for both long and short waves. Hence, with a given model of baroclinic westerlies, the speed v_x^* of the undisturbed westerlies at the level of nondivergence is approximately the same parameter for long and short waves. The speed of all such waves, which are superimposed on the same westerly current, therefore varies with wave length according to the formula,

$$c = v_x^* - \frac{2\Omega a \cos^3 \phi}{n^2}. \quad (3)$$

Short waves (large n) move almost with the speed of the air at the level of nondivergence. Long waves move eastward more slowly than short ones, and may also retrograde. Table II, applied to the case $c = 0$, gives us a survey of the wind at the level of nondivergence in stationary waves. At 70° latitude the west wind must be quite light for waves to be stationary, and the 180° wave length seems to be the most likely one for standing waves. Proceeding to lower latitudes, we find that the 180° stationary wave requires stronger westerlies than are ever known to occur. Assuming that v_x^* would never be greater than 30 m sec^{-1}, we see that below 60° latitude the 180° stationary waves would never occur, below 50° the 120° stationary waves also become impossible, and so on. This dependence of the long-wave pattern on geographical latitude usually leads to the establishment of only two or three standing waves per earth's circumference near the pole and stationary patterns with higher wave numbers in lower latitudes. In the latitudes of pattern transitions, complicated cases of wave interference occur.

The waves in the westerlies associated with the moving extratropical cyclones are by necessity of short wave lengths, say thirty degrees longitude. According to (3) and Table II, such waves move with a speed c, only slightly smaller than v_x^*, the speed of the westerlies at the level of nondivergence.

Figure 1 shows the position of the pressure minimum at sea level relative to that of the upper trough. The axis of minimum pressure of the closed low tilts towards the coldest side, which is usually to the west or northwest of the location of the surface center. The tropospheric part of the upper trough is also displaced westward with height, but not as far per unit height as the subjacent center of low pressure.

The friction layer of the closed vortex (up to about 1-km height) has horizontal convergence. Above the influence of surface friction the eastern half of the vortex has horizontal convergence of mass and the western half, horizontal divergence. This holds true also for the upper trough up to the level of nondivergence, beyond which divergence and convergence exchange positions. The tendency equation (1) applied to the schematic cyclone cross section of Fig. 1 gives an answer to the two questions: How can the cyclone move eastward as most middle latitude cyclones do? and, How can it deepen despite the frictional convergence?

The eastward displacement of the cyclone is assured if the vertical integral of horizontal mass divergence in the tendency equation is determined as to sign by the atmosphere above the level of nondivergence. The deepening of the pressure minimum likewise depends on the influence from above the level of nondivergence. Because of the westward tilt of the axis of the cyclone a vertical air column located at the surface center will show horizontal convergence in its lower portion, where it passes through the forward part of the vortex, and horizontal divergence where it traverses the upper wave pattern east of the wave trough. Deepening of the surface center will occur only if this upper-air divergence overcompensates the low-level convergence.

In all parts of the cyclone the surface pressure tendency represents a small change in the weight of the local vertical column, resulting from the difference between accumulation and depletion of air, each of which represents much greater weight changes. The natural adjustment of the pressure tendencies to the observed moderate values can be visualized as follows. In the low-level vortex, the convergence in the front and the divergence in the rear are more strongly developed the faster the vortex is forced to move. We have also seen from Table II that the level of nondivergence in the upper wave rises to a higher elevation, and the divergence values above that level decrease, when the wave speed increases. Therefore, a supposed increase in speed without a change in the structure of the cyclone would lead to a weakening of the high-level contribution and a strengthening of the low-level contribution to the change in weight of air columns. This would be tantamount to a decrease in pressure tendencies. Quite analogously, it can be shown

that a supposed slowing down of the cyclone without change in its structure would lead to increasing surface pressure tendencies. From this it can be concluded that the speed of a given cyclone is stable as long as its total three-dimensional structure remains the same. The pressure tendencies are then also stable although they are made up as small differences of large opposite contributions of horizontal divergence and convergence.

A real increase in the divergence effects of the upper wave would come from a lowering of the level of nondivergence and an inherent increase of that part of the atmosphere in which the air current is supercritical. According to (3), this may take place through one of the following changes of parameters in the upper wave: (1) an increase of the speed v_x of the upper westerlies, (2) a decrease of angular wave length $2\pi/n$, and (3) travel towards higher latitudes. In all these cases the compensating mass-divergence effects from low levels automatically become greater as the speed of the cyclone increases. The occurrence of excessive values of barometric tendencies is thus automatically avoided.

The slowing down which is normally observed in deep and extensive cyclones is associated with the great depth of atmosphere moving in a closed cyclonic flow pattern. If that pattern has a subcritical eccentricity, which is the more frequent case, it will maintain mass convergence in the eastern half and mass divergence in the western half. The influence of an upper wave pattern can then only barely overcompensate the divergence effects below, and the resulting pressure tendencies will be small. A final reversal of tendencies, and a retrograding of the cyclone, will result if the low-level mass-divergence effects overcompensate those from the upper wave pattern.

The distribution of horizontal divergence also determines the vertical motion, which in the "smoothed" cyclone model always goes upward in the front half and downward in the rear (Fig. 1). This model feature agrees with the observed distribution of cloudiness and precipitation in the cyclones. Modern aerological analysis of the field of vertical motion carried on by the Department of Meteorology at New York University has corroborated these findings. A summary of the careful and extensive work in that field was published in 1948 by Miller [20]. From that report we also know that the vertical motion of extensive air masses usually is less than 3 cm sec^{-1} even at the level of nondivergence, where the maximum upward and downward values of momentum occur. Higher values of vertical motion up to 10 cm sec^{-1} should occur over narrow zones near fronts, while the occurrences of updrafts and downdrafts of several meters per second are restricted to small parts of individual convective clouds.

Typical patterns of temperature distribution in the extratropical cyclone can be seen from the sample cyclones described later in this article. The most frequent development of the temperature pattern in the lower tropospheric part of the vortex can be represented schematically by the maps and profiles in Fig. 2. The incipient cyclone (Fig. 2*a*) coincides with the apex of a warm tongue formed on a "front" across

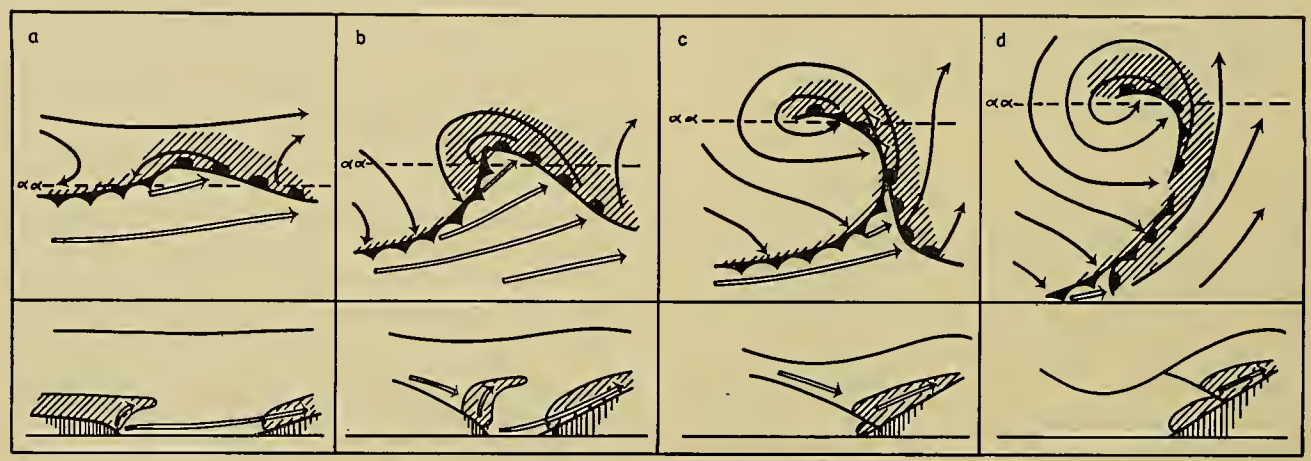


FIG. 2.—Successive stages of development of a frontal wave to an occluded vortex.

which the horizontal temperature gradient reaches a maximum. The frontal surface rises towards the cold side at an angle of inclination averaging around one in a hundred. It conserves its identity from day to day and moves along at a speed determinable from the winds through the kinematic boundary condition. The wave amplitude increases as the cyclone matures (Fig. 2*b*) and the central pressure decreases. Next follows the "occlusion" process (Fig. 2, *c* and *d*) during which the warm tongue is lifted from the ground, first near the center, later also farther out. The occluded front formed at the junction of the two cold wedges tends to wrap around the cyclone center as part of a spiral, and the same shape is found for the warm tongue in all levels of closed cyclonic circulation. In the profile in Fig. 2*c*, which is placed at a short distance south of the cyclone center, the occluded front is a warm-front type, that is, the cold wedge behind the occlusion is less cold than the one in front. This is also true of Fig. 2*d*, but it can usually be assumed that farther south the occlusion is a cold-front type. Where the transition from one occlusion model to another takes place the occluded front on the map must show a little

gap. The lifting of a tongue of warm air relative to a colder environment, illustrated in Fig. 2, can be assumed to furnish a great part of the increase in kinetic energy during the cyclonic development from wave to vortex.

The tropopause is also shown in the profiles. It has a crest over the warm-front surface and a trough over the cold-front surface, and the amplitude of the tropopause oscillation increases with the growth of the cyclone. In Fig. 2*d* the tropopause has a deep depression almost coinciding with the cyclone center, which is at that stage surrounded by air of cold origin up through the whole troposphere. Details of tropopause structure, such as the frequent subdivision into multiple tropopauses, have been left out in Fig. 2.

Hatched areas in Fig. 2 indicate the location of the main precipitation areas of the cyclone. The largest area is covered by the warm-front rain, where the air from the warm tongue climbs the receding wedge of cold air and condenses much of its moisture. A more narrow zone of precipitation accompanies the cold front where some air from the lower part of the warm tongue is lifted by the advancing cold wedge. Higher portions of the warm tongue move faster than the cold-front wedge and are not lifted by it. The described upward motion of the warm air next to the frontal surfaces should be visualized as being superimposed on the general pattern of vertical motion, upward in the front half and downward in the rear half of the cyclone (Fig. 1). This general, upward motion is sometimes sufficient to cause rain where it is not called for as a consequence of upgliding on frontal surfaces. Some extensive warm-sector rains and also the rain in the front half of a cold trough or a cold vortex are probably to be explained by the general upward motion shown in Fig. 1.

To complete the precipitation picture of the cyclone the air-mass precipitation should also be added, that is, the drizzle in the warm moist parts, caused by condensation from low clouds formed by the cooling of the warm air over cold surfaces (mainly ocean surfaces), and the convective showery precipitation formed through the heating from the ground, or through lifting of convectively unstable air at fronts.

While the thermal pattern of the cyclone near the ground is the result mainly of horizontal advection and nonadiabatic gain or loss of heat exchanged with the ground, the pattern in the free atmosphere is also influenced by the slow but systematic vertical displacement of the air shown in Fig. 1 and by the heat transfer of penetrative convection. However, the dominant process for the shaping of the upper-tropospheric temperature field is horizontal advection. The development of the thermal pattern of the waves in the upper westerlies follows roughly the advective scheme shown in Fig. 3. A warm tongue forms in the part of the wave with advection from the south, and a cold tongue in the part with advection from the north. In this early stage of the wave the pressure crests and troughs must tilt westward, as shown for the pressure trough in Fig. 1. In the further development, both warm and cold tongues grow in amplitude and move forward relative to the pressure wave, because the eastward motion of the air exceeds that of the wave. If the wave motion were entirely horizontal, a thermal pattern of permanent structure relative to the moving wave would be reached when the isotherms have adapted to the shape of the relative streamlines. For the idealized case of v_x = constant in each level of the sinusoidal

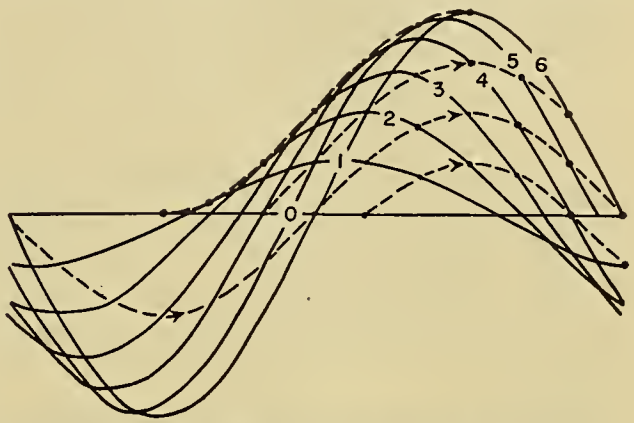


FIG. 3.—Advective formation of the thermal upper wave by the winds blowing relative to the moving pressure wave.

wave pattern, the ratio of the amplitude A_R of the relative streamline to that of the streamline A_S would be

$$\frac{A_R}{A_S} = \frac{v_x}{v_x - c}. \tag{4}$$

Hence, close to the level where v_x is equal to the wave speed c, the relative streamlines, and with them the advectively transported isotherms, would acquire a much greater amplitude than that of the streamline. The ratio A_R/A_S will decrease from that level upward to the tropopause, where v_x has its maximum.

Under the influence of the upward motion ahead of the pressure trough and downward motion behind it, the ratio in (4) would be reduced, as shown by Miller [20]. Synoptic experience shows that A_R/A_S stays positive in all tropospheric levels where $v_x > c$, also under the joint influence of vertical motion and horizontal advection. In other words, after a period of thermal transformation of the type shown in Fig. 3, the waves in the upper troposphere tend to become thermally symmetric, with warm tongues coinciding with pressure crests and cold tongues with pressure troughs.

With the reversal of the meridional temperature gradient from troposphere to stratosphere, the advective effects on temperature in the upper waves are also reversed. Hence the stratospheric pressure crests are cold and the pressure troughs warm. It then follows indirectly that the wave pattern of pressure crests and troughs rapidly loses amplitude with height in the stratosphere. In the stably stratified stratosphere the local warming and cooling through vertical motion are stronger than in the troposphere, and are quite

often stronger than the temperature change by advection.

Vorticity Analysis of the Extratropical Cyclone. An analysis of the vorticity distribution and the history of vorticity change of individual parcels in the vortex and wave will reveal more of the dynamics of the cyclone. The vertical component of the vorticity ζ may be identified on the horizontal streamline maps as a particle rotation about a vertical axis, partly due to curvature, v/r_s, where r_s is the radius of curvature of the streamline, and partly due to shear, $-\partial v/\partial n$:

$$\zeta = \pm \frac{v}{r_s} - \frac{\partial v}{\partial n} = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}. \tag{5}$$

The rules for determining the algebraic sign can always be decided upon by referring to the Cartesian component form of vorticity, added as an alternate expression in (5). The convention used here is to let the positive direction of the coordinate n point to the left of the wind, to consider v and r_s always positive, and to use the positive sign in front of v/r_s for cyclonic and the negative sign for anticyclonic curvature of the streamlines.

The vorticity change of the individual traveling particle [11, 13, 16, 26, 27] is given by the equation,

$$\frac{d\zeta}{dt} = \frac{\partial p}{\partial x}\frac{\partial \alpha}{\partial y} - \frac{\partial p}{\partial y}\frac{\partial \alpha}{\partial x} - (\zeta + 2\Omega \sin\phi)\,\mathrm{div}_H\, \mathbf{v} - \frac{(2\Omega \cos\phi)}{a} v_y + \frac{\partial v_z}{\partial y}\left(2\Omega \cos\phi + \frac{\partial v_x}{\partial z}\right) - \frac{\partial v_z}{\partial x}\frac{\partial v_y}{\partial z}. \tag{6}$$

The first two terms on the right represent in component form the effect of isobaric-isosteric solenoids on the change of vertical vorticity. These terms are always found to be insignificant compared to the following ones. The divergence term shows how horizontal expansion (divergence) creates negative (anticyclonic) vorticity, and horizontal contraction creates positive (cyclonic) vorticity. The next term shows the effect on relative vorticity ζ of the displacement of the air, either towards the polar regions where the vertical component of the earth's vorticity $2\Omega \sin\phi$ is great, or towards the equator where $2\Omega \sin\phi$ is zero. In the absence of the other factors, poleward movement would entail a decrease of relative cyclonic vorticity or an increase of relative anticyclonic vorticity, and movement away from the pole would entail an increase of relative cyclonic vorticity or a decrease of relative anticyclonic vorticity. The last two terms on the right-hand side describe the influence of the vertical motion in changing the vorticity about the vertical. The term involving $\partial v_z/\partial y$ represents, in part, the fact that the infinitesimal disk of air, whose rotation decides the value of ζ, arrives at a horizontal position from earlier positions with meridional tilt. In terms of relative-vorticity change, this is equivalent to a change in latitude in addition to that by horizontal meridional advection, as can be seen from the analogy between the terms $(2\Omega \cos\phi)\partial v_z/\partial y$ and $-(2\Omega \cos\phi)v_y/a$. Furthermore, the term in $\partial v_z/\partial y$ represents the effect of rotating the air in the yz-plane so that the vorticity about the y-axis, $\partial v_x/\partial z$, acquires a vertical component at the rate $(\partial v_z/\partial y)(\partial v_x/\partial z)$ per unit time. Analogously, the last term in (6) represents the rate of change of vorticity about the z-axis resulting from a rotation of the air in the xz-plane. Usually the terms in $\partial v_z/\partial y$ and $\partial v_z/\partial x$ are considered to be insignificant in relation to the divergence term and the meridional advection term, but possible exceptions will be mentioned below.

In the frontal wave of the lower troposphere, the cold air enters the moving cyclone along the warm front (Fig. 4). Near the front it has initial cyclonic shear

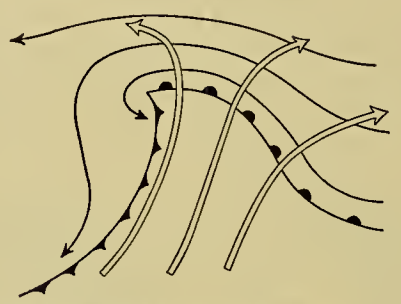


FIG. 4.—Motion of warm and cold air relative to the moving frontal wave.

which had been acquired during the period of frontogenesis (see p. 590). The increase of the cyclonic vorticity in the cold air on its way toward the wave apex is due to the horizontal convergence, which extends all over the front half of the cyclone (see Fig. 1). This creation of relative cyclonic vorticity is somewhat reduced by the effect of the poleward component of air travel. Behind the wave apex the air returns southward and as a result its relative cyclonic vorticity should increase. At the same time, however, the air enters the region of horizontal divergence, which has the opposite effect upon vorticity change. The result is that the air which had acquired maximum cyclonic shear along the warm front maintains cyclonic vorticity after passing the wave apex, but with a simultaneous shift from shear to curvature vorticity. The cold air passing at greater distance from the center changes from moderate cyclonic to anticyclonic vorticity. During the growth of the cyclone more and more of the cold air is able to maintain its cyclonic vorticity after passing the wave apex.

The warm air in the frontal wave enters the cyclone from the southwest. Its speed in lower levels just barely exceeds that of the cyclone in its eastward motion. Upon arrival at the warm front the warm air climbs the receding cold wedge. Again, one branch of anticyclonic and another of cyclonic vorticity may be discerned. Farthest away from the center, where the horizontal convergence is moderate or nonexistent, the warm air gains anticyclonic relative vorticity through the poleward component of movement. Closer to the center, where the horizontal convergence is stronger, the warm air acquires cyclonic relative vorticity despite its dis-

placement polewards. This latter development, which does not start until the cyclone is past the nascent stage, gains in magnitude with the growth of the cyclone.

In the westerly wave of the upper troposphere, as represented in Fig. 1, the air enters the cyclone from the northwest and leaves it toward the east, across the wave crest ahead of the surface cyclone. As long as the upper wave does not degenerate, the relative vorticity changes sign at the longitude of the inflection points, thus changing from anticyclonic to cyclonic relative vorticity in the middle of the upper zone of convergence, and from cyclonic to anticyclonic in the middle of the upper zone of divergence. This vorticity change by divergence is supported by the effect of meridional advection.

The factor $(\zeta + 2\Omega \sin \phi)$ in the divergence term of (6) is equal to the absolute vorticity ζ_a of the air relative to a nonrotating coordinate system. When ζ is positive (cyclonic), ζ_a is large and the individual vorticity change with time becomes quite sensitive to horizontal convergence or divergence. If the air is subject to a sustained process of horizontal convergence, its cyclonic vorticity will increase without any theoretical upper limit. The cyclonic bends of an upper sinusoidal westerly are therefore frequently seen to become strongly curved. On the other hand, when ζ acquires large negative (anticyclonic) values, the absolute vorticity may go to zero or even become negative. This happens almost exclusively in the upper troposphere and lower stratosphere where the wind velocities are very strong. On wave crests where ζ_a reaches values close to zero, the vorticity change is only feebly influenced by horizontal divergence, and obeys mainly the term of meridional advection in (6). This would be equivalent to motion under approximately constant absolute vorticity $\zeta_a \approx 0$ or $\zeta \approx -2\Omega \sin \phi$. If a wave crest in the upper atmosphere has developed to that extreme stage, the particles overtaking the crest would maintain their anticyclonic vorticity (in the form of curvature and/or shear) for a long period thereafter. Figure 5 illustrates that case schematically. From an initial flow pattern of sinusoidal westerlies (streamline 1) a "meandering" westerly current develops through the growth of the wave crest and the deepening of the next downwind wave trough (streamlines 2 and 3). This development towards meandering flow is not dependent on the absolute vorticity's actually having reached zero. With absolute vorticities still positive, but numerically small, the vorticity change begins to react sluggishly to horizontal convergence with the result that the sinusoidal perturbation of the westerlies begins to degenerate. The meandering development may also start from an initially straight current with anticyclonic shear close to the value $-2\Omega \sin \phi$. Any small wave impulse may then develop into meandering wave patterns.

It is obvious that the meandering phenomenon, once started in regions of excessive anticyclonic vorticity in the upper atmosphere, will also have a profound influence on the total cyclone picture down to the ground. The deepening of the upper wave trough is associated with the deepening of the cyclonic vortex underneath, and usually also entails a southward component added to the normal eastward displacement of the cyclone. In all cases of such deepening the initial upper disturbance must start through the build-up of excessive anticyclonic curvature on the wave crest to the west of the cyclone. Above the level of nondivergence, the divergence term and the meridional advection term in (6) are of the same phase, so that the additional terms in $\partial v_z/\partial y$ and $\partial v_z/\partial x$ are not likely to affect the general pattern of $d\zeta/dt$ very much. The two levels where their influence may be expected to count are (1) close to the level of nondivergence, and (2) at the localities where

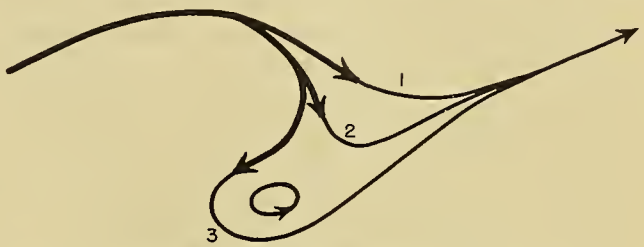


FIG. 5.—Successive (1→3) degeneration of sinusoidal wave-pattern caused by excessive anticyclonic vorticity on wave crest.

$\zeta - 2\Omega \sin \phi$ is near zero. In both cases the competition with the divergence term is almost eliminated. Furthermore, v_z and its horizontal derivatives reach their maximum in the upper troposphere (about two kilometers above the level of nondivergence where $|\rho v_z|$ has its maximum).

The most compelling reason for admitting a perceptible influence of the terms in $\partial v_z/\partial y$ and $\partial v_z/\partial x$ on the variations of ζ lies in the fact that neither the divergence term, nor the meridional advection term, nor their sum, can account through equation (6) for the occurrence of negative absolute vorticity. Analyses of observational data do show areas of negative absolute vorticity on pronounced upper wave crests and/or south of pronounced "jet streams" (see Fig. 7). A vertical motion effect of the right sign to explain the growth of $-\zeta$ beyond $2\Omega \sin \phi$ would be found north of the maximum of upward velocity in the cyclone ($\partial v_z/\partial y < 0$, $\partial v_z/\partial z > 0$). The result in terms of a large anticyclonic vorticity, and occasionally a negative absolute vorticity, can then be expected to accrue on the upper wave crest to the east of the cyclone.

The above reasoning about the vertical motion terms in equation (6) has been developed by L. Sherman and will appear under his authorship.

Inertial Motion in Isentropic Surfaces. In slow-moving long waves of the upper westerlies, the streamlines relative to the waves almost coincide with the streamlines relative to the earth, and the isotherms will be moved advectively so as to coincide more or less with the isobars. Around the inflection points of such long waves we find the best approximation to the relatively simple conditions of straight baroclinic flow. Frontogenesis and frontal cyclogenesis are frequent

under the straight southwesterly currents of long waves, and for a study of these phenomena we will here consider the theory of adiabatic, inertial motion in tilting, stationary isentropic surfaces. Adiabatic (or pseudo-adiabatic) changes of state of particles in the free atmosphere can justifiably be assumed, because non-adiabatic temperature changes are so slow and uniformly distributed that they do not appreciably affect a relatively rapid phenomenon such as cyclogenesis.

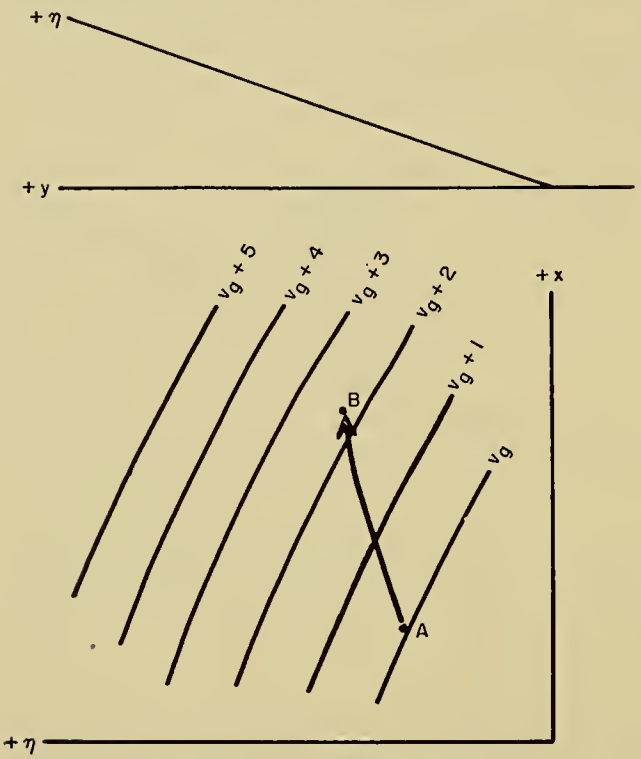


FIG. 6.—Profile of sloping isentropic $x\eta$-plane and sample distribution of the isovels of v_g in $x\eta$-coordinates.

The following simplified analysis has been inspired by the theoretical studies on "dynamic instability," which go back to the classical paper of Helmholtz, "Über atmosphärische Bewegungen," published in 1888 [12]. Several recent contributions by theoretical meteorologists to the same field have been included in the list of literature references [6–9, 14, 15, 17–19, 21, 31, 32]. The synoptic applications of equation (11) in this article to the problems of frontogenesis and cyclogenesis have, to my knowledge, not been attempted before.

The adiabatic movement of a particle parallel to an isentropic surface (sloping or horizontal) is not opposed by buoyancy forces and is left to the stable or unstable control of the horizontal pressure gradient and the Coriolis force. The same is true for particles of entire isentropic sheets moving in unison, when they are far enough from the ground to be independent of boundary effects. For the purpose of this article we shall consider the environmental field of pressure and potential temperature constant while the sample particle (or sample isentropic sheet) moves isentropically through the field. We shall consider part of an isentropic surface of sufficiently small extent to be treated as a sloping plane (Fig. 6), on which the horizontal direction will be called the x-direction (due eastward) and the direction of steepest slope (due northward) will be called the η-direction. The η-axis is supposed to form an angle with the horizontal y-axis of the order of one in a hundred or less. The undisturbed air motion is supposed to be zonal, geostrophic, and horizontal, which allows the isentropic surface to stay fixed in space. Furthermore the fundamental motion is constant along each streamline, $\partial v_x/\partial x = 0$, and represents a steady state, $\partial v_x/\partial t = 0$. Later application to the long wave, where the quasi-straight flow is not exactly zonal, will be done without any strict mathematical treatment. The disturbed motion of the sample particle is supposed to be contained in the isentropic surface and to have an initial upslope component v_η superimposed on the general horizontal motion characteristic of the environment.

The x-component of acceleration of the disturbed particle amounts to

$$\frac{dv_x}{dt} = 2\Omega_z v_y \cong 2\Omega_z v_\eta, \tag{7}$$

and makes the particle speed up in the positive x-direction while it climbs the isentropic slope. The wind of the environment v_g, which is geostrophic and directed along the x-axis in the whole field, changes in value along the path of climb. An observer following the disturbed particle would see the environmental geostrophic wind, relative to the earth, change by

$$\frac{dv_g}{dt} = v_\eta \frac{\partial v_g}{\partial \eta}. \tag{8}$$

The x-component of the speed of the disturbed particle was assumed to be equal to the geostrophic wind v_g at the initial time. Depending on whether $dv_x/dt > dv_g/dt$ or $dv_x/dt < dv_g/dt$, the disturbed particle will move eastwards faster or slower than its new environment after a time differential of climbing. In the first case, the y-component of acceleration of the disturbed particle $dv_y/dt = -2\Omega(v_x - v_g)$ will be directed down the isentropic slope opposite to the initial disturbance. A stable inertia type of oscillation will then result. In the second case, the y-component of acceleration will point in the same direction as the initial disturbance velocity v_η, so that an exponential growth of the disturbance will follow. The instability case is thus $dv_x/dt < dv_g/dt$. In order to make the instability criterion applicable also for the downward directed disturbance, it should be written $\left|\frac{dv_x}{dt}\right| < \left|\frac{dv_g}{dt}\right|$.

Now, provided that the substitution of v_η for v_y in (7) is justified by a sufficiently small inclination of the isentropic surface, the instability criterion derived from (7) and (8) takes the simple form,

$$\frac{\partial v_g}{\partial \eta} > 2\Omega_z. \tag{9}$$

The observed increase of westerly geostrophic wind from the lower to the higher portion of an isentropic

surface sometimes satisfies this instability criterion, as will be shown later.

If we drop the initial conditions of $\partial v_x/\partial x = 0$ and $\partial v_x/\partial t = 0$, no exact treatment can be offered, because then the fundamental current is not a steady-state one. The following reasoning should however be valid, provided that the long-wave deformations of the fundamental current are much slower than the short-wave developments on the isentropic surface. This condition is usually fulfilled.

In Fig. 6 an element of the isentropic surface is shown in $x\eta$-coordinates. Isobars on that surface are parallel to the x-direction, while the distribution of the speed of the geostrophic wind v_g is shown by slanting scalar curves. Thus v_g has a gradient in the x-direction in addition to the much stronger gradient in the η-direction. The disturbed path of a sample particle along the isentropic surface is supposed to go from A to B during the time differential. The x-component of the acceleration (parallel to the isobars) of the sample particle is again, to a first approximation, $dv_x/dt = 2\Omega_z v_\eta$, while the change of geostrophic wind encountered along the path is

$$\frac{dv_g}{dt} = v_\eta \frac{\partial v_g}{\partial \eta} + v_x \frac{\partial v_g}{\partial x} + \frac{\partial v_g}{\partial t}. \tag{10}$$

The acceleration of the particle in the η-direction is supposed to be zero at the initial point of the trajectory A. The acceleration in the η-direction will also be zero at the end of the trajectory B if $dv_x/dt = dv_g/dt$, or

$$2\Omega_z v_\eta = v_\eta \frac{\partial v_g}{\partial \eta} + v_x \frac{\partial v_g}{\partial x} + \frac{\partial v_g}{\partial t},$$

that is,

$$v_\eta = \frac{v_x \dfrac{\partial v_g}{\partial x} + \dfrac{\partial v_g}{\partial t}}{2\Omega_z - \dfrac{\partial v_g}{\partial \eta}}. \tag{11}$$

Specializing now for the condition $2\Omega_z - \partial v_g/\partial \eta > 0$ and for $v_x \partial v_g/\partial x + \partial v_g/\partial t > 0$, we find that the particle given an initial speed component v_η greater than the value found in (11) will have an acceleration component dv_η/dt opposite to v_η. Given a smaller positive initial v_η, or a negative initial v_η, the particle would accelerate towards the value for v_η given in (11). This value of v_η therefore represents the η-component of a stable upgliding motion in which all the particles of the isentropic surface may join. The η-component of the stable upgliding motion approaches infinity when $2\Omega_z - \partial v_g/\partial \eta$ goes to zero. In other words, in the case of inertial indifference any finite initial v_η would increase exponentially.

Quite analogous reasoning in the case $2\Omega_z - \partial v_g/\partial \eta > 0$ and $v_x \partial v_g/\partial x + \partial v_g/\partial t < 0$ reveals the existence of stable downgliding motion in which the η-component is also given by (11).

The difference between the cases $v_x \partial v_g/\partial x + \partial v_g/\partial t = 0$ and $\gtrless 0$ is thus the following: In the former case the departures from the geostrophic wind remain of a stable oscillatory nature until the anticyclonic isentropic shear reaches the critical value of $-2\Omega_z$. In the latter cases there is a sustained stable departure from the geostrophic wind, represented by v_η, which has finite values also when $2\Omega_z - \partial v_g/\partial \eta > 0$. In addition there may be inertial perturbations superimposed on the current representing the vector sum of geostrophic motion v_g and isentropic motion v_η; such perturbations will be stable as long as $2\Omega_z - \partial v_g/\partial \eta > 0$.

The demonstration of the occurrence of anticyclonic shear in the upper atmosphere, which approaches or even surpasses the critical limit of dynamic instability, is due to recent research work at the University of Chicago. The first profile of the westerlies showing such conditions was analyzed by Palmén [23] in 1948. In the same paper it is also shown how we must treat zonal flow as curved flow (radius of curvature $r = a \operatorname{cotan} \phi$) in order to arrive at a satisfactory accuracy of an isovel profile which is to correspond to an observed meridional pressure profile. In the following discussion we shall use the model of straight baroclinic westerlies with a stationary polar front published by Palmén and Newton [24] and reprinted here as the left part of Fig. 7. It represents an isotherm-isovel profile based on an averaging of twelve eastern North American meridional profiles made during December 1946. In the right-hand part of Fig. 7 we have added a diagram of the computed quantity $2\Omega_z - \partial v_g/\partial \eta$, with η interpreted as the curvilinear isentropic coordinate (positive direction northwards). In most of the field $2\Omega_z$ is greater than $\partial v_g/\partial \eta$, indicating inertial stability; but in a narrow zone south of the maximum upper westerlies, $2\Omega_z - \partial v_g/\partial \eta$ is negative, indicating inertial instability. Furthermore, in the frontal zone below 600 mb, where $\partial v_g/\partial \eta$ has been measured along saturation isentropes, the values of $2\Omega_z - \partial v_g/\partial \eta$ indicate only a slight amount of inertial stability. We shall focus our attention first on that part of the profile.

The small positive (or in some individual cases negative) values of $2\Omega_z - \partial v_g/\partial \eta$ are located in a narrow frontal zone, while in the adjacent parts of the warm and cold air masses, $2\Omega_z - \partial v_g/\partial \eta$ is positive and far from zero. Since in (11) the component of stable upgliding or downgliding is inversely proportional to $2\Omega_z - \partial v_g/\partial \eta$, it follows that the air in the narrow frontal zone has a much greater possibility for isentropic up- or down-displacements than the air masses on either side.

The quantity $v_x \partial v_g/\partial x + \partial v_g/\partial t$, representing the numerator in the expression for v_η in (11), cannot be judged from the data of one profile alone. It will be large and positive (1) where the isobars of the horizontal pressure distribution converge, and (2) where the gradient wind increases locally with time. The first condition is fulfilled, for example, along the axis of kinematic dilatation extending eastward from a col of the pressure field. This synoptic situation is known to be frequently associated with frontogenesis and subsequent maintenance of a sharp front. The second condition, local increase of gradient wind parallel to the frontal zone, frequently occurs during frontogenesis,

but the cause of such an increase of gradient wind is not necessarily attributable to the frontal mechanism.

Whenever $v_x \partial v_g/\partial x + \partial v_g/\partial t$ has the same sign in each of the two air masses, v_η will also have the same sign in the whole field; but its maximum numerical value, as far as the lower troposphere is concerned, will be found in the frontal zone where $2\Omega_z - \partial v_g/\partial \eta$ is at a minimum.

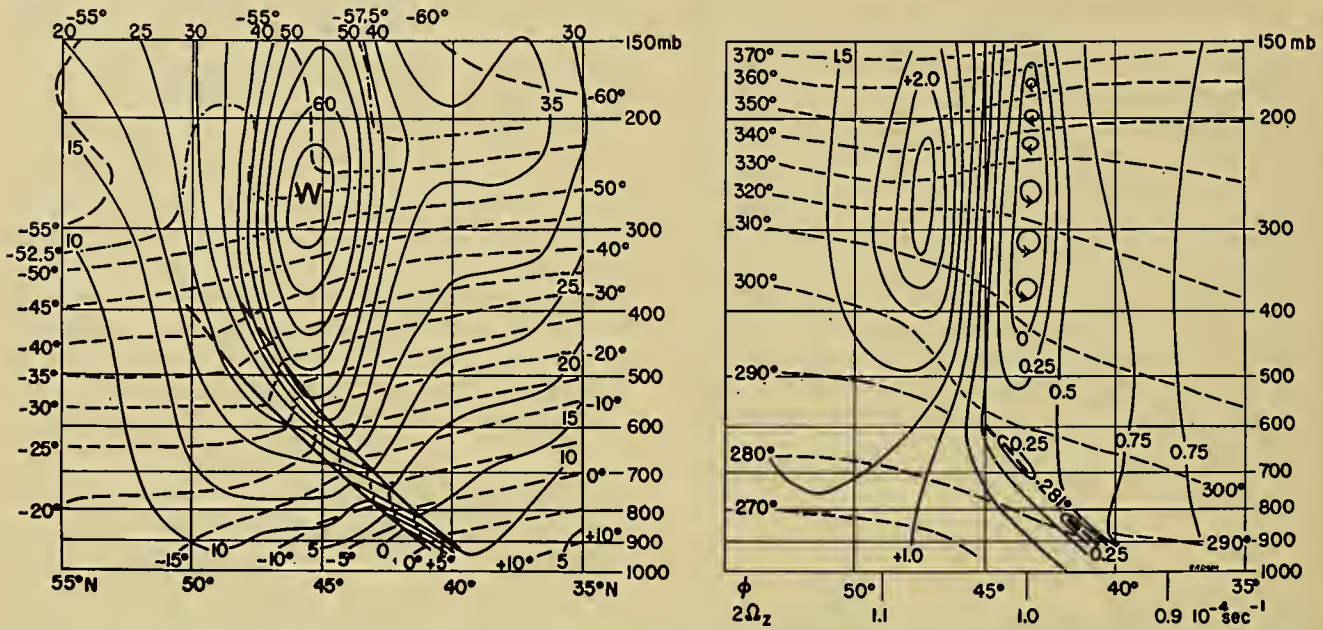


Fig. 7.—Meridional profiles through a model of straight westerlies with quasi-stationary polar front (Palmén and Newton [24]). Left: Dashed lines show isotherms (degrees centigrade), and solid lines isovels (m sec^{-1}) of zonal geostrophic wind. Right: Dashed lines show the field of dry-isentropes (degrees absolute), and the saturation-isentrope of 281° in the frontal zone. Solid lines represent the quantity $2\Omega_z - \partial v_g/\partial \eta$ in units of 10^{-4} sec^{-1}.

As shown in Fig. 7, the warm air over the lower and intermediate portion of the polar-front surface has anticyclonic isentropic shear, increasing to great values in the upper troposphere, whereas the air above the upper part of the frontal surface has cyclonic shear, likewise increasing to high values in the upper troposphere. The dividing line between anticyclonic and cyclonic shear runs almost vertically through the maximum of west-wind velocity, which in the average condition represented by Fig. 7 is located above the place where the frontal surface intersects the 500-mb level. Isentropic upgliding or downgliding as defined by equation (11) will reach larger values south of the velocity maximum than north of it. It is likely that this difference in v_η values north and south of the velocity maximum does give rise to important horizontal divergence effects because the η-component represents a nongeostrophic part of the total wind. The v_η-divergence effect in the jet-stream region should work out as shown schematically in Fig. 8. Where there is "confluence" of the winds into the western beginning of a "jet stream," equation (11) indicates a superimposed isentropic upgliding $v_\eta > 0$ and "isentropic convergence" $\partial v_\eta/\partial \eta < 0$ north of the line of maximum $|v_\eta|$. Where the wind velocity decreases along the streamlines in the "delta" of a jet stream, equation (11) indicates isentropic downgliding $v_\eta < 0$ and isentropic divergence $\partial v_\eta/\partial \eta > 0$, north of the line of maximum $|v_\eta|$. In the stratosphere the terms downgliding and upgliding must be interchanged, because of the opposite tilt of isentropic surfaces, but the statement about the isentropic divergence remains identical for stratosphere and troposphere. At the line of maximum $|v_\eta|$ values the isentropic divergence $\partial v_\eta/\partial \eta$ changes sign, as shown by the hatching in Fig. 8.

In figuring out the effect of the isentropic divergence in changing the pressure field we may think of the distribution of $\partial v_\eta/\partial \eta$ as representing in the first approximation a field of $\partial v_y/\partial y$, where v_y is the nongeostrophic y-component of motion. Assuming that the distribution of $\partial \rho v_y/\partial y$ can also be qualitatively represented by the hatched and unhatched areas in Fig. 8, we have in that diagram an outline of the contribution of isentropic divergence to the total horizontal divergence. The isentropic divergence, acting in the same sense through

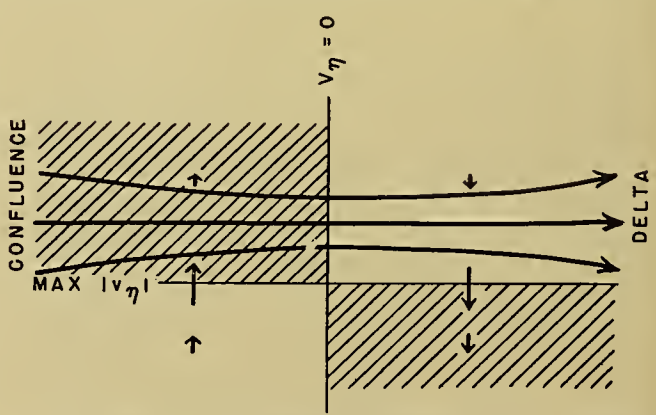


Fig. 8.—Isentropic convergence (hatched) and divergence (unhatched) in the regions of jet-stream coufluence and delta.

the stratosphere and the upper half of the troposphere, may be an important effect to consider together with the divergence effects represented in Fig. 1. A cyclonic storm traveling along the jet-stream zone would come under the influence of superimposed upper mass divergence from the time when it passes the place of greatest constriction of the upper streamlines. A complete theoretical treatment of this case, which calls for a combination of the divergence effects of Fig. 1 and Fig. 8, is not available; but there seems to be considerable empirical evidence for strong cyclonic deepening under the described circumstances. Such synoptic evidence has mainly been gathered by Scherhag [30]. Scherhag points to Ryd [28, 29] as the originator of the idea that mass divergence of importance for cyclone deepening should occur in upper delta patterns. Ryd's theoretical contributions appeared in 1923 and 1927 when there were as yet no upper-air maps.

Returning to Fig. 7, we see that complete inertial instability $\partial v_g/\partial \eta > 2\Omega_z$ may at times extend from 150 mb down towards the 500-mb surface. It may also extend over a thousand kilometers' length of current, but the width of the zone of such unstable shear is hardly more than three hundred kilometers at any one point. Inside that volume of current the geostrophic wind, with its superimposed component of isentropic upgliding or downgliding, does not represent a stable flow. However, with stable neighboring flow on either side, no very large unstable deviations from geostrophic flow will be able to develop. The most likely system of perturbations in the unstable part of the current will be helical cellular circulations, as indicated in Fig. 7. Such circulations would serve the purpose of exchanging momentum across the zone of unstable shear and thus lessen that shear. The height of each cell would have to be small, probably less than one kilometer, so that the solenoid field set up by the cellular circulation should not grow strong enough to reverse the initial circulation. An indirect indication of the existence of the helical cellular circulations is seen in the observed "multiple tropopauses," each one probably representing a cell wall between superjacent circulation rolls. According to Palmén [22], these multiple tropopauses are quasi-isentropic as would be expected if they are formed as circulation-cell boundaries.

SYNOPTIC EXAMPLE OF AN EXTRATROPICAL CYCLONE

The weather situation over North America during November 7–10, 1948, has been selected to illustrate the principles of this article. A large occluded cyclone which was located over the Hudson Bay region during this period can serve as a model of the most frequent structure of old cyclones, while over the central United States the atmosphere displays all the successive stages of frontogenesis and the early life history of a growing frontal wave cyclone. Our description will begin with the evolution of the long-wave background pattern of the upper layers, represented by a set of 300-mb maps, then the advective frontogenesis in the lower troposphere will be illustrated by a sequence of ground-level and 850-mb maps as well as selected profiles, and finally the three-dimensional structure of the frontal wave cyclone will be shown by a synoptic set of maps from the ground to 300 mb.

Synoptic Evolution of the Upper Layers. The six 300-mb maps at 12-hr intervals in Fig. 9 all show the semipermanent Hudson Bay cyclone. Through the whole troposphere this cyclone is a cold-core vortex and therefore shows up as a deep center on the 300-mb maps. Equally permanent is the crest of high pressure extending northwards from a warm anticyclone over the eastern North Pacific. Both the Hudson Bay low and the eastern Pacific high are typical features of the general circulation but they have more than average strength during November 7–10. The westerly current meandering through between them is quite strong over a narrow zone, while the pressure gradients in the high and the low are quite weak. The trough located over the western United States on November 7 moves slowly to the central states and deepens gradually from November 7 to November 9. This upper-air process plays an important role in the formation of the frontal cyclone which takes place under the pre-trough southwesterly current (without producing any separate low-pressure center at 300 mb).

The deepening of the upper trough may be caused in two ways (see equation (1)): either through a sinking component of motion at the 300-mb level, or through horizontal mass divergence in the column above 300 mb. In the former case the temperature in the trough at 300 mb ought to be rising with time. This is not borne out by the observations during November 7–9, so that we are left with the horizontal mass divergence as the probable cause of the deepening of the trough. The mass divergence must be operated through the feeding of air into the trough with such a high velocity that the Coriolis force and centrifugal force overcompensate the initial pressure gradient. The mechanism for producing such a strong jet in the northerly current behind the trough must be sought on the anticyclonic bend to the west.

The maximum curvature of the 28,400-ft contour of the 300-mb surface is represented on the November 7th map by an arc of a circle with radius r_i. At the same place in western Canada the maximum possible curvature of a steady-state anticyclonic current, flowing under the influence of the observed pressure-gradient force, is represented by another arc of a circle with radius $r_{\min}$.[3] The curvature analysis on the 300-mb

3. The value of $r_{\min}$ is obtained from the equation of anticyclonic circular motion

$$-v^2/r = -2\Omega_z v - \partial\Phi/\partial r = -2\Omega_z(v - v_g), \qquad (12)$$

in which Φ stands for the geopotential in the pressure topography, r and v are positive, $-\partial\Phi/\partial r$ is the outward-directed pressure gradient, $-2\Omega_z v$ is the inward-directed Coriolis force, and $-v^2/r$ is the centripetal acceleration. When equation (12) is applied to a selected point on the map, Ω_z and $\partial\Phi/\partial r$ are

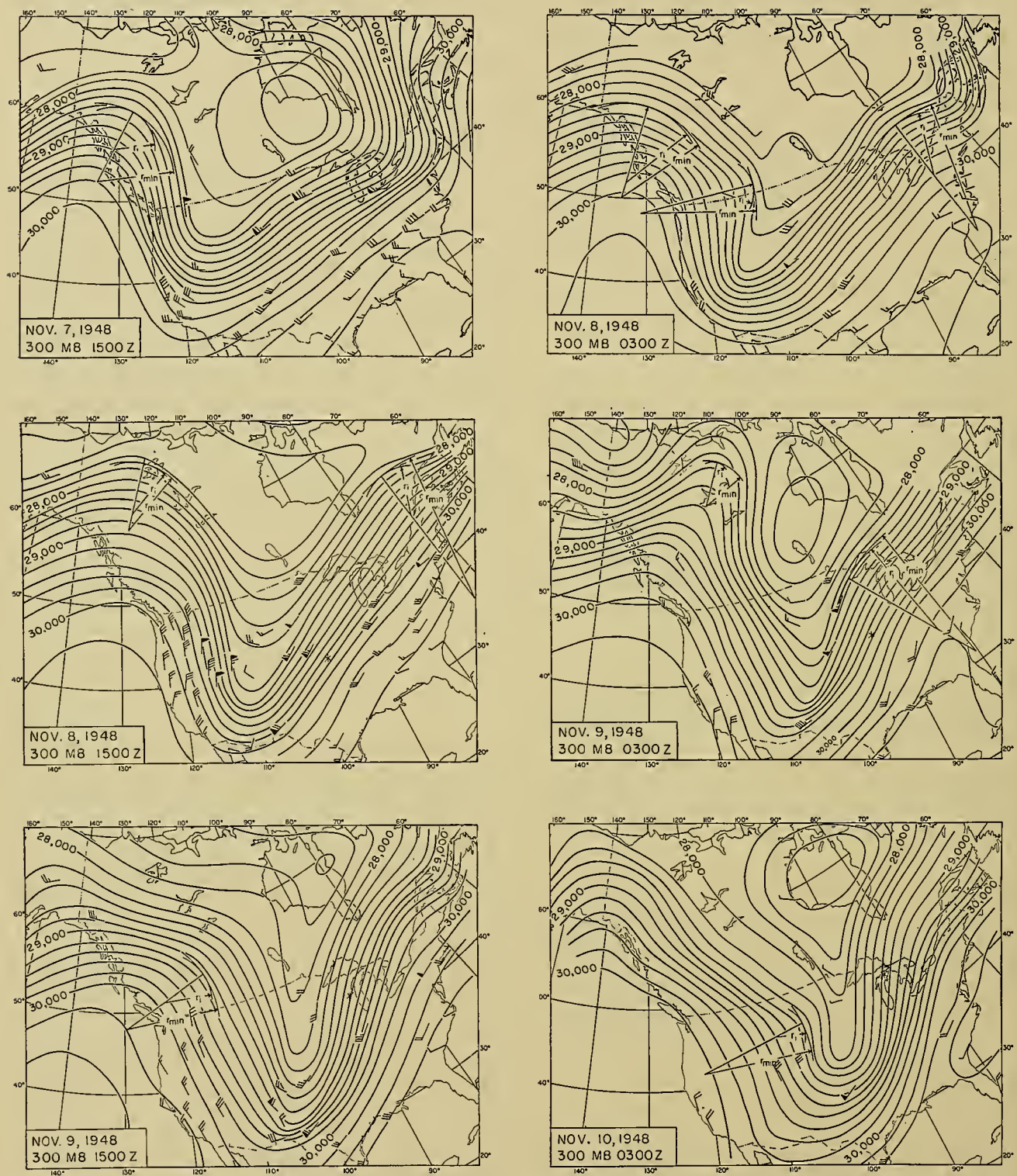


FIG. 9.—Twelve-hourly sequence of 300-mb maps during November 7–10, 1948 (r_i = radius of isobaric curvature, $r_{\min} = -\frac{\partial \Phi}{\partial r}/\Omega_z^2 = 2v_g/\Omega_z$). Asterisk marks position of apex of frontal wave on sea-level map. Half barb 5 m sec^{-1}, full barb 10 m sec^{-1}, triangular barb 50 m sec^{-1}.

anticyclonic bend over western Canada on November 7 and the following days shows that instances of $r_i < r_{min}$ are quite frequent, in other words, that with the given pressure gradient and contour curvature the paths of particles often cannot have as small a radius of curvature as that of the isobaric contours. If that applies to a quasi-stationary pressure ridge like the orographic one over the Canadian Rockies, where the radius r_s of streamline curvature is equal to the radius r of path curvature, the air would have to cross the isobars towards low pressure while making the anticyclonic turn. This must imply a forward acceleration of the particle leading up to maximum speed at the end of the anticyclonic sweep. If such fast-moving air is fed directly into a pressure trough downstream, in which the pressure gradients were adapted to smaller wind speeds, an intensification of the trough should follow. The deepening of the large pressure trough over western and central United States during November 7–9, 1948, should probably be interpreted this way. Measured winds are unfortunately not available for a complete check of these ideas. The only part of the phenomenon that can be well demonstrated is the general occurrence of wind components towards high pressure, resulting from the supergeostrophic velocity of the air that is leaving the anticyclonic bend (see winds over the western United States on the 300-mb map of November 8, 1500Z).

The flow around the quasi-stationary anticyclonic bend over western Canada cannot be a steady-state one, although the major features of that part of the map do remain unchanged. The 300-mb maps show how one moving wave perturbation after the other appears on top of the large stationary crest of high pressure in the west. The first of these traveled about 1500 km during twelve hours (35 m sec^{-1}) and is found on the second map with its wave crest in the northerly current at the Canadian-United States border. The 300-mb contour curvature at that time and place defines an r_i which is much smaller than r_{min} .

The radii of curvature of streamlines, r_s , and of air trajectories, r, in a moving sinusoidal wave, are related to each other by the formula

$$r_s = r\,\frac{v - c}{v}, \tag{16}$$

---

constants. Solving (12) for

$$r = \frac{v^2}{2\Omega_z v + \partial\Phi/\partial r} = \frac{v^2}{2\Omega_z(v - v_g)}, \tag{13}$$

and seeking the value of v for which r is a minimum, we obtain

$$v = \frac{-\dfrac{\partial\Phi}{\partial r}}{\Omega_z} = 2v_g. \tag{14}$$

The corresponding values of r_{min} and v are thus

$$r_{min} = -\frac{\dfrac{\partial\Phi}{\partial r}}{\Omega_z^2} = \frac{2v_g}{\Omega_z}, \qquad v = 2v_g. \tag{15}$$

where c is the speed of the wave. No measured wind velocities are available at 300 mb in western Canada during the days under consideration. Theoretical estimates of v must lie between v_g and $2v_g$, and are most likely closer to the lower than the upper limit, as will be shown later. Assuming tentatively for the moving pressure crest at the Canadian–United States border on November 8, 0300Z, $v = 1.1v_g = 49$ m sec^{-1}, we would have from (13)

$$r = \frac{49^2}{1.08 \times 10^{-4}(49 - 44.5)} = 4900 \text{ km},$$

and would arrive at the following estimate of r_s :

$$r_s = 49 \times 10^5\,\frac{49 - 35}{49}\,\text{m} = 1400 \text{ km},$$

which is much longer than the measured radius of contours $r_i = 440$ km. These estimates and measurements are of course subject to great errors, but even so, the conclusion seems to be that also on the moving pressure crests the streamlines will fail to adapt to the strong curvature of the isobars. It then also follows that the moving pressure crests at the 300-mb level are preceded by a velocity maximum. When the air from that velocity maximum enters the slow-moving low-pressure trough, a pulse of deepening by centrifugal action would result. The rapidly moving upper wave cannot be seen to continue its propagation on the front side of the deep slow-moving trough. Hence all its wave energy must have been absorbed in the large trough.

With the above estimate of $v = 49$ m sec^{-1} and $r_s = 1400$ km, the anticyclonic vorticity due to curvature $-v/r_s$ amounts to -3.5×10^{-5} sec^{-1}. This is numerically much less than $2\Omega \sin 50° = 1.1 \times 10^{-4}$ sec^{-1}, so it does not seem likely that the complete anticyclonic vorticity $-v/r_s - \partial v/\partial n$ reaches the critical value of $-2\Omega_z$ anywhere in the rapid wave at 300 mb. The described manifestation of instability through cross-isobaric flow on the anticyclonic bend thus takes place independently of the fulfillment of the criterion $-v/r_s - \partial v/\partial n + 2\Omega_z < 0$.

On November 8, 1500Z, when the most unstable part of the anticyclonic flow was found far north (again marked by $r_i < r_{min}$), a growing crest and a downwind, deepening trough formed simultaneously. When that perturbation caught up with the slow-moving trough ahead, another deepening occurred (see November 9, 1500Z), this time in the north-central United States while the southern end of the trough was losing depth.

These fast-moving unstable waves on the 300-mb maps are, of course, at times connected with disturbances in the lower atmosphere. The first of the upper waves was formed on November 7, 1500Z, as an occluded front was approaching from the west; it is likely that the excessive anticyclonic curvature resulted from a superposition of the upper wave crest (associated with the occluded cyclone) upon the semipermanent anticyclonic bend produced orographically by the northern Rocky Mountains. Once formed, the unstable

upper wave separates from the frontal disturbance by virtue of its superior speed (35 m sec^{-1}). The second unstable upper wave had no clear connection with any frontal disturbance.

During the selected period, the flow of air east of the big slow-moving trough turned gradually from west-southwest towards south-southwest while increasing a little in strength. On November 8, 1500Z, after the cold trough of the Hudson Bay cyclone had moved off to the northeast, the upper current over the eastern half of the United States and Canada became almost straight. On November 9, 0300Z, when the growing frontal cyclone (marked by an asterisk on the 300-mb maps) began to exert influence high up, the upper current became slightly S-shaped. The newly formed upper wave moved along with the cyclone center below at a speed of only 9 m sec^{-1}. The best estimate of the wind speed on the anticyclonic bend is probably 80 m sec^{-1} (see below) and hence $r_s = r(80 - 9)/80 = 0.9r$. Even with $r = r_{min} = 1950$ km, r_s would be 0.9×1950 km = 1760 km, which is greater than the measured $r_i = 1350$ km. The streamlines will consequently not be able to adapt to pressure contours around the anticyclonic bend.

The tentative assumption of $r = r_{min}$ given above is really predicated on the further assumption that the wind maintains a speed of $2v_g$ around the anticyclonic bend. We can in this case show convincingly that the wind does not reach such a speed and therefore that the air trajectory must have a radius of curvature considerably longer than r_{min}.

The geostrophic wind in the strongest part of the straight southwesterly current on November 8, 1500Z amounted to about 70 m sec^{-1}, and on the chart for November 9, 0300Z a measurement of the geostrophic wind in the Great Lakes region gives nearly the same value. Even at the geostrophic speed of 70 m sec^{-1}, which gives a speed of 70 − 9 = 61 m sec^{-1} relative to the wave, it would take only 4½ hours for each air parcel to cover the 1000-km distance along which there is anticyclonic curvature. Suppose a particle passes the inflection point at 70 m sec^{-1} and from then on experiences a forward tangential acceleration $(dv/dt)_t = 2\Omega_z v_r$ on the anticyclonic bend. If v_r, the wind component directed outward normal to the isobars, reaches the high average value of 10 m sec^{-1} on the anticyclonic bend, the speed of the particles would increase at a rate of 10 m sec^{-1} per three hours, and at most by 15 m sec^{-1} during the whole travel from inflection point to inflection point. This increase in speed would thus go only one-fifth of the way from v_g to $2\,v_g$. This reasoning justifies the earlier assumption of the moderately supergeostrophic wind of 70 + 10 = 80 m sec^{-1} on the middle of the anticyclonic bend.

Another effect of the transisobaric wind component on the anticyclonic bend is also worth considering. A flow component across anticyclonic contours towards low pressure is usually synonymous with horizontal divergence of mass, and offers in that way a contribution to pressure fall (see equation (1)). The basic pattern in Fig. 1 of horizontal divergence in westerly waves would thus, in the levels of strongest westerlies, show the divergence extending forward beyond the ridge of highest pressure. The implications of this divergence on the pressure ridges of the upper atmosphere for the storm development in the lower atmosphere will be considered on p. 597.

Frontogenesis. Figure 10 illustrates three stages of frontogenesis, 24 hours apart, represented by simultaneous sea-level and 850-mb maps. At the first map time the Hudson Bay cyclone is also shown. Its thermal structure is that of an old cyclone with the occluded front beginning to wrap around the center. The upper warm tongue, extending east and north of the Hudson Bay center from the warm-air reservoir over the Atlantic, is shown clearly in the 850-mb isotherms. The pressure trough pointing southwards from the Hudson Bay cyclone is not of frontal nature, as can be seen from the weak temperature gradient at 850 mb in that region. The same nonfrontal trough continues up to the 300-mb level (Fig. 9), where its orientation approaches northwest-southeast, and in those upper layers it actually extends out over the Atlantic, producing a bend in the warm sector current. Such troughs always move more slowly than the air, and it is kinematically impossible for fronts to develop in them.

Historical continuity made it obvious that the cold front from the Hudson Bay cyclone had reached Florida by November 7, 1500Z, but the 850-mb map shows how the cold air, after arriving over the Gulf States, must have subsided and thereby effaced most of the frontal temperature contrast. The bundle of isotherms running along the northern part of the cold front bends westward over the Carolinas and northern Georgia and there marks the intersection of the 850-mb map with the tilting surface of subsidence. The 850-mb winds in that region blow across the isotherm bundle from cold to warm, but fail to produce any local fall in temperature because of the simultaneous sinking. The deviations from geostrophic flow are quite striking in that sinking air mass. While descending from the levels of strong west winds the air particles must be retarded and, in order to do so, they must move with a component towards high pressure, as shown on the 850-mb map of November 7, 1500Z. The same type of geostrophic departure is found on that day over the southeastern United States all the way up to 300 mb (Fig. 9). On the following day (November 8), the geostrophic departures characteristic of the front side of moving anticyclones can be seen on the 850-mb map over New England. In the rear of the moving anticyclone the opposite geostrophic deviation is observed. In that part the air is ascending and accelerating and must have a horizontal component towards low pressure. This phenomenon is actually part of the process of frontogenesis over the central United States which will be considered next.

Frontogenesis by horizontal advection operates when a field of deformation is maintained in a baroclinic air mass. Optimum efficiency in this process is achieved when the axis of dilatation of the field of deformation coincides with the direction of the isotherms. The

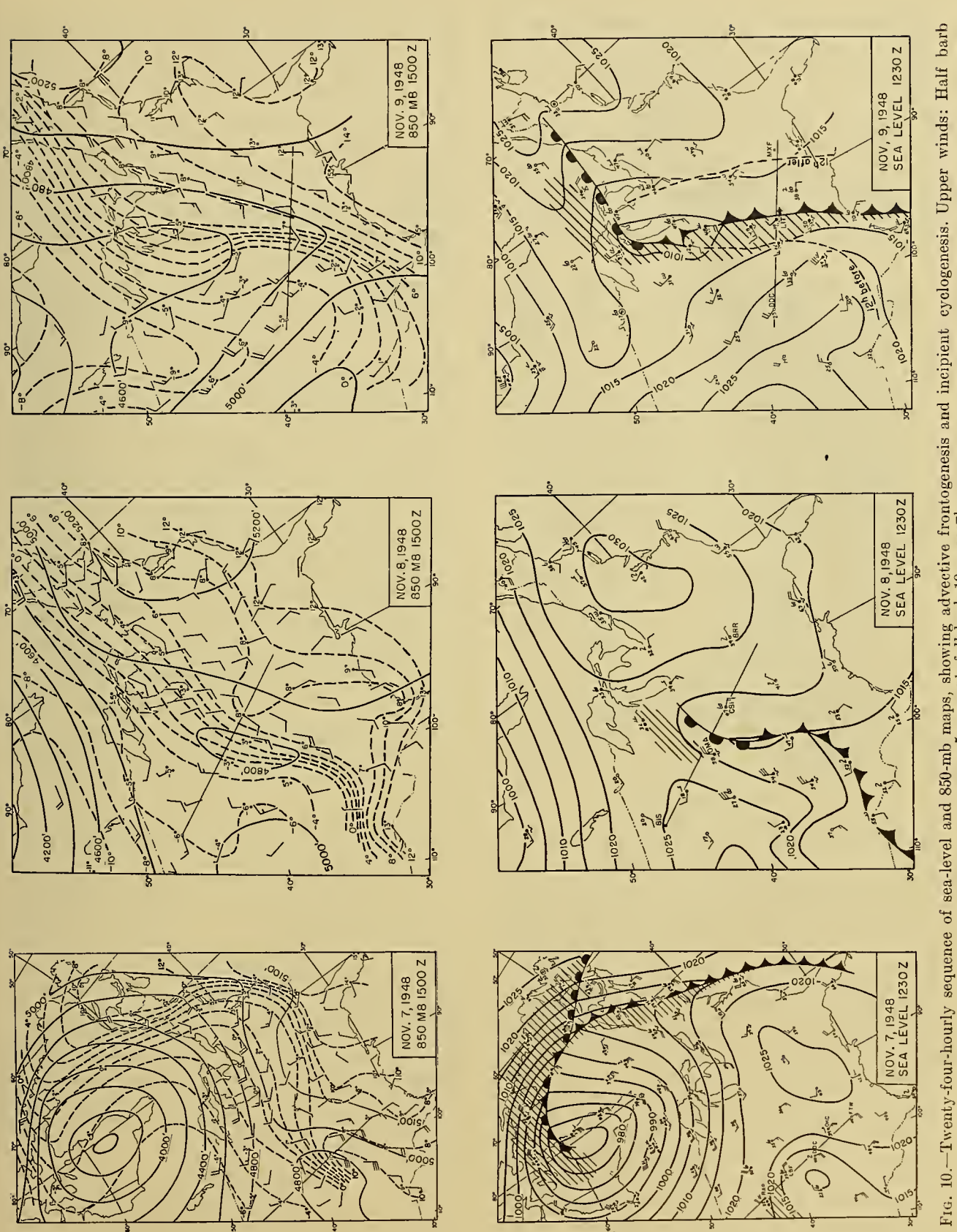


FIG. 10.—Twenty-four-hourly sequence of sea-level and 850-mb maps, showing advective frontogenesis and incipient cyclogenesis. Upper winds: Half barb 5 m sec^{-1}, full barb 10 m sec^{-1}.

streamlines in the col over the central United States on November 7, 1500Z, fulfill that condition fairly well. The mentioned geostrophic departure towards low pressure on the warm side of the col also favors the transportation of the isotherms towards the axis of dilatation. As a result of these processes a surface front has formed on November 8, 1500Z, over the region previously occupied by the col, while frontogenesis is in progress over the Great Lakes region where the col has now arrived. A frontal wave has already formed over the state of Missouri and is represented in the pressure field by a small elongated low. During the following twenty-four hours the new cyclone deepens and moves along the front northeastward. A new field of deformation is active on November 9 along the cold front of that cyclone and helps maintain the frontal temperature contrast.

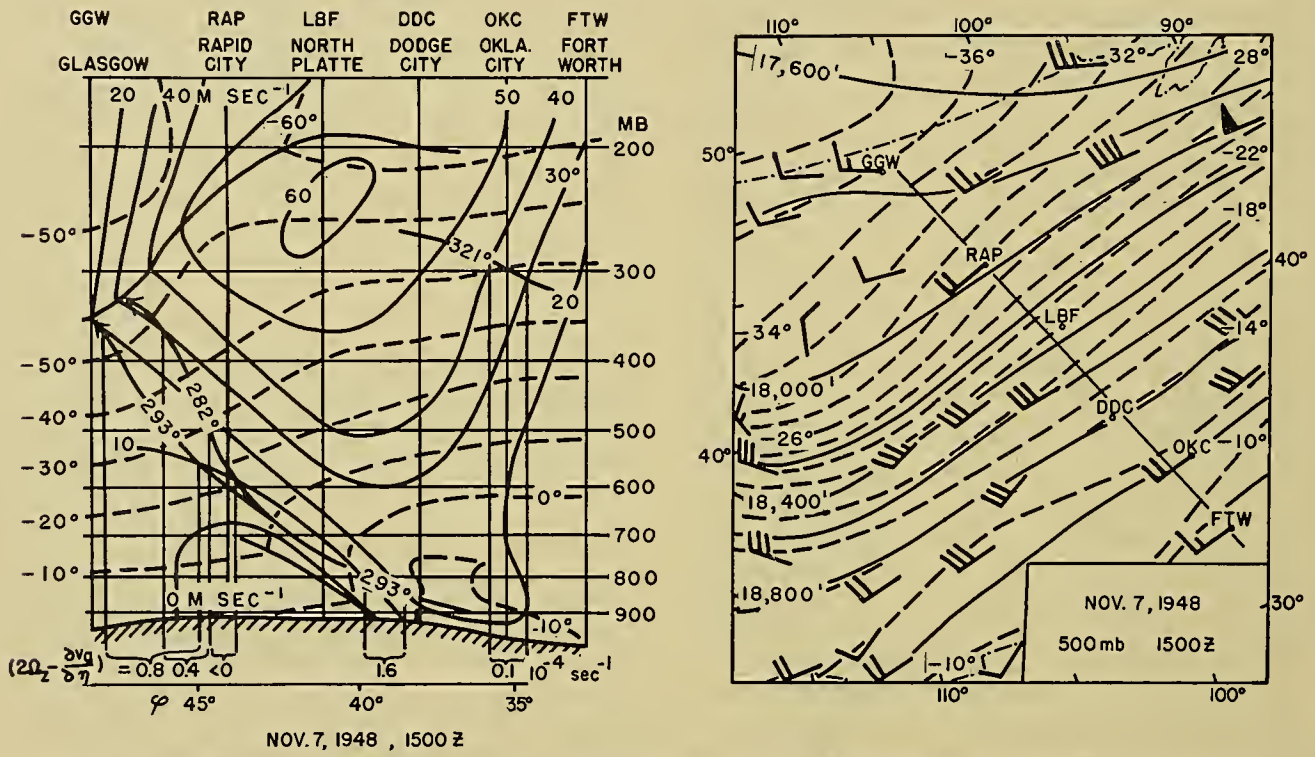


FIG. 11.—Profile of early frontogenesis, November 7, 1948, 1500Z, and part of the 500-mb map for the same time. Sample evaluations of $2\Omega_z - \partial v_g/\partial \eta$ below diagram. Upper winds: Half barb 5 m sec^{-1}, full barb 10 m sec^{-1}, triangular barb 50 m sec^{-1}.

The described frontogenetical development near the ground conforms with the advective rules set forth by Bergeron [1] and Petterssen [25]. We shall here add a study of the dynamical conditions for isentropic upgliding in the free atmosphere, which is an important part of the process of frontogenesis.

The rate of frontogenesis near the ground is increased considerably if the air of the lower part of the frontal zone is removed by upgliding. The dynamical possibilities for that process are considered in Fig. 11, which contains a profile across the zone of frontogenesis during its early stage on November 7, 1500Z. At that time no clear-cut front was yet discernible on the surface map, but on the 850-mb map there is great crowding of isotherms between the cold air over North Platte and the foehn air over Dodge City. The foehn on the warm side of the col gives the frontogenesis a good start in the layers below the level of the continental divide, but the process also takes place higher up, as shown by the isothermalcy between 600 and 560 mb at Rapid City.

The thermodynamics of upgliding in the sloping frontal zone can be tested through an inspection of the 293° dry-isentrope which has been inserted in Fig. 11. It shows that a particle could be brought dry-adiabatically along the profile from near the ground in Oklahoma to the tropopause at 350 mb over Montana without being subject to stabilizing gravity effects. If we take into account that condensation would begin in such a particle at 700 mb, the ascent from there on would follow the saturation isentrope of 282°, which climbs more steeply than the dry-isentrope and likewise reaches the tropopause. Actually no single particle undergoes such far-reaching isentropic displacements inside the profile; but the isentropes can still be used as indicators of the direction of the component of stable upgliding (see p. 585), which the particles may have in addition to their much stronger geostrophic component of motion normal to the profile.

A study of the values of the isentropic upgliding (equation (11)),

$$v_\eta = \frac{v_x \frac{\partial v_g}{\partial x} + \frac{\partial v_g}{\partial t}}{2\Omega_z - \frac{\partial v_g}{\partial \eta}},$$

along the different sections of the inserted isentropes, will reveal the dynamical possibilities for frontogenesis.

The denominator in the expression above can be determined uniquely from the data contained in the profile, whereas the numerator depends on derivatives of the geostrophic wind normal to the profile and derivatives in time. Let us consider the denominator first.

Along the lower part of the 293° isentrope the geostrophic shear $\partial v_g/\partial \eta$ is negative, and hence $2\Omega_z - \partial v_g/\partial \eta$ is large. Between the points of intersection of the 293° isentrope with the isovels of 20 m sec^{-1} and 10 m sec^{-1} the value of $2\Omega_z - \partial v_g/\partial \eta$ can be computed to be 1.6×10^{-4} sec^{-1}, as indicated at the bottom of the profile. Farther up along the 293° isentrope, $\partial v_g/\partial \eta$ changes sign and $2\Omega_z - \partial v_g/\partial \eta$ decreases despite the northward increase of $2\Omega_z$. In the baroclinic field between Rapid City and Glasgow $2\Omega_z - \partial v_g/\partial \eta$ has decreased to 0.8×10^{-4} sec^{-1}. Still smaller values are found along the 282° saturation isentrope, and a negative $2\Omega_z - \partial v_g/\partial \eta$ results in the section near Rapid City. The latter value indicates dynamic instability or, in other words, the condition of upgliding without dynamic brake action. Actually only small volumes of saturated air (altostratus and cirrus) occur in the region under consideration during the early stage of frontogenesis, and friction of such air against the dry environment probably exerts enough of a brake action to preclude violent developments. The dry-adiabatic upgliding thus still applies to the greater part of the baroclinic upper-tropospheric air.

The numerator in the expression for v_η can be judged from an inspection of the 500-mb map (in Fig. 11). If $v_x \partial v_g/\partial x$ is measured on the map just north of Rapid City, it amounts to $20 \times 2.6 \times 10^{-5}$ m sec^{-2} $= 5.2 \times 10^{-4}$ m sec^{-2}. The large value of the term comes from the convergence of the 500-mb contours and that feature, in turn, is inherent in the structure of the large pressure trough to the west with its central area of weak pressure gradient bordering on strong pressure gradients to the south. The large value of $\partial v_g/\partial x$ is, of course, also corroborated by measured wind velocities, which increase from 10 m sec^{-1} to 50 m sec^{-1} along the streamline from northern Wyoming to Green Bay, Wisconsin. In addition, the 300-mb map (Fig. 9) shows the same convergence of contours a little farther north.

An evaluation of v_η at the 500-mb level just north of Rapid City gives

$$v_\eta = \frac{5.2 \times 10^{-4}}{0.8 \times 10^{-4}} = 6.5 \text{ m sec}^{-1}.$$

Here $\partial v_g/\partial t$ has been neglected as insignificant in comparison with $v_x \partial v_g/\partial x$. The corresponding v_z would be about one-hundredth of v_η, hence 6.5 cm sec^{-1}. Corresponding determinations of v_η lower down on the frontal slope result in smaller values, and consequently $\partial v_\eta/\partial \eta$ is positive. With $\partial v_\eta/\partial \eta$ and $\partial v_x/\partial x$ both positive in the frontal zone, there is stretching in both the η-direction and the x-direction, so that frontogenesis progresses under optimum conditions.

In the upper troposphere south of the maximum westerlies $\partial v_g/\partial \eta$ assumes large values approaching those of $2\Omega_z$. In Fig. 11, $2\Omega_z - \partial v_g/\partial \eta$ measured at 300 mb over Oklahoma City is only 0.1×10^{-4} sec^{-1}. However, $v_x \partial v_g/\partial x + \partial v_g/\partial t$ is also small at that place (see Fig. 9) so that no great v_η-component results. Farther east near the Atlantic, where the anticyclonic isentropic shear is equally great and $v_x \partial v_g/\partial x$ has a large negative value, v_η is observed to have a large negative value (directed towards high pressure) at all reporting upper-wind stations. This is an example of the systematic nongeostrophic wind components at the "delta" of an upper jet stream derived in Fig. 8. The horizontal convergence resulting in the southern half of the delta is instrumental in providing the pressure rise ahead of the moving high in the southeastern United States (Fig. 10).

Figure 12 shows a profile through the zone of frontogenesis twenty-four hours later. The frontal slope has become steeper ($\frac{1}{50}$ in the lowest portion) and the frontal shear $-\partial v_x/\partial y$ is now characterized by a sharp 180° wind shift. Negative v_g values (northeast wind) stronger than 10 m sec^{-1} now occur in the lower part of the cold wedge near the front. Applying the v_η-formula to particles in the northeast current, we find conditions set for isentropic downgliding because the numerator $v_x \partial v_g/\partial x + \partial v_g/\partial t$ is now negative. A numerical estimate of the downgliding along the 281° isentrope near the cold edge of the frontal zone at 850 mb follows:

$$v_\eta = \frac{v_x \dfrac{\partial v_g}{\partial x} + \dfrac{\partial v_g}{\partial t}}{2\Omega_z - \dfrac{\partial v_g}{\partial \eta}} = \frac{-10 \times 1.4 \times 10^{-5} - 10^{-4}}{0.96 \times 10^{-4} - 0.1 \times 10^{-4}} = -2.8 \text{ m sec}^{-1}.$$

The resulting dry-isentropic descent traverses the frontal zone with a component from the cold to the warm side. This nongeostrophic component towards the frontal trough, together with the frictional flow component in the same direction, accounts for the subgeostrophic displacement of warm fronts in general. In some cases the nongeostrophic component normal to the front may permit the cold wedge to advance against a moderate geostrophic component from warm to cold.

In the upper part of the frontal zone, isentropic upgliding continues on November 8 just as on November 7, as can be seen from the convergence of contours between Omaha and Bismarck on the 500-mb map (Fig. 12). The same contour convergence is found right over Omaha on the 700-mb map (not reproduced). In the profile the 284° saturation isentrope approximately follows the warm edge of the frontal zone; $\partial v_g/\partial \eta$ measured along that isentrope gives values greater than $2\Omega_z$ from the condensation level up to 700 mb. This lower portion of the frontal zone is thus dynamically unstable; while higher up, where the 284° isentrope turns parallel to the v_g-isovels, finite speeds of upgliding can be determined. An estimate of the upgliding on the saturation isentrope of 284° at the 500-mb level gives

$$v_\eta = \frac{23 \times 1.4 \times 10^{-5}}{(1.0 - 0.1) \times 10^{-4}} = 3.6 \text{ m sec}^{-1}.$$

Exploring the whole frontal zone for nongeostrophic isentropic motion, we find the conditions for downgliding limited upwards by the zero isovel, while upzone is about $\frac{1}{50}$ in the upper portion, and it is quasivertical near the ground, while an intermediate portion around 700–800 mb tilts only by $\frac{1}{130}$. The general

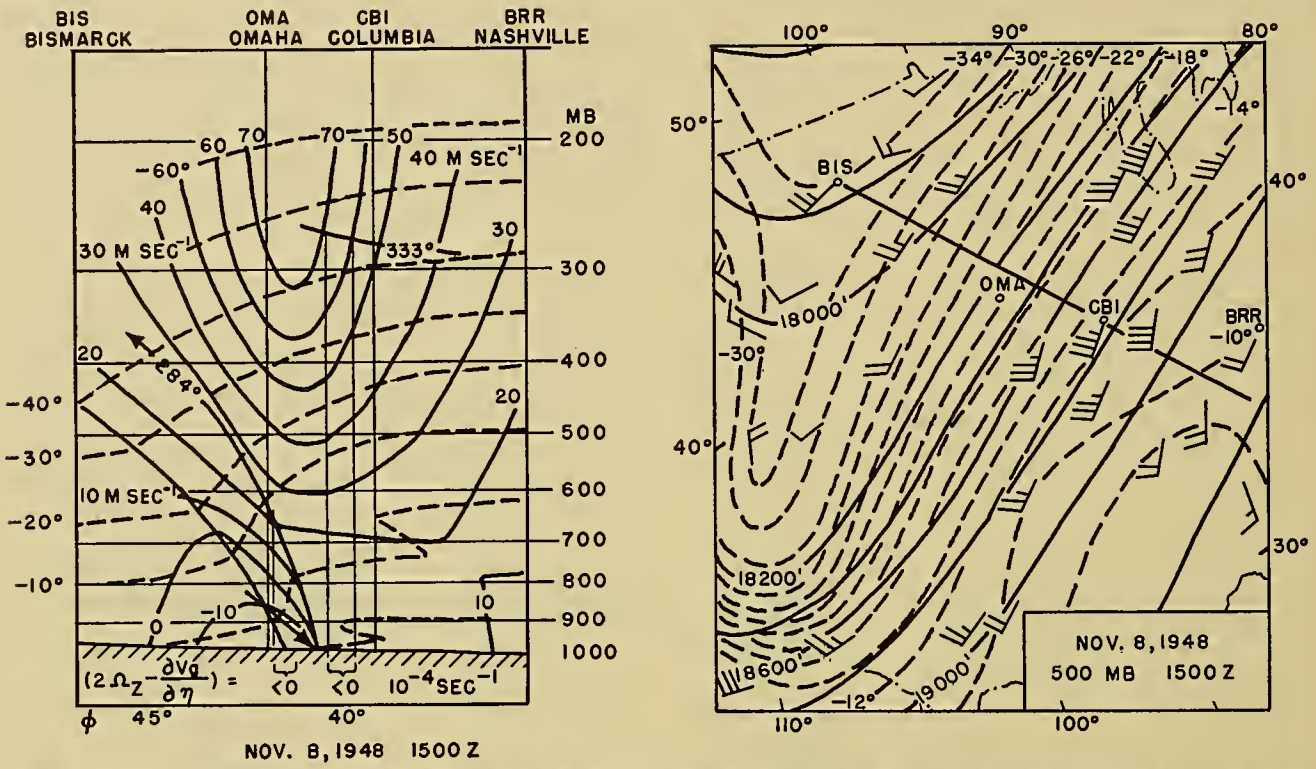


FIG. 12.—Profile of the warm front of the incipient wave on November 8, 1948, 1500Z, and part of the 500-mb map for the same time. Sample evaluations of $2\Omega_z - \partial v_g/\partial \eta$ below diagram. Upper winds: Half barb 5 m sec^{-1}, full barb 10 m sec^{-1}.

gliding begins above that line. Moreover, we find that the dry-isentropes indicate a downgliding with a horizontal component across the front from cold to warm, while the upgliding along saturation isentropes remains almost parallel to the frontal slope. The frontal profile therefore must begin to bulge forward from cold to warm in the lower layers while remaining rather unchanged higher up. That is what happens when the apex of the frontal wave goes by, as will be discussed in connection with the maps in Fig. 14.

The anticyclonic shear south of the westwind maximum has grown to the state of dynamic instability as shown in Fig. 12 by means of the differentiation of v_g along the 333° isentrope. Shears rather close to the instability limit also extend down to the 500-mb level and make possible the rather large and systematic wind component towards low pressure observed on the 500-mb map southeast of the jet stream.

Figure 13 shows a profile across the cold front from Dodge City (Kansas) to Maxwell Field (Alabama) on November 9, 1500Z. The corresponding sea-level and 850-mb maps are to be found in Fig. 10. The profile shows that strong horizontal temperature gradients have formed all the way to the top of the diagram. The frontogenetical process by horizontal advection has actually been operating through the whole troposphere (in Fig. 14 the resulting frontal zone shows up well even at the 300-mb level). The inclination of the frontal shape of the profile can be understood as the result of the bulging forward of the lower portion of the cold wedge after the passage of the frontal wave apex. The

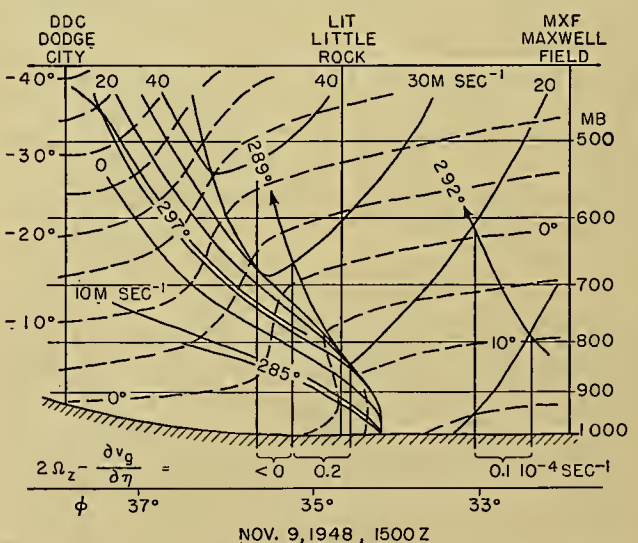


FIG. 13.—Profile of the cold front on November 9, 1948, 1500Z. Sample evaluations of $2\Omega_z - \partial v_g/\partial \eta$ below diagram.

nongeostrophic downgliding responsible for that process was found to be dynamically justified from the study of the frontal profile 24 hours earlier, $v_x \partial v_g/\partial x$ +

$\partial v_g/\partial t$ being negative in the lower portion of the cold wedge. The sea-level map for November 9, 1500Z (Fig. 10) shows a positive $\partial v_g/\partial x$ along the whole cold front (the geostrophic wind component parallel to the front in the cold air is increasingly negative as we pass towards the negative x-direction), but the actual wind component v_x, parallel to the front, is about zero in the forward part of the cold wedge; $\partial v_g/\partial t$ also is small. Hence it follows that the isentropic downgliding v_η should be insignificant except where $2\Omega_z - \partial v_g/\partial \eta$ is zero or negative. The profile in Fig. 13 shows an almost perfect parallelism of v_g-isovels and dry-isentropes in the cold wedge, so that $\partial v_g/\partial \eta = 0$. Therefore, no dynamic instability occurs inside the cold air, not even in the upper part where the frontal zone is quite steep. The only place for dynamic instability to occur inside the cold air is right at the quasi-vertical part of the frontal surface near the ground, where a real temperature discontinuity exists. However, the air volume involved is too small to appear on a profile of the scale used in Fig. 13. The release of the dynamic instability at the cold front is responsible for maintaining the downdraft, which is always observed in a strip of a few kilometers' width following the frontal passage. This cold-air downdraft is instrumental in giving the cold front a greater speed of displacement than would have been indicated by the geostrophic wind determined from the sea-level pressure distribution. In Fig. 10 the computed twelve-hour geostrophic displacement of the cold front is represented by a short arrow of 80-km length, while at the same place the preceding twelve-hour displacement was 210 km and the subsequent one 460 km. The geostrophic wind component normal to the front computed from the 850-mb map is large enough to account for the frontal displacement, and we must assume that air from this level enters the frontal downdraft and carries westerly momentum to the surface layer. In the layers above 850 mb the cold-air current is cyclonically curved and hence subgeostrophic. With increasing height, the wind in the frontal zone also becomes more and more parallel to the frontal boundary, so that the frontal displacement is less there than at the ground.

The prefrontal air in the profile in Fig. 13 has a lapse rate slightly less than the saturation adiabatic, but with the existing horizontal temperature gradient a saturation-adiabatic ascent is possible at the rather steep angle of $\frac{1}{50}$, as shown by the sample saturation isentropes of 289° and 292°. The measured values of $2\Omega_z - \partial v_g/\partial \eta$ show a close approach to dynamic indifference and even some dynamic instability. A saturation-isentropic upgliding is in order at 850 mb, as can be seen from the convergence of contours ($v_x \partial v_g/\partial x > 0$) in the warm current intersected by the profile (Fig. 10). The same is true for the 700-mb map (not reproduced) but not for the 500-mb and 300-mb maps (Fig. 9). The frontal upgliding should therefore be confined to the layers under 500 mb. This is also verified by the Little Rock sounding, which goes up through the cold-front rain but shows a 35 per cent relative humidity at 500 mb. Thunderheads growing up from the cloud mass of the cold front would, of course, go well beyond 500 mb, but such phenomena of "vertical instability" were not reported in the case under consideration. Intermittent, light prefrontal rain, which was reported as far as 300 km ahead of the cold front, can be accounted for quite well by the saturation-isentropic upgliding. In the warm season such upgliding in the tropical air current may be sufficient to trigger thunderstorm formation, which in turn may develop prefrontal squall lines.

Structure of the Maturing Frontal Cyclone. Figure 14 presents the sea-level, 700-mb, 500-mb, and 300-mb maps for November 10, 0300Z, depicting the structure of the maturing frontal cyclone. The amplitude of the frontal wave on the surface map has now increased considerably, and the cold air from the rear begins to encircle the cyclone center. It is clearly seen how this occlusion process has had its inception only at the ground, while the isotherm patterns of the upper maps still indicate an open wave of small amplitude. This shows that while the wave travels along the front the frontal slope increases to a maximum at the wave apex. At that point the smooth wave "breaks" and a relatively shallow cold outbreak fans out along the ground. The mechanism of this breaking of the smooth wave probably lies in the dynamic instability of the lower part of the frontal zone, described in its stage of inception in Fig. 12 and continuing in the form of the frontal downdraft in Fig. 13. During the process, the cold front part near the cyclone center undergoes frontolysis through cold-air downgliding. This is always noticeable in the surface-map analysis, and, on November 10, 0300Z, it also shows up in a weakening of the frontal temperature gradient on the 700-mb map. Farther south, where the cold front passes through the frontogenetical field of deformation, the frontal temperature gradient remains rather strong. The strongest frontal temperature gradient, however, is found at the 500-mb level where the breaking of the frontal wave has not yet started. The 300-mb map also shows fairly strong temperature gradients across the pressure trough, which may justify the use of the term "front" even at that level. Particularly striking is the crowding of isotherms in the southwestern corner of the map, probably an effect of frontogenesis in a 300-mb col in the unmapped area west of Mexico. The 300-mb map under consideration is just tangent to a tropopause depression over the cold tongue of tropospheric air in the west. The map intersects the tropopause along the zone of lowest temperature across Labrador and northern Ontario. The temperature maximum east of the Hudson Bay cyclone is stratospheric. The cold air to its north is tropospheric and has been brought from lower latitudes as part of the warm sector shown in Fig. 9. On the 500-mb map, which is entirely tropospheric, the Hudson Bay center has a cold core with warmer surroundings both to the north and to the south.

The pressure minimum of the frontal wave at the Great Lakes continues up to 700 mb with some westward tilt, but at that level it has already shrunk so as to be a minor feature in the pressure field compared to

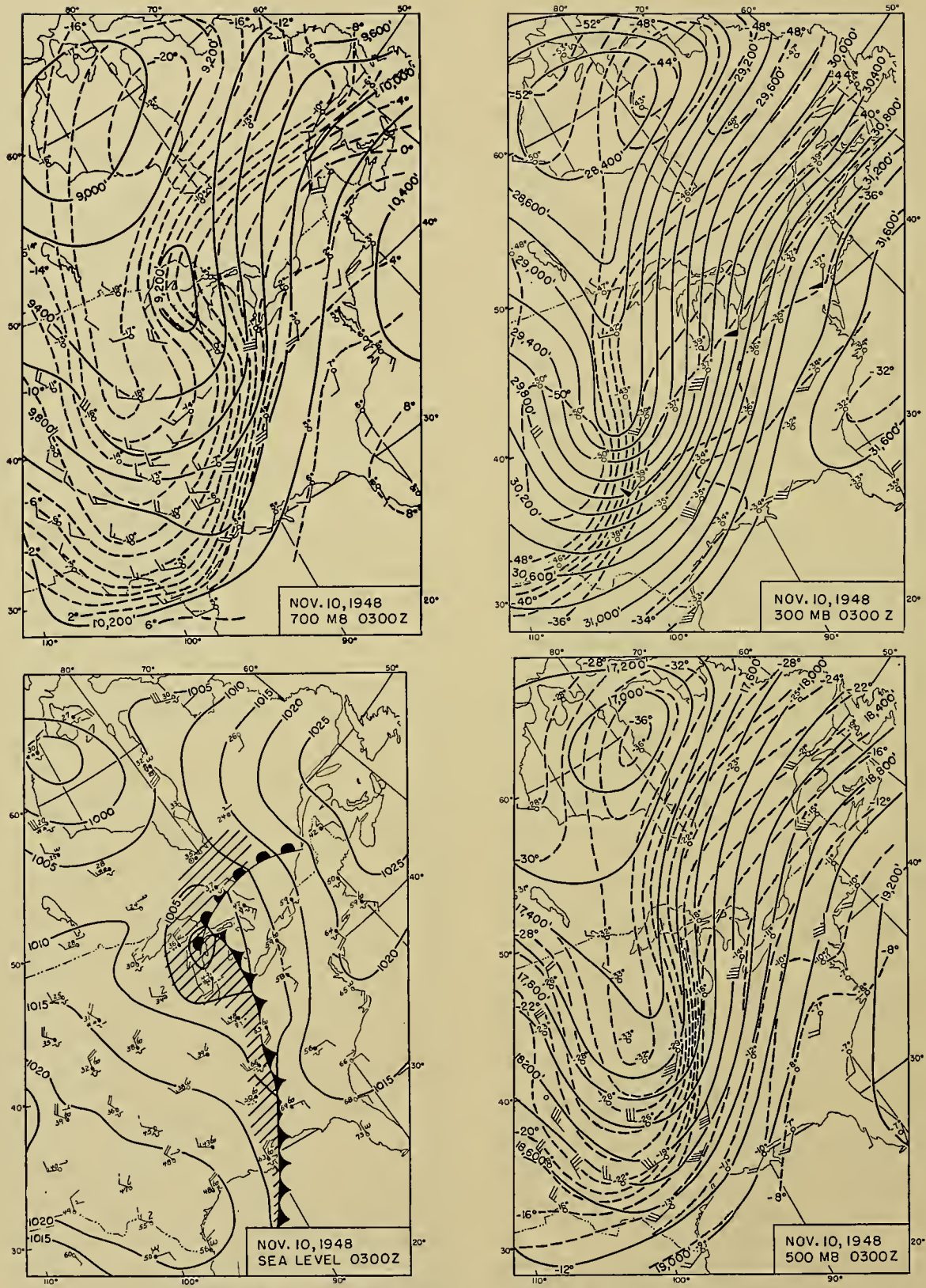


FIG. 14.—Structure of maturing frontal cyclone at sea level, 700 mb, 500 mb, and 300 mb, on November 10, 1948, 0300Z. Upper winds: Half barb 5 m sec^{-1}, full barb 10 m sec^{-1}, triangular barb 50 m sec^{-1}.

the old stationary Hudson Bay minimum. But the young cyclone over the Great Lakes has all the potentialities of development inherent in the solenoid field concentrated along the frontal wave all through the troposphere. As a hydrostatic consequence of the wave-shaped pattern of isotherms, the upper current above the closed center is likewise wave-shaped with an anticyclonic bend over the warm-front area of upgliding.

The frontal cyclone deepened at the rapid rate of 14 mb per twelve hours during November 10, and ended up as a storm center of 975 mb over northern Labrador on November 12. We shall briefly consider the application of the various theories of upper-air divergence which may claim to explain the deepening. Considering first the formation of the upper pressure crest which precedes the surface cyclone, we shall see the nature of the interplay between lower and upper layers. The fact that the upper pressure crest moves at a speed of only 9 m sec^{-1} in a current of 70–80 m sec^{-1} shows that the upper wave cannot be a free one. The time and place of the first appearance of the pressure crest on the 300-mb map (November 9, 0300Z) make it likely that the upper crest was formed by the upward motion connected with the beginning frontal upgliding in the nascent frontal wave. Once the upper wave is established, upper divergence will be located in the area between the pre-existing upper pressure trough and the new upper pressure crest ahead of it. With the trough in north-northeast–south-southwest orientation and the pressure crest in northwest–southeast orientation, the half wave length between them shortens northward and makes the upper divergence particularly strong in that region. This is also the region where the rapid deepening of the frontal cyclone takes place. So far, the dynamics involved in the deepening process conform with the principles of Fig. 1. With that basic pattern accepted, we must also admit the existence of the following modifying processes.

The upper pressure crest is continually being fed from below by the rising motion in the front half of the cyclone and therefore is forced to maintain the same slow speed of propagation as the surface vortex. When the curvature of the anticyclonic bend of upper isobars becomes sufficiently strong, the fast upper current is unable to follow the isobars in gradient wind fashion, and horizontal divergence must result on both sides of the upper crest line (see p. 590). This component of upper divergence modifies the model in Fig. 1 in the sense of extending the pressure fall farther ahead of the surface center.

The upper-air divergence of the Ryd-Scherhag theories (p. 586) is also superimposed on the divergence and convergence inherent in the wave pattern. It can best be judged by comparing the geostrophic wind at two successive inflection points. The average geostrophic wind between the 30,000-ft and 29,000-ft contours at the inflection point southwest of the Great Lakes was 60 m sec^{-1}, and between the same contours at the inflection point over Labrador it was 42 m sec^{-1}. Therefore it can be concluded that the Great Lakes cyclone was situated under a "delta" of the upper current, and that an upper divergence pattern in the style of Fig. 8 would be superimposed. This upper pattern would tend to produce pressure falls in the northern half and pressure rises in the southern half of the delta. It may be counted in favor of this reasoning that the Great Lakes cyclone proceeded northeastward to northern Labrador and not along the 300-mb contours ahead of the storm which would have meant east-northeastward propagation.

The sample cyclone described above had its early development near the ground, where the frontal upgliding of the warm air and the downgliding of the cold were a direct consequence of the preceding frontogenesis. Such wave cyclogenesis is typical for all the frontogenetical areas of the middle and high latitudes. Another type of cyclogenesis occurs at times independently of any pre-existing low-level front. The cyclogenetic mechanism in that case seems to lie in the dynamic instability of the upper tropospheric westerlies, which leads to the meandering of the current and the subsequent formation of an upper cold low. Cyclogenesis of this kind is described by means of two synoptic examples in the following article by E. Palmén. In one example, that of November 4, 1946, the upper low did not reach a frontogenetical area, and did not extend down to the surface. In the other case, that of November 17–18, 1948, a frontal wave action can be discerned but, in contrast to the case of November 7–10, 1948, the wave motion started first in the upper troposphere and later extended to low levels.

CONCLUSION

Summing up our state of knowledge of extratropical cyclones, we may classify their formation as being due either to unstable frontal wave action or to unstable growth of an upper wave trough. A subsequent combination of both processes is quite frequent, and all the strongest cyclones on record seem to have that double origin. This article has dealt mainly with the unstable frontal wave action, which is in itself a large subject. The descriptive study of cyclogenesis and cyclone growth is carried out daily by all the weather forecasters of middle latitudes, and an adequate report on their experiences would have exceeded by far the scope of this article. The theoretical study of the model of frontal cyclogenesis has been the work of a small number of scholars of dynamical meteorology. Viewed in retrospect, the contribution of H. von Helmholtz to this field of research appears to be of outstanding importance and to represent the foundation upon which contemporary and future theories should build. The Helmholtzian concept of dynamic instability of the inertial motion in isentropic surfaces points out which properties of the atmospheric fields of mass and motion on a rotating earth are conducive to the growth of perturbations from infinitesimal to finite amplitudes; and it seems obvious that such a fundamental theoretical principle must have its applications to the early phases of the life cycle of cyclones. However, there is still a wide gap to be bridged between the existing theory of dynamic instability and the applications called

for in daily synoptic practice. Helmholtz considered only the background pattern of an atmosphere moving zonally at all longitudes. What synopticians need are the criteria of dynamic instability for large-scale atmospheric currents which are nonzonal and nonpermanent, because they are part of a slowly moving, long-wave pattern, but still offer the propitious environment for the quick growth of perturbations of cyclone size.

Once the frontal wave has occluded, dynamic instability in the Helmholtz sense is eliminated, because of the accomplished spread of cyclonic vorticity to the whole cyclonic area. Indirect effects of dynamic instability may, however, be brought to the cyclone from the neighboring upwind anticyclonic part of the upper-tropospheric westerlies. In the absence of such a rejuvenating influence, the cyclone is left to decay by frictional inflow at the surface unopposed by deepening effects in the upper troposphere. This sketch of the last part of the life cycle of cyclones is even less intimately related to exact dynamic theories than is the existing theory of frontal cyclogenesis. The most hopeful approach to a solution of the problems of the old cyclone is most likely to be through the study of the energy transformations in the cyclone and its farther environment. The major old cyclones occupy such big fractions of the volume and weight of the atmosphere that the "farther environment" must be taken to mean the whole hemispherical circulation.

REFERENCES

1. Bergeron, T., "Über die dreidimensional verknüpfende Wetteranalyse." *Geofys. Publ.*, Vol. 5, No. 6 (1928).
2. Bjerknes, J., "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54: 462–466 (1937).
3. —— and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
4. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4: 135–162 (1947).
5. Cressman, G. P., "On the Forecasting of Long Waves in the Upper Westerlies." *J. Meteor.*, 5: 44–57 (1948).
6. Eliassen, A., "The Quasi-static Equations of Motion with Pressure as Iudependent Variable." *Geofys. Publ.*, Vol. 17, No. 3, pp. 30–34 (1949).
7. Ertel, H., Jaw, J.-J., und Li, S.-z., "Tensorielle Theorie der Stabilität." *Meteor. Z.*, 58: 389–392 (1941).
8. Fjørtoft, R., "On the Deepening of a Polar Front Cyclone." *Meteor. Ann.*, 1: 1–44 (1942).
9. —— "On the Frontogenesis and Cyclogenesis in the Atmosphere. Part I—On the Stability of the Stationary Circular Vortex." *Geofys. Publ.*, Vol. 16, No. 5 (1946).
10. Haurwitz, B., "The Motion of Atmospheric Disturbances." *J. mar. Res.*, 3: 35–50 (1940).
11. —— *Dynamic Meteorology*. New York, McGraw, 1941. (See p. 234)
12. Helmholtz, H. v., "Über atmosphärische Bewegungen." *Meteor. Z.*, 5: 329–340 (1888).
13. Hesselberg, T., und Friedmann, A., "Grössenordnung der meteorologischen Elemente." *Veröff. geophys. Inst. Univ. Lpz.*, 1: 147–173 (1914).
14. Höiland, E., "On the Interpretation and Application of the Circulation Theorems of V. Bjerknes." *Arch. Math. Naturv.*, 42: 25–57 (1939).
15. —— "On the Stability of the Circular Vortex." *Avh. norske VidenskAkad.*, No. 11 (1941).
16. Holmboe, J., and others, *Dynamic Meteorology*. New York, Wiley, 1945. (See pp. 324–325)
17. Holmboe, J., "On Dynamic Stability of Zonal Currents." *J. mar. Res.*, 7: 163–174 (1948).
18. Kleinschmidt, E., "Stabilitätstheorie des geostrophischen Windfeldes." *Ann. Hydrogr., Berl.*, 69: 305–325 (1941).
19. —— "Zur Theorie der labilen Anordnung." *Meteor. Z.*, 58: 157–163 (1941).
20. Miller, J. E., "Studies of Large-Scale Vertical Motions of the Atmosphere." *Meteor. Pap. N. Y. Univ.*, Vol. 1, No. 1, 49 pp. (1948).
21. Moltschanow, P., "Bedingungen des Gleichgewichts und der Stabilität der Luftmassen nach der Horizontalen und der Vertikalen." *Petermanns Mitt.*, Erg. 47, Heft 216, SS. 62–67 (1933).
22. Palmén, E., "Aerologische Untersuchungen der atmosphärischen Störungen." *Acta Soc. Sci. fenn.*, Phys. Math., Vol. 7, No. 6 (1933).
23. —— "On the Distribution of Temperature and Wind in the Upper Westerlies." *J. Meteor.*, 5: 20–27 (1948).
24. —— and Newton, C. W., "A Study of the Mean Wind and Temperature Distribution in the Vicinity of the Polar Front in Winter." *J. Meteor.*, 5: 220–226 (1948).
25. Petterssen, S., *Weather Analysis and Forecasting*. New York, McGraw, 1940. (See pp. 238–248)
26. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semipermanent Centers of Action." *J. mar. Res.*, 2: 38–55 (1939).
27. Rossby, C.-G., "Kinematic and Hydrostatic Properties of Certain Long Waves in the Westerlies." *Dept. Meteor. Univ. Chicago, Misc. Rep.* No. 5 (1942).
28. Ryd, V. H., "Meteorological Problems, II—Travelling Cyclones." *Medd. danske meteor. Inst.*, No. 5, 124 pp. (1923).
29. —— "The Energy of the Winds." *Medd. danske meteor. Inst.*, No. 7 (1927).
30. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose*. Berlin, Springer, 1948. (See pp. 184–187)
31. Solberg, H., "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones." *P. V. Météor. Un. géod. géophys. int.*, Edimbourg, 1936. II, pp. 66–82 (1939).
32. Van Mieghem, J., "Forme intrinsèque du critère d'instabilité dynamique de E. Kleinschmidt." *Bull. Acad. Belg. Cl. Sci.*, 5e sér., 30: 19–34 (1944). (For several subsequent papers see references in "Hydrodynamic Instability" by J. Van Mieghem in this Compendium.)

THE AEROLOGY OF EXTRATROPICAL DISTURBANCES

By E. PALMÉN

Academy of Finland

The present article deals mainly with the role of extratropical disturbances as links in the general atmospheric circulation and as cells for meridional exchange of air masses. Because of limited space we cannot discuss the subject in detail. Since there already exists a large amount of literature on the structure and behavior of cyclones in the surface layers, most emphasis has been placed in this article on the structure of disturbances in the free atmosphere. Most of the ideas and viewpoints expressed here can be found in the extensive writings dealing with extratropical cyclones. However, it has not been possible to refer directly to more than a portion of this material.

This article was prepared in final form while the writer was under appointment as visiting professor at the University of Chicago. The writer is indebted to Mr. C. W. Newton for his valuable help in preparing many of the figures in the article and especially for his analysis of the cyclone of November 17–19, 1948.

CHARACTERISTIC TEMPERATURE AND WIND DISTRIBUTION IN THE WEST-WIND ZONE

The formation of extratropical cyclones is connected with fronts separating air masses of different origin.[1] The most important front is that which separates air masses of polar origin from those of tropical origin. This "polar front" can, at least during the colder season, be followed relatively easily around half or more of the Northern Hemisphere on daily synoptic surface maps. The breaks in the polar front are of great importance. In their classical paper on the polar-front theory, J. Bjerknes and H. Solberg [7] pointed out that these breaks in the polar front are necessary in order to permit an exchange of air masses between the polar and tropical source regions.

From studies of synoptic charts for different levels one can conclude that the outflow of polar air from the polar source region occurs particularly in the regions between two polar fronts separated by a cold anticyclone. The polar outflow here appears in the southern parts as a relatively shallow layer of subsiding air masses. On the other hand, the inflow of tropical air into the polar region at higher latitudes usually appears as currents of warm air in the upper atmosphere. This can be seen from statistical studies of the frequency of the two principal air masses at different levels [35]. Frontal analyses of upper-air charts indicate that the area occupied by tropical air increases with height at all latitudes. Since the characteristic difference between polar and tropical air at a given latitude could not be maintained for a long time without advection, the only possible explanation for the above-mentioned fact seems to be the existence of a systematic meridional circulation of the type outlined here.

It is necessary to emphasize that the meridional circulation appears as a statistical fact. The average meridional circulation is naturally not the same at all longitudes nor at all times during a certain season. The circulation has a tendency to break up into several cells with quasi-vertical axes, depending upon the geographical distribution of continents and oceans and the formation of disturbances [11]. Thus the paths of individual air parcels are extremely complicated and the different air parcels carry out various kinds of complicated oscillating movements before they have performed a complete meridional circulation.

If there is a statistically opposite meridional movement in the lower and upper troposphere, obviously there must also be a surface of zero average meridional movement, separating the northward flow in the upper troposphere from the southward flow in the lower troposphere. This surface has a tendency to become a frontogenetical surface. Since the meridional movements discussed here are average displacements for longer periods, the separating surface naturally cannot have the character of a real front. However, synoptic experience shows that in nature there is a strong tendency to concentrate the contrasts between the air masses of tropical origin and those of polar origin in a relatively thin layer of transition. This transitional layer must have a certain inclination toward the horizontal surfaces because of the earth's rotation. On a nonrotating earth, frontal surfaces of the type observed in the atmosphere, with an inclination of the order of $\frac{1}{50}$–$\frac{1}{200}$, could never form.

It is obvious that there must be a transformation of polar air into tropical air at low latitudes and that the opposite must be true at high latitudes. The outflow of polar air into the subtropics takes place in the lower troposphere and the inflow of tropical air into the polar regions takes place in the middle and upper troposphere, and perhaps partly in the lower stratosphere. Therefore there cannot be any continuous frontal surface around the whole hemisphere in the vicinity of the earth's surface or in the upper atmosphere. However, in other parts of the troposphere there is nothing in principle against the idea of having a continuous polar front around the entire hemisphere.[2]

1. If taken in the broadest meaning of the word, cyclones or depressions of different types can form without any preexisting fronts or frontal zones. These types of irregular, generally weak depressions will not be discussed here.

2. The concept of a front is here used in the broad sense; it is obvious that a distinct upper polar front can be expected only in special situations favoring an extreme concentration of the air-mass distinctions.

Another point of importance should be noted here. The source region of polar air is smaller in area than the source region of tropical air. Therefore any outflow of polar air into lower latitudes must have a tendency to break up into "streams" embedded in tropical air. These streams of polar air often select "preferred regions" determined by geographical factors. Several such preferred regions for polar outflow can be found in the Northern Hemisphere. It is through these regions of preferred polar-air outflow that the principal frontogenetical regions on the earth's surface are determined, as Bergeron pointed out in his outline of dynamic climatology [3].

Analyses of synoptic charts for different levels indicate that in the region between two cyclone families, where no surface polar front exists, pronounced frontal zones appear on the charts for the middle troposphere (*e.g.*, at the 700- and 500-mb surfaces). In such cases the polar front appears to be interrupted only in the surface layers and not at higher levels in the atmosphere (compare Figs. 5 and 6). Since this difference between the polar front on the surface and in the free atmosphere is essential for an understanding of the three-dimensional air movement in cyclones, it is necessary to discuss the problem of meridional movement of air masses before we go further in our description of the characteristic structure of the atmosphere.

If ζ denotes the vertical component of the relative vorticity, ϕ the latitude, Ω the angular velocity of the earth's rotation, and D the depth of a given air mass (here regarded as incompressible), the principle of conservation of potential vorticity, according to Rossby [52], can be expressed by the equation

$$\frac{d}{dt}\left(\frac{\zeta + 2\Omega \sin \phi}{D}\right) = 0. \tag{1}$$

From (1) it follows that the individual change of vorticity is

$$\frac{d\zeta}{dt} = -\frac{2\Omega \cos \phi}{a} v + \frac{1}{D}\frac{dD}{dt}(\zeta + 2\Omega \sin \phi), \tag{2}$$

where v is the meridional wind component and a the radius of the earth.

The first term on the right in (2) gives the change of vorticity due to the meridional motion; the second term, the effect resulting from the change in depth of the air column. If we assume that the air moves without change in depth, the vorticity increases for movement toward the south. If, in order to simplify the discussion, we assume that the increase in vorticity appears as increasing curvature, the air parcel continuously curves to the left if v is negative (north wind). It is then easy to show that any air column moving without shrinking must ultimately bend back after reaching a lowest latitude. According to equation (1), this latitude is determined by the initial latitude, meridional wind component, and vorticity of the air parcel.

The change in depth of an air column moving southward can be computed from equation (2) if the vorticity along the trajectory is known. In the special case of a straight flow without horizontal shear (relative vorticity $\zeta = 0$) we obtain

$$\frac{1}{D}\frac{dD}{dt} = \frac{v}{a} \operatorname{ctn} \phi. \tag{3}$$

This equation determines the shrinking of a polar air column moving with the north-south velocity v along a trajectory of zero relative vorticity.

If we use either equation (1) or (3) for $\zeta = 0$, it is possible to compute the depth of a polar air mass at different latitudes as a function of its original depth at a given latitude. If we assume a depth of 8 km at latitude 60°N, we obtain the following values for D at lower latitudes:

Latitude N (deg)	60	50	40	30	20	10
Depth (km)	8.0	7.1	5.9	4.6	3.1	1.6

The foregoing values are very approximate. If we permit an increase in relative vorticity, higher values of the final depth of the polar air result. On the other hand, however, air which descends from the level of 8 km to, for example, 5 km will undergo a temperature increase of about 30C. In other words, the air will no longer be very cold at that lower level. Thus there must be some limit, determined by the original temperature, for the sinking of the polar air.

From the very simplified reasoning above, it follows that there must be some southern limit for the extension of characteristic polar air in the upper-air charts. This southern limit for the 500-mb level is somewhere around latitude 30°N. Systematic analyses of circumpolar 500-mb charts indicate that air masses with the characteristic polar air temperature very seldom reach latitude 30°N, and that 20°–25°N represents the southernmost limit for real polar air at the 700-mb level.[3]

It is now easier to understand the structure of the atmosphere in the area between two cyclone families or surface polar fronts. The breaks in the surface polar front which permit outflow of polar air into the tropics are not necessary at upper levels where the polar air flows southward as a subsiding current. At the 500-mb level, for example, a southern boundary for the polar air can be maintained around the hemisphere, at least in principle.

Many meteorologists who regularly use fronts on surface maps are inclined to reject the idea of upper fronts existing around the hemisphere. Their opinion is founded on the idea that surface friction is necessary for the formation of distinct fronts. It is obvious that surface friction contributes to the sharpening of fronts. In the free atmosphere, however, other frontogenetical factors can be as effective as those in the surface layer, or sometimes even more effective (*cf.* Bergeron [2] and Petterssen [46]). On the other hand it is evident that the great dynamic significance of fronts, confirmed in

3. In this reasoning the nonadiabatic processes have been completely disregarded. A southward-moving polar air mass naturally undergoes a gradual change and eventually completely loses its polar character because of the heat transfer from below.

daily synoptic work, could hardly be explained if fronts were phenomena restricted to the surface layer. Systematic use of the frontal method of analysis of upper-air charts, at least up to 500 mb, has shown that fronts are not only surface phenomena.[4]

The very valuable daily synoptic weather maps published by the United States Weather Bureau in cooperation with the Army, Navy, and Air Force confirm the existence of a band of strong isotherm concentration at the 500-mb level. Over large parts of the Northern Hemisphere this band shows all the characteristics of a real frontal zone; in other parts, however, it is more diffuse. Since this band of crowded isotherms undergoes strong meridional displacements from day to day, it

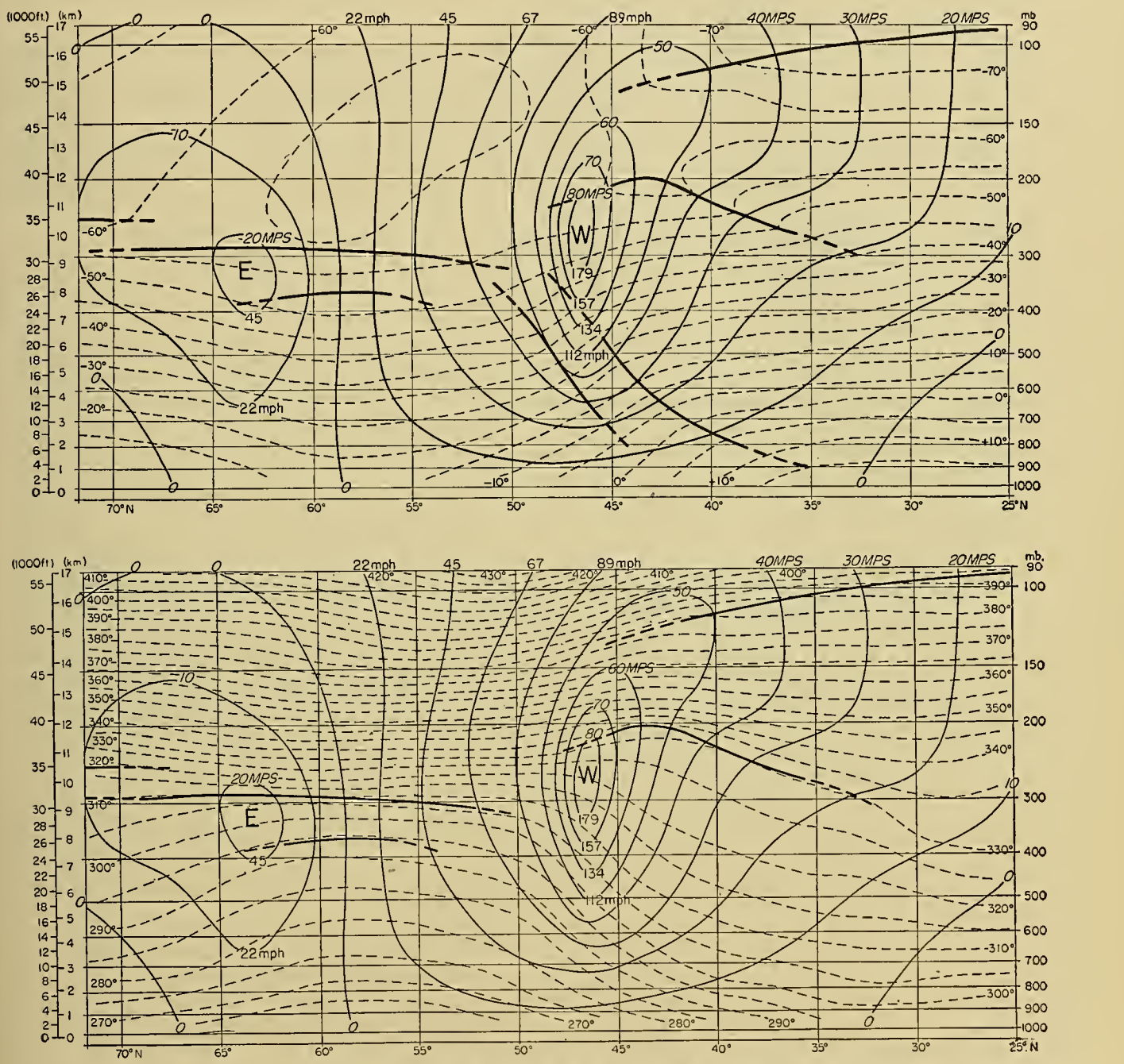


FIG. 1.—Mean cross section for 0300 GMT, November 30, 1946, showing average distribution of geostrophic westerly wind and of temperature over North America from latitude 25° to 75°N in a case of approximately straight westerly flow. Heavy lines indicate boundaries of the principal frontal layer (polar front) and tropopauses. Thin solid lines indicate velocity of westerly wind component (meters per second and miles per hour), and dashed lines temperature (degrees centigrade) in top picture and potential temperature (degrees absolute) in lower picture.

4. The method of drawing frontal contour charts, first used in the analysis of selected European cyclones [9, 10] and later extended to daily weather charts in Canada [18], is very useful for a simple description of actual atmospheric conditions.

cannot show up very clearly in mean cross sections and charts.

Numerous vertical cross sections normal to the average air flow in the free atmosphere have provided detailed information concerning the structure of temperature and wind fields in different synoptic situations. Figure 1 presents a situation with almost zonal flow over North America. The geostrophic wind computed from the height of the standard isobaric surfaces shows the strong concentration in a wind maximum over 80 m sec^{-1} between the 300- and 200-mb levels. The maximum wind (center of the "jet stream") seems to be associated with a typical break in the tropopause as can be seen in Fig. 1. In this special case there seem to be three different tropopauses: a tropical one around 100 mb, a subtropical one between 200 and 250 mb, and a polar one (with multiple structure) over the northern part of the section.

In Fig. 2 one other situation is represented. In this case the vertical cross section is computed along a line from southwest to northeast in Europe. The principal air flow here is from the northwest and the polar front extends from northwest to southeast. In this special case the geostrophic northwest wind shows an even stronger concentration into a jet than in Fig. 1. Isotherms, isentropes, frontal boundaries, and tropopause discontinuities show almost the same characteristics as in the previous figure.

If the polar front zone at the 500-mb surface is used as a reference for coordinates upon which the mean data are computed, it is possible to maintain some of the essential features of the meridional temperature and wind field found in individual situations. Such a computation has been made by Palmén and Newton [43] for the meridian 80°W. The average temperature and wind were correlated with the polar front at the 500-mb surface. The cross section can be regarded as representative of cases with predominant zonal air flow.[5] However, if the cross section is to be used on a hemispherical basis, it must be noted that the average concentration of both meridional temperature gradient and west wind is generally more pronounced in the meridian used for this cross section than along most other meridians.

Characteristic of the temperature distribution in the vicinity of the polar front is the sloping layer of maximum horizontal temperature gradient and pronounced vertical stability, and also the rather strong slope of the isotherms (strong baroclinity) both above and below the frontal surface. This pronounced concentration of the meridional temperature gradient in the polar front itself and in the layers above and below the sloping frontal surface must be associated with a strong concentration of the zonal wind in the upper troposphere and the lower stratosphere as can be seen from Fig. 1.

The equation for the vertical increase of the west wind u (approximately equal to the gradient wind) can be written

$$\frac{\partial u}{\partial z} = -\frac{g}{2\left(\Omega \sin \phi + \frac{u}{a} \tan \phi\right)} \frac{1}{T}\left(\frac{\partial T}{\partial y}\right)_p. \quad (4)$$

Here g is the acceleration of gravity, T the absolute temperature at the level z, and $(\partial T/\partial y)_p$ the meridional temperature gradient measured in an isobaric surface p at the same level. The west wind increases up to the level where the meridional temperature gradient changes its sign (generally at the tropopause level). If u_{max} denotes the zonal wind at that level, one can write

$$u_{max} = u_0 - \frac{g}{2} \int_0^{z_t} \frac{\left(\frac{\partial T}{\partial y}\right)_p}{T\left(\Omega \sin \phi + \frac{u}{a} \tan \phi\right)} dz, \quad (5)$$

where u_0 is the surface wind and z_t the height of the tropopause. Since in most cases the gradient wind at the surface is relatively weak, the wind at the tropopause must be especially strong over the frontal zone and relatively weak outside this zone. Thus any pronounced front with an inclination not too small must be accompanied by a strong upper-level wind concentrated into a relatively narrow band.

The foregoing equations, which are an expression for the assumed balance between gravity, pressure force, and the resultant of Coriolis and centrifugal forces, do not explain the formation of the strong west-wind belt at upper levels, but only establish the fact that the phenomena appear to be parallel. In cases of strong concentration of the meridional temperature gradient in the frontal zone, the wind concentration in the upper troposphere and lower stratosphere is also strong. In such cases it is justifiable to use the name "jet" or "jet stream" for the phenomenon.

The concentration of the pre-existing meridional temperature contrasts into a real frontal zone or layer (frontogenesis) and the concentration of the zonal wind in a jet stream run parallel. Thus any theory for frontogenesis should also be a theory for the formation of an upper jet stream.

The frontal layer in Fig. 2 is separated from the two principal air masses by surfaces of discontinuity for temperature gradient and wind shear, not for temperature and wind as in the classical case treated by Margules [34]. If the angle of inclination of the surface separating the warm air from the air in the frontal layer is denoted by ψ, the change in horizontal wind shear $\partial u/\partial y$ is given [41] by

$$\frac{\partial u}{\partial y} = \frac{\partial u'}{\partial y} - \frac{g \tan \psi}{2\Omega T \sin \phi}\left[\left(\frac{\partial T'}{\partial y}\right)_p - \left(\frac{\partial T}{\partial y}\right)_p\right]. \quad (6)$$

In this equation $\partial u'/\partial y$ and $(\partial T'/\partial y)_p$ represent the wind shear and temperature gradient in the warm air, respectively, and $\partial u/\partial y$ and $(\partial T/\partial y)_p$ represent the same quantities in the frontal zone.

The horizontal wind shear in the frontal zone becomes positive (anticyclonic) if the expression on the right

5. This average cross section is reproduced in Fig. 7 in "Extratropical Cyclones" by J. Bjerknes in this Compendium (see p. 586).

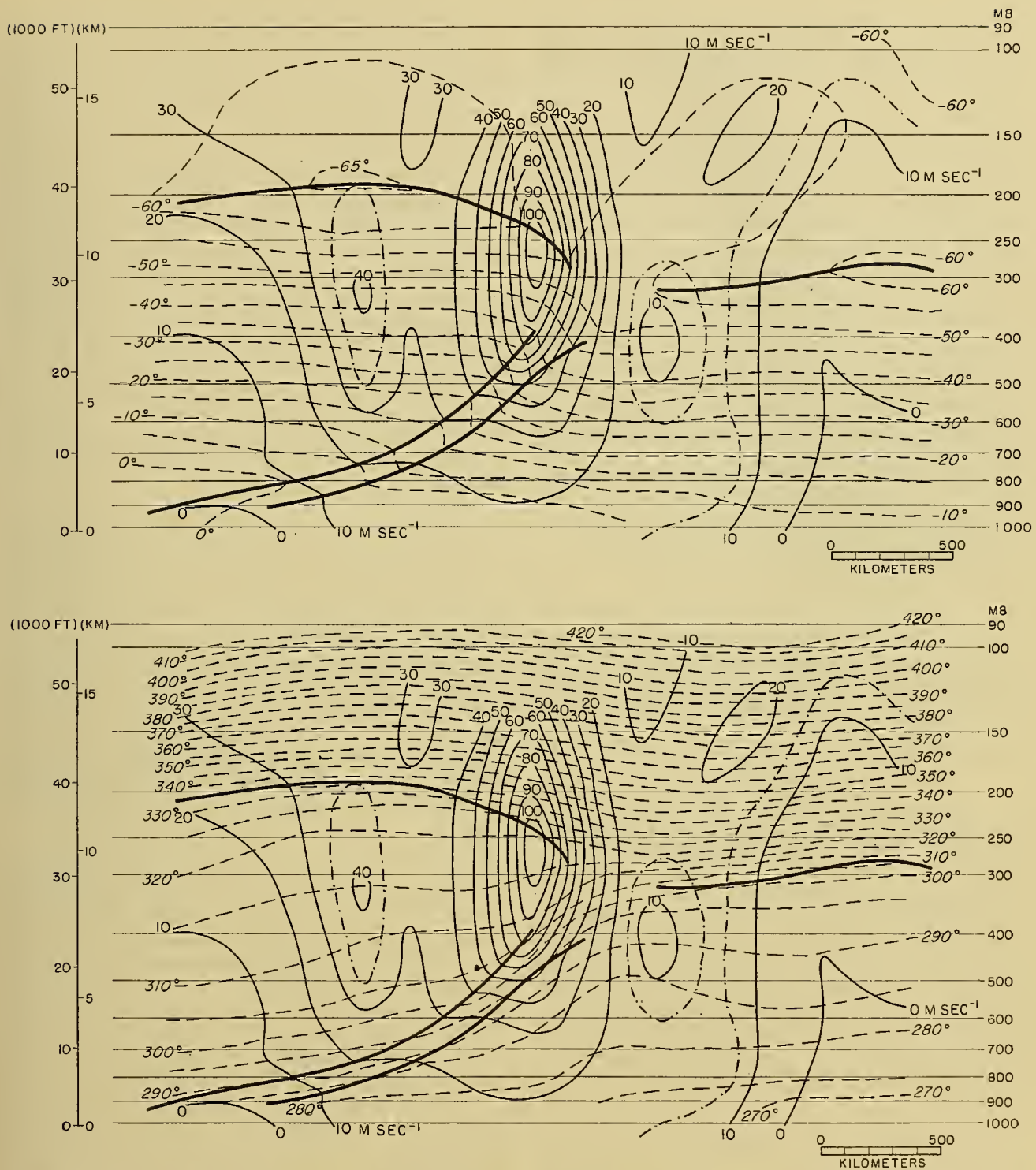


FIG. 2.—Mean cross section through northwesterly flow over Europe. Constructed from all European observations on morning of November 29, 1942 (observations from *Wetterber. Dtsch. Seewarte*). Legend as in Fig. 1 (winds in meters per second). Dash-dotted lines indicate intermediate isovels.

side of equation (6) is greater than zero. Thus, the inequalities

$$\frac{g \tan \psi}{2\Omega T \sin \phi}\left[\left(\frac{\partial T'}{\partial y}\right)_p - \left(\frac{\partial T}{\partial y}\right)_p\right] \gtrless \frac{\partial u'}{\partial y} \begin{matrix}\text{(cyclonic)} \\ \text{(anticyclonic)}\end{matrix} \qquad (7)$$

express the condition for cyclonic or anticyclonic shear in a frontal zone. Hence, the result above can be expressed in the following way: If there is strong anticyclonic shear in the upper warm air, and this anticyclonic shear is greater than the negative value of the second term on the right in (6), the west wind increases toward the north in the frontal zone.

This result concerning the horizontal wind shear in a frontal zone is of considerable importance for the cy-

clone theory. As long as fronts were regarded as surfaces of discontinuity of the order zero (Margules' type) every front was characterized by cyclonic shear. Since real fronts always have a certain width, the principle of cyclonic shear can no longer be sustained in its original form. Analyses of charts for different levels show that although most fronts are characterized by cyclonic shear, fronts with anticyclonic wind shear are not uncommon at the level around 700 mb, whereas practically all fronts in the vicinity of the earth's surface and in the middle and upper troposphere (above 600–500 mb) are characterized by cyclonic shear.

Another characteristic of the westerlies is that there seems to be *an upper limit for the anticyclonic shear of the west wind.* This limit is determined by the rule that the absolute angular momentum about the earth's axis should not increase northward along a horizontal surface. This rule of "dynamic stability," first formulated by Helmholtz [26], has been further developed and used by Solberg [58], Høiland [28], Kleinschmidt [31], Van Mieghem [61, 62], and others, and has been combined with the hydrostatic stability.

The upper limit for the meridional shear in the case of a dynamically stable zonal motion is given by

$$\frac{\partial u}{\partial y} = 2\Omega \sin \phi + \frac{u}{a} \tan \phi. \tag{8}$$

The second term on the right depends upon the horizontal component of the radius of curvature for zonal flow. Since the absolute vorticity ζ_a about a vertical axis is expressed by

$$\zeta_a = 2\Omega \sin \phi + \frac{u}{a} \tan \phi, \tag{9}$$

equation (8) also expresses the rule that the absolute vorticity is negative for a dynamically unstable zonal motion.

In the use of upper-air charts the real wind is commonly identified as the geostrophic wind. In many cases the geostrophic wind gives a sufficiently good approximation to the real wind. However, one must be careful in using the geostrophic approximation in determining the upper limit for the anticyclonic wind shear. If equation (8) is expressed by the geostrophic wind u_g, the following equation for the upper limit of the shear of the geostrophic wind [41] should be used:

$$\frac{\partial u_g}{\partial y} = 2\Omega \sin \phi + \frac{3u_g}{a} \tan \phi. \tag{10}$$

The difference is considerable. If, for example, at latitude 45°N the geostrophic wind at the tropopause level is 80 m sec^{-1} (not an unusually large value), the critical shear for the geostrophic wind is 1.41×10^{-4} sec^{-1}. The corresponding gradient wind (radius of curvature of the trajectory equals $a/\tan \phi$) is only 73 m sec^{-1} and the critical shear for the gradient wind is 1.14×10^{-4} sec^{-1}.

The foregoing results are valid only if the isentropic surfaces are horizontal. In other cases the shear should be determined along a corresponding isentropic surface or, in cases of condensation, along a surface of constant wet-bulb potential temperature. In cases where the slope of these surfaces is not too great the formulas for the critical shear given above can be used; in cases where the slope is greater there should be a corresponding correction. If the shear exceeds the critical value, the air flow is dynamically unstable because in such a case the stabilizing influence of the inertial force vanishes. The importance of this type of instability will not be discussed further here since the problem is treated by J. Bjerknes in his contribution to the Compendium. It might, however, be pointed out that the critical shear represents an interesting analogy to the adiabatic lapse rate of temperature as the upper limit for the vertical temperature gradient in the atmosphere.

The foregoing rule concerning the upper limit for the anticyclonic shear south of the zone of maximum wind should be modified if applied to disturbances, since the curvature of the air flow then becomes of major importance.

Since the existence of a well-marked jet stream in the upper troposphere is associated with the existence of a strong concentration of the meridional solenoid field, a real jet appears on hemispheric upper-tropospheric charts only in regions where the polar front in the middle troposphere is relatively well marked. This means that the jet in actual cases should not be considered as a hemispheric phenomenon without interruptions. A distinct jet stream, as well as a distinct polar front, must be the result of a certain type of disturbance acting frontogenetically. If we use the concept of a hemispheric polar front or hemispheric jet, we do not intend to emphasize that both phenomena should necessarily be fully established around the whole hemisphere.

In mean upper charts and meridional cross sections a relatively well-marked jet stream appears, especially in winter, at latitudes around 20°–40°N according to Namias and Clapp [39]. Also, in mean meridional cross sections studied by Willett [63], Hess [27], and Chaudhury [15], and for the Southern Hemisphere by Loewe and Radok [32], a similarly low latitude for the maximum westerlies was found. Since the average position of the polar front at the 500-mb level appears to be around latitude 45°–50°N, one should expect to find the maximum west wind near this higher latitude. The belt of maximum west wind associated with the upper polar front, however, undergoes such strong meridional displacements and other changes that it could barely be recognized in mean cross sections. Therefore the southern jet on mean cross sections or charts is in some measure another phenomenon probably associated with the northern boundary of the subtropical cell of meridional circulation. Only in cases of very strong southward displacement of the polar-front jet stream are both phenomena combined in one very pronounced belt of maximum west wind. This combination of two upper west-wind maxima appears to be rather common at some longitudes, especially in the vicinity of the meridians 80°W and 120°E.

UPPER LONG WAVES AND CYCLONES

As has been pointed out, the idealized picture of a zonal flow is strongly disturbed in all actual situations. Therefore the real pattern of air flow can be regarded as the superposition of different wind disturbances upon the zonal flow pattern.

There is still no satisfactory theory for the formation of the disturbances of the upper westerlies. However, three general sources for large disturbances must be considered. The first depends upon the topography of the earth's surface, the second is thermodynamical and depends upon the distribution of secondary heat and cold sources connected with the geographical distribution of continents and oceans, and the third is connected with a possible instability of the zonal baroclinic current.

The orographic disturbances are not limited to the regions of mountains. Once formed in such a region, a disturbance has a tendency to influence other regions around the entire hemisphere, as has been pointed out by Rossby [53], Yeh [64], and Charney and Eliassen [14]. The thermodynamical disturbances depend upon the fact that the continents act as secondary cold sources in the winter season and as secondary heat sources in the summer season. The opposite can be said about the oceans. In the cold season the continents are regions of surface outflow or divergence and the oceans are regions of surface inflow or convergence. Correspondingly, there must be upper convergence over continents and upper divergence over oceans. Because of the westerly flow in the middle and upper atmosphere there must be a tendency for the formation of upper troughs over the eastern part of continents, and ridges over the eastern part of oceans. During the summer season the opposite must be true.[6]

The combined effect of the "orographic" and "thermodynamic" disturbances presents a very complicated problem. The location of the hemispheric disturbances dependent on both of these effects is also influenced by the strength of the west wind and the location of the belt of the wind maximum. And because an "orographic" or "thermodynamic" disturbance, once formed, must influence other regions outside its source region, there are a great number of possible combinations for perturbations around the entire hemisphere. Since the influence of these geographical factors can never be eliminated, it is difficult to get any idea of what kind of atmospheric disturbances would form in an atmosphere without any geographical factors influencing the atmospheric processes.

A comparison with the Southern Hemisphere could give some clues for the solution of this problem. However, the upper-air data from the Southern Hemisphere are still so sparse that no satisfactory hemispheric upper-level analyses can be made. Although the influence of mountain barriers and other orographic effects is not as great in the Southern Hemisphere, the continents of Africa, Australia, and especially South America, and in addition the whole Antarctic, have a considerable disturbing influence upon the zonal air flow.

The third type of disturbance, that caused by the supposed instability of the zonal current, has been the subject of a large number of theoretical studies in the last twenty-five years. The instability of "infinitely small" disturbances superimposed upon the westerly current has been the usual starting point in these investigations. A study of these attempts to solve the cyclone problem, however, gives the impression that there still does not exist any theoretical solution fully applicable to the cyclone problem. Therefore the question arises whether it is, in principle, permissible to start from infinitely small perturbations in discussing the cyclone problem. If such an infinitely small perturbation could cause the development of strong cyclones, it would indicate that the atmosphere is extremely unstable. How such an instability associated with storing of useful potential energy could develop in an atmosphere where rather strong perturbations of all kinds are always present seems difficult to understand.

If we consider this, it seems more likely that extratropical cyclones are induced by rather large migrating disturbances which were already in existence. These large disturbances must always be present. Since cyclones and anticyclones are probably cells for transforming potential energy into kinetic energy, the preexistence of a large amount of useful potential energy is necessary for a strong cyclonic development. However, the pre-existing situation must correspond to some kind of *potential instability* that can be released only by the influence of finite perturbations.[7]

The available potential energy of the atmosphere is concentrated primarily in the solenoid field of the polar front. It is also a well-known fact that the polar-front region is the principal birthplace of extratropical cyclones. The question then arises whether a migrating disturbance would induce an irreversible process of the type observed in the development of cyclonic disturbances. The cyclone problem then would become more of a problem of the instability of certain large disturbances always observed on synoptic charts, whereas the ultimate cause of these large perturbations could be neglected. Similar ideas have been expressed earlier by many meteorologists (see, for example, von Ficker [23]).

The 500-mb level has certain advantages for a study of such large disturbances because at that level the upper zonal wind is well developed and the polar-front zone is still not too diffuse.

On circumpolar charts for the 500-mb surface, the belt of the strongest westerlies has the appearance of a "meandering river" situated, on the average, just to the south of, or in the zone of, the strongest horizontal temperature gradient. This zone has already been de-

6. The northeastern part of the North American continent (the Hudson Bay region) acts as a cold source in summer also; in this respect there is a remarkable difference between this region and eastern Siberia.

7. In his contribution to the Compendium, J. Bjerknes discusses in some detail the interaction between migrating upper disturbances and polar-front cyclones.

fined as the intersection between the sloping polar-front layer (surface) and the 500-mb surface. The polar front at the 500-mb level thus shows latitudinal displacements of the same extent as the belt of strongest west wind.

These considerations lead to the assumption that there must be a very close connection between the formation and further development of the long waves and the latitudinal displacement of the tropical and polar air masses in the free atmosphere. The upper troughs are regions for the maximum southward displacement of polar air, and the upper ridges are regions for the maximum northward displacement of tropical air. The existence of a really well-marked polar-front surface in cross sections drawn through formations of this type has been shown by a large number of detailed studies of characteristic situations.

The horizontal dimensions of upper long waves are usually considerably larger than the dimensions of polar-front cyclones on surface maps. Upper long waves, as defined by Rossby [55], correspond more to the individual cyclone families first described by J. Bjerknes and H. Solberg [7] in their original paper on the polar-front theory. The individual members of a cyclone family appear at the 500-mb level as "minor" wavelike perturbations superposed upon the pattern of the long waves. The scheme is presented in Fig. 3, which contains four similar upper long waves at the 500-mb surface, the 500-mb polar front (here considered as a continuous line around the hemisphere), and the four polar fronts at the ground, with their minor cyclonic disturbances. The minor irregularities in the upper polar front and in the contour lines at the 500-mb surface correspond to cyclonic disturbances at the surface. The surface fronts are placed approximately in the same regions as in Fig. 124 in Petterssen's textbook [47, p. 269].

Figure 3 represents a combination of the original polar-front scheme of Bjerknes and Solberg with the scheme for the long waves in the upper westerlies. Both the long waves and the polar fronts on the surface move, on the average, slowly in a west-east direction. The motion of the air at the 500-mb level is faster than the movement of the individual cyclone waves, and the movement of the latter is faster than the eastward displacement of the long waves.

It might be pointed out that the number of four long waves in Fig. 3 is an arbitrary one. Actually the wave number varies from time to time. Also the eastward speed of the waves varies considerably; statistical studies by Cressman [16] indicate that on the average the speed is in fair agreement with Rossby's formula,

$$c = U - \frac{2\Omega a \cos^3 \phi}{n^2}, \qquad (11)$$

for long waves in a barotropic atmosphere (n is the number of hemispheric waves), if U is determined as the mean velocity of the west wind at the level of 600 mb (average level of nondivergence). From equation (11) it can also be concluded that there must be a tendency for a larger wave number n at lower latitudes than at higher latitudes. Thus the scheme in Fig. 3, with four long waves at all latitudes, cannot be maintained; at lower latitudes the waves must show a tendency to be split into a greater number.

The air movement is not strictly horizontal. A complicated vertical movement is superposed upon the horizontal movement expressed approximately by the contour lines in Fig. 3. This vertical movement determines a vertical circulation pattern which is essential for the understanding of the process of development. As a consequence the polar front at the 500-mb surface undergoes continuous changes. These processes will be discussed more in detail in a later section.

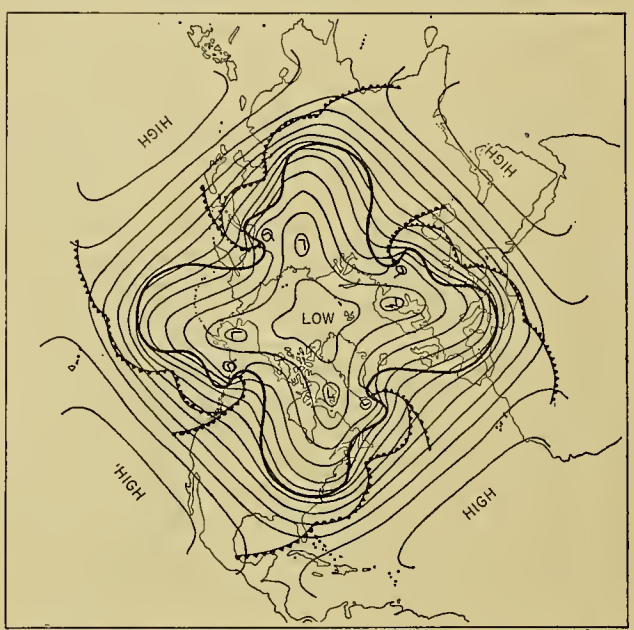


FIG. 3.—Schematic circumpolar chart for the 500-mb level showing four long waves. Heavy line is the polar front at 500 mb and thin lines are contours of 500-mb surface. Fronts at the earth's surface are indicated by the usual symbols.

The changes of the upper polar front and the upper flow pattern can be attributed to some kind of instability of the perturbations in the westerlies. The instability appears, for example, in a rapid increase of the amplitude of the polar-front contour and of the streamlines. At the 500-mb surface, the band of strongest isotherm concentration, which during the coldest season is normally situated in the vicinity of latitude 50°N, may be pushed southward to latitudes 40°–30°N. In the lower troposphere (between about 1000 and 800 mb) this process is connected with a pronounced anticyclonic outflow of polar air masses. The surface polar air thus forms a cold anticyclone, whereas the upper polar air forms an intensified upper trough or, in extreme cases, a closed upper cyclone. The process can be regarded as a combined consequence of increase of cyclonic vorticity due to latitudinal displacement and to convergence in the upper parts of the subsiding cold air.

An opposite process takes place during the formation and strengthening of upper ridges. These ridges are

characterized by cyclonic inflow of tropical air in the surface layers; the upper tropical air forms an intensified upper ridge or, in extreme cases, a closed upper anticyclone, as a result of the latitudinal effect and the divergence in the northward ascending air.

On daily 500-mb charts the formation, intensification, and final degeneration of the upper troughs and ridges can be studied. The disturbances, superficially studied, give the impression of being large waves, and the air flow gives the impression of being a horizontal movement along sinusoidal trajectories. A careful analysis of the real movement of selected air parcels, however, emphasizes the importance of the vertical components of the three-dimensional air flow and the irreversibility of the processes. In the most extreme form the irreversibility of the process appears in cases in which an upper trough develops into a closed "high-level" cyclone to the south of the strongest westerlies, or in which an upper ridge ends up in a closed "high-level" anticyclone to the north of the strongest westerlies. Thus the life history of the upper disturbances shows an irreversibility similar to that which is characteristic of the polar-front cyclones.

Figure 4 presents some of the principal types of upper-air disturbances associated with the deformations of the polar front at the 500-mb level. This figure is a further development of Fig. 3. The polar-front disturbances on the corresponding surface map are indicated by the common symbols for warm, cold, and occluded fronts. Because of technical difficulties in presenting the smaller upper disturbances corresponding to the polar-front waves and cyclones, they have been omitted in drawing the boundary of the polar air and the 500-mb contour lines. The shaded regions indicate the area occupied by polar air at the 500-mb surface.

The scheme presented here consists essentially of four irregular long upper waves as in Fig. 3. The wave troughs are situated on the east coast of Asia, over the Gulf of Alaska, on the east coast of North America, and over eastern and southern Europe. The schematic chart does not intend to present any climatological distribution of the disturbances in question. However, the different types of disturbances are to some extent placed in regions of the Northern Hemisphere where they can frequently be observed on daily synoptic charts. It might, however, be pointed out that not all of the disturbances in Fig. 4 would necessarily appear at the same time over the Northern Hemisphere. The chart corresponds to typical winter situations. For other seasons the intensity of the disturbances is usually weaker, but in principle of the same type.

In addition to the four large upper troughs of different shape, the chart also contains three cold cyclones on the south side of the polar front. These cyclonic disturbances are formed from previous cold troughs of the same type as the four large troughs in Fig. 3. The formation is associated with a gradual "cutting-off" of the polar air from its original source region in the north and will be described later.

The average latitude of the polar front at the 500-mb surface is about 48°N in Fig. 4. The strongest wind at the 500-mb surface, corresponding to the smallest distance between the contour lines, can be observed chiefly in the vicinity of the upper polar front or just to the south of it. Exceptions to this rule, however, can be seen in connection with the strong northward displacement of the polar front from its average position.

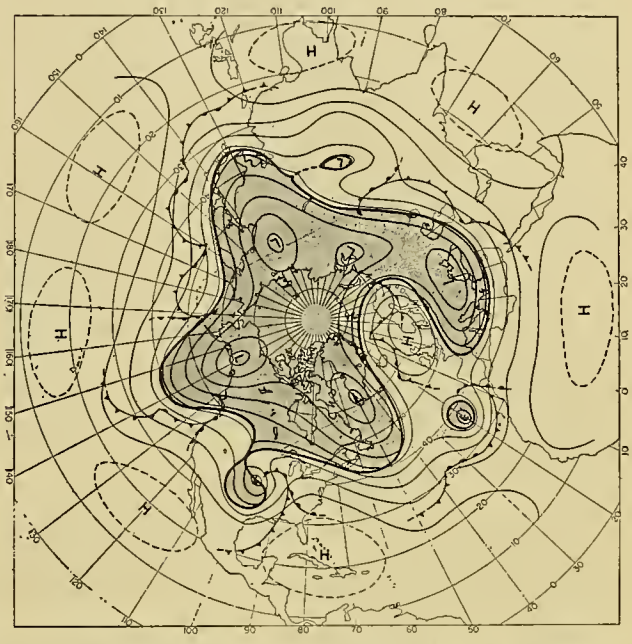


FIG. 4.—Schematic circumpolar chart for the 500-mb level showing several types of disturbances. Heavy line is polar front at 500 mb; thin lines are contours of 500-mb surface. Fronts at earth's surface indicated by usual symbols. Area covered by polar air at 500-mb level is indicated by stippling.

Our schematic figure does not intend to give more than a rather general idea of the development and structure of different types of disturbances. A more detailed description of some of them will be presented in the subsequent sections. There the following special types will be discussed: (1) the cyclone families of the regular type (Fig. 3, or over the Pacific Ocean in Fig. 4); (2) the occluded surface cyclone (represented, for example, by the cyclone over the central parts of the United States); (3) the practically symmetric upper low (east of the Azores); and (4) the warm upper anticyclone (over northwestern Europe) with its series of cyclonic perturbations.

In the schemes presented in Figs. 3 and 4 there is a general difference in scale between the polar-front disturbances connected with individual surface cyclones and those associated with the long upper waves. However there are also cases where one upper long wave is connected with only one large surface cyclone. This type is illustrated in Fig. 4 by the cyclone over the United States.

It seems difficult to maintain the idea of a complete difference in nature between the "long upper waves" and the cyclonic perturbations. Every cyclone naturally influences the upper-air flow and appears as a disturbance on the upper polar front. During the development and deepening of a surface cyclone there is also a

development and intensification of the upper disturbance. However, it is not possible to conclude anything about the causal connection between these phenomena without a detailed analysis.

In a recent paper by Berggren, Bolin, and Rossby [4], the conclusion was drawn that extratropical cyclones comprise a heterogeneous collection of different types of disturbances, including typical frontal waves as well as dynamically quite dissimilar major storms associated with the deepening of planetary wave-troughs in the upper west-wind belt. Our schematic picture is intended to include some of these types.

The surface polar front and the associated perturbations are, in our schematic figures, marked only on the southeastern sides of the large, upper, cold troughs. However, synoptic experience both in North America and in Europe indicates that cyclonic disturbances can often develop also along the southwestern sides of a cold trough. Especially over western Europe during the cold season such disturbances can develop into strong cyclones, though they are usually weaker than the cyclones formed on the southeastern side of a trough. It was not possible to include that type of disturbance in our scheme.

It is not quite clear why there is a preference for the southeastern side for the development of large cyclones. The preference could be associated with the availability of moisture and also with the meridional variation of the Coriolis parameter.

Through the "cutting-off" process associated with the formation of "high-level" cyclones on the south side of the strongest westerlies, a previous large upper trough is gradually eliminated and must ultimately disappear. The number of large upper waves then decreases. Because this process of deformation and ultimate elimination of upper troughs always goes on there must also be an opposite process, namely, the formation of new major troughs. Sometimes this process is a rapid one, an intensifying smaller disturbance developing into a major disturbance. In other cases the process is more gradual, and the new large wave is formed in a region where there is always a general tendency for such a formation, for example, on the east coasts of North America and Asia.

CYCLONE FAMILIES AND THE DEVELOPMENT OF INDIVIDUAL CYCLONES

The general scheme presented in Fig. 3 can be considered the "normal" pattern for the relationship between the surface polar fronts and the long upper waves. Some phenomena associated with this normal pattern for disturbances in the westerlies will be discussed in the following paragraphs. For that purpose a part of the scheme containing two surface fronts with their cyclonic perturbations and the cold anticyclone separating them is reproduced on a larger scale in Fig. 5. In this figure the distance between the two large troughs has been somewhat increased and other small changes have been made.

In Fig. 5 the surface fronts, one over the Atlantic and the other over the Pacific Ocean, with the separating anticyclone over the North American continent, together represent a synoptic situation very common to this part of the Northern Hemisphere. Three frontal perturbations representing different stages of development are marked on both surface fronts. In Fig. 6 the corresponding 500-mb chart with schematic fronts and contour lines is presented. In order to emphasize the connection between both figures, the positions of the surface fronts as well as the contours of the fronts at 700 mb and 500 mb are marked on both the surface chart and the 500-mb chart.

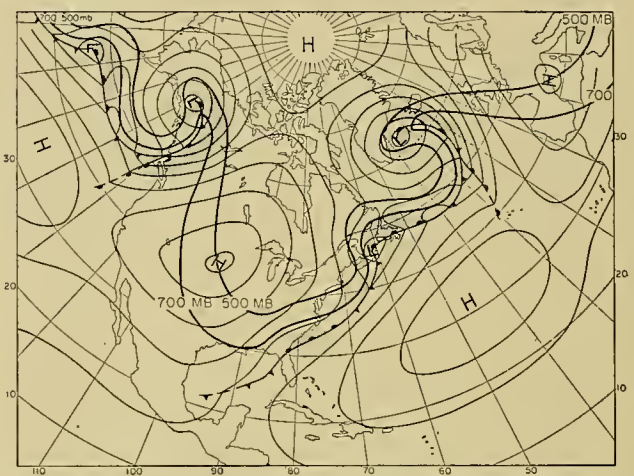


Fig. 5.—Schematic surface map showing cyclone families commonly observed with fairly symmetrical long waves. Heavy lines are 500- and 700-mb contours of polar front, and surface fronts; thin lines, isobars. Precipitation shown by stippled areas.

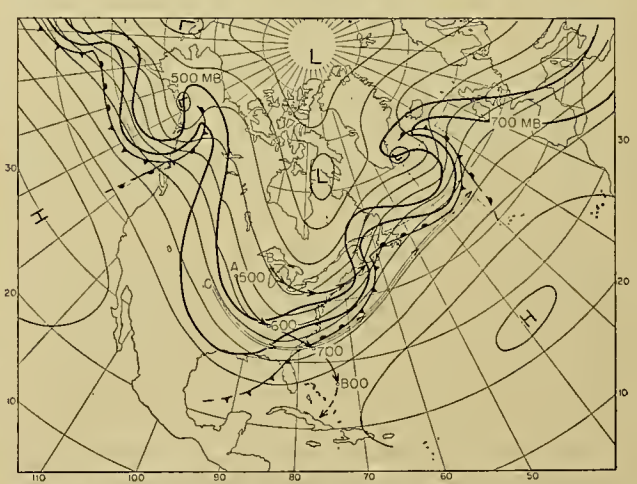


Fig. 6.—Schematic 500-mb map corresponding to Fig. 5. Frontal contours as in Fig. 5; thin lines are contours of 500-mb surface. Arrows starting at *A*, *B*, and *C* indicate approximate 48-hour trajectories of air particles in polar air and in tropical air, respectively. On trajectory *A* are indicated pressure levels reached by particle in frontal layer at approximate 12-hourly intervals.

Analyses of a great number of upper-air charts indicate, as has already been mentioned, that the polar-front zone, interrupted in the surface layers, appears as a rather well-marked zone of transition between the

polar and tropical air higher up in the troposphere. Many times the front is especially well marked only on the southwestern side of the large upper trough. The slope of the front over the western parts of an upper trough is smaller than over the eastern parts. Especially weak is the slope in the region above southwestern parts of the cold surface high, where the cold air appears as a shallow mass. A typical case has been studied in detail by Palmén and Nagler [44].

It might be pointed out that in the region of the surface anticyclone there is often a tendency toward the formation of a new surface front between the colder polar air (continental air or "arctic air") in the northeast and the somewhat warmer polar air (maritime air) in the southwest. This phenomenon is especially common over the United States, where the mountains prevent the cold surface air from penetrating to the western coastal regions. Thus the continental anticyclone is ordinarily not as symmetrical as it has been drawn in Fig. 5; on the contrary, the anticyclone is very often split into two parts separated by a secondary front, making possible the formation of disturbances similar in character, but weaker than those on the polar front. This secondary front has been disregarded in our schematic figures in order to avoid complications.[8]

The general flow pattern at the surface and at 500 mb presented in Figs. 5 and 6 cannot be fully understood if the vertical components of the air motion are not introduced into the picture. In Fig. 6 the upper warm and cold air masses have an almost parallel geostrophic flow. Upon this geostrophic flow is superposed a complicated pattern of nongeostrophic motions. The pattern of vertical movement in upper troughs and associated cyclones has been subjected to a careful study at New York University by Miller [36, 37] and Fleagle [24]. According to their results, there is, on the average, descending motion on the west side of the upper trough and ascending motion on the east side. This vertical motion is accompanied by low-level divergence and upper-level convergence on the west side, and low-level convergence and upper-level divergence on the east side. On the average, the vertical components reach their maximum values at the level of nondivergence (which is not necessarily horizontal). In the divergence field the vorticity of the air parcels moving from west to east undergoes systematic changes; the absolute vorticity has its highest value at the trough line and its lowest value at the ridge line.

This relatively simple pattern for vertical motion and divergence, however, does not describe very well the real movement in a disturbance which necessarily must contain a component of an *irreversible process*, as has been pointed out earlier. Studies of synoptic charts indicate that there is on the average an outflow of cold air from the surface anticyclone and consequently also a general subsidence in the cold air, extending very far up into the upper cold trough. The cold surface anticyclone and the cold part of the upper trough represent regions where cold polar air flows to lower latitudes. In order to maintain continuity the same amount of warmer air must flow into the polar regions. The outflow of cold air on the average must represent a descending current; the inflow of warm air, an ascending current. A vertical circulation therefore is superposed upon the general picture of the horizontal air movement. This circulation is energy producing and serves to maintain the kinetic energy of the westerlies.

In order to obtain a general picture of the three-dimensional movement in synoptic situations of the type discussed here, it would be necessary to follow the three-dimensional air trajectories for air parcels in different parts of our picture. That could to some extent be done by methods developed in the above-mentioned investigation at New York University. However, in order to give results applicable to a general cyclone theory, the method should be combined with a careful three-dimensional frontal analysis and applied to synoptic situations of well-defined types.

Since no such investigations have yet been made, we must restrict the discussion to some rather general qualitative results concerning the three-dimensional trajectories of air parcels in selected parts of our schematic Figs. 5 and 6.

An air parcel initially situated at point A in Fig. 6 is a polar air parcel. It moves with a descending component relative to a warm air parcel at point C of the same 500-mb chart. Both parcels are descending as long as they are on the west side of the trough line. The cold air moves along a path of the type marked in Fig. 6, where the successive positions and pressures are indicated by small circles. The trajectory starts out with a slight cyclonic curvature, corresponding to the cyclonic curvature observed in the upper trough, but the curvature changes gradually to an anticyclonic one when the air parcel reaches levels where $\partial w/\partial z < 0$. At the end of the trajectory the polar air moves anticyclonically as a shallow layer of air which rapidly loses the properties of a cold air mass because of heating from below and adiabatic heating due to subsidence.

The warm air parcel at point C at first descends on the west side of the trough, but gradually overtakes the trough line and then starts ascending. At the time when the cold air flows out in the subtropics as an anticyclonic surface current the warm air parcel has moved very far to the northeast and has eventually left the upper trough in which it started. There is no possibility of giving the exact position of the air, but it follows the general geostrophic flow in the upper troposphere with an average tendency to deviate a little to the left, because of the average southward displacement of the polar air underneath.

If we assume that after thirty-six hours the cold air parcel is at the point marked by 800 mb in Fig. 6, the mean velocity since it started from point A is about

8. This secondary front is usually called the "arctic front" in the United States because the cold Canadian continent is regarded as belonging to the arctic source region in winter. In Europe a similar front often separates the maritime polar air masses from the continental polar air masses over the eastern parts of the continent. Since the source region of the latter air masses in Europe is in the east, not in the north, the corresponding front is not identified with the arctic front.

21 m sec^{-1}. The corresponding vertical velocity is then about -2.8 cm sec^{-1}. This is the order of magnitude of the vertical wind components in the vicinity of the polar front.

Air particles (B in Fig. 6) situated at greater distances from the principal front probably move almost horizontally and thus follow approximately the contour lines on the corresponding chart. Since the air at the 500-mb level usually moves faster than the upper disturbances, a given particle leaves one trough and moves into the next trough. Thus the idea of a quasi-horizontal large-scale wave motion is much more applicable to the air masses far away from the polar front.

The three-dimensional movement outlined here has some resemblance to the helical circulation proposed by Mintz [38]. In our scheme, however, there is no complete left-handed helix, since the transformation of polar air into tropical air, and vice versa, is a much more complicated process than that assumed in the simple model for helical motion.

The movement of a tropical air parcel can be studied in the same way. In this case it is more practical to start from the surface map and discuss the three-dimensional trajectory of a tropical air parcel situated at the beginning in the vicinity of the polar front. A study of the large areas of condensation and precipitation around the principal front shows that the warm air in that vicinity must be subjected to vertical movements of an order of magnitude varying between 1 and 20 cm sec^{-1}. Thus the ascending warm air can rise from the surface layers to the level around 600–500 mb in one day or less. In many cases this rapid ascending motion stops before the air reaches this level, but a careful study of 500-mb charts indicates that moist, almost saturated air, with a potential wet-bulb temperature characteristic of the surface tropical air farther to the south, can be observed in the region above the polar-front surface. The principal area of extended precipitation on the eastern side of the upper trough and in the region of the upper ridge indicates where the principal ascent of the tropical air takes place.

Synoptic experience thus gives some clues concerning the nature of the vertical circulation necessary to maintain the kinetic energy of the atmosphere. The vertical circulation is mainly associated with cyclones and anticyclones. The principal regions of descending polar air are the cold anticyclones separating cyclone families (or individual cyclones) and their counterparts in the free atmosphere, the cold troughs. The tropical air ascends in the regions where condensation and precipitation are observed.

The scheme presented in Figs. 5 and 6 is not a steady-state situation. It represents a certain stage in the development which gradually leads to a degeneration and ultimate elimination of the upper trough. Before we go on to a discussion of that process, however, we must study the development of the typical frontal cyclones.

According to the generally accepted theory for the development of individual cyclones, kinetic energy is produced by a solenoidal circulation connected with sinking of cold air and the ascent of warm air during the occlusion process. The occlusion process in lower layers is associated with convergence and an increase of cyclonic circulation (vorticity). The ascending warm air must form a diverging current in the upper troposphere. The convergence at lower levels and the divergence at upper levels associated with a general ascending movement represent one branch of the solenoidal circulation, the other branch being the combined lower divergence and upper convergence associated with descending movement in the surrounding cold anticyclonic areas.

It is a well-known fact that during the occlusion process the cyclonic circulation gradually spreads to deeper layers of the atmosphere. In a fully developed occluded cyclone, which in the lower troposphere is cold compared with the surrounding air, a strong cyclonic circulation occupies the whole troposphere and the lower stratosphere. However, increasing cyclonic circulation presupposes convergence, according to V. Bjerknes' circulation theorem, and the ascending warm air in an occluding system must be subjected to upper divergence.

This dilemma has been discussed by Brunt [12] in an article on cyclones. In his article, Brunt points out that the upper-level divergence necessary to produce the pressure fall in the central parts of a deepening cyclone must result in increasing anticyclonic circulation in the upper troposphere. The dilemma can be solved, according to Brunt, only if the three-dimensional structure of the cyclone is such that the upper current can remove the air accumulating in the region of lower convergence. This removal is not possible if the extratropical cyclones are symmetrical and, at the same time, the cyclonic circulation is increasing in depth. Here we find one of the principal differences between extratropical and tropical cyclones.[9] In order to find a model which combines the increase in cyclonic circulation with the divergence necessary for removal of the air from the deepening cyclone, we apply J. Bjerknes' tendency equation [6] to a vertical air column situated over the momentary surface center of the cyclone. The pressure change at a level h is then given by

$$\left(\frac{\partial p}{\partial t}\right)_h = -g\int_h^\infty \mathrm{div}_H\,(\rho \mathbf{v})\,dz + (g\rho w)_h. \qquad (12)$$

At the surface, where the vertical component w is zero, the pressure tendency is given by the integrated mass divergence over the whole air column. Since the pressure in a deepening cyclone is falling, the integral of the mass divergence must be positive; since there is low-level convergence during the occlusion process, the upper-level mass divergence must be somewhat stronger than the low-level convergence.

9. The central parts of tropical cyclones are warm compared with the outer parts [51]. The cyclonic circulation therefore decreases upward in tropical cyclones whereas it increases upward in fully developed occluded extratropical cyclones.

Equation (12) can be written

$$\left(\frac{\partial p}{\partial t}\right)_h = -g\int_h^\infty \rho \,\mathrm{div}_H\, \mathbf{v}\, dz - g\int_h^\infty \left(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y}\right) dz + (g\rho w)_h. \tag{13}$$

In the first approximation we can neglect the second term on the right which represents the contribution of advection to the pressure change. A pressure tendency at the surface of -1 mb hr^{-1} corresponds to an average divergence of -0.3×10^{-6} sec^{-1}. Direct measurements from wind observation by Houghton and Austin [29] and further by Sheppard [1] give values for the divergence of the order of magnitude of 10^{-5} sec^{-1} or even more for specific levels. The disagreement between the average divergence in a whole vertical air column determined from pressure changes and the values for the divergence in different levels computed from actual wind observations shows that vertical circulations are essential for an understanding of the atmospheric processes. Therefore we can state that the surface pressure change is the relatively small sum of two large terms with different signs and that the influence of the upper divergence must be somewhat greater than the influence of the lower convergence if $(\partial p/\partial t)_0 < 0$. Furthermore, we can conclude that there must be a certain level of nondivergence approximately coinciding with the level of maximum vertical wind velocity $|w|$.

We can now use the vorticity equation

$$\frac{d\zeta_a}{dt} = -(\zeta + 2\Omega \sin\phi)\,\mathrm{div}_H\, \mathbf{v} \tag{14}$$

in order to investigate the influence of the divergence field on the upper-air flow above the center of the cyclone. The individual absolute vorticity of the air parcels moving above the cyclonic center decreases under the influence of the field of divergence. If we expand the expression in equation (14), we obtain

$$\frac{d\zeta_a}{dt} = \frac{\partial \zeta_a}{\partial t} + v_s\frac{\partial \zeta_a}{\partial s} + w\frac{\partial \zeta_a}{\partial z}, \tag{15}$$

where v_s is the horizontal velocity component along the momentary streamline. The local change $\partial\zeta_a/\partial t$ is, on the average, positive above the surface center of the cyclone, since the upper trough is approaching. The term $w\partial\zeta_a/\partial z$ is small in the upper troposphere where $|w|$ is small. Since the cyclone and the upper trough west of the surface center are moving slowly compared with the upper wind one can put

$$\frac{\partial \zeta_a}{\partial s} \approx -\frac{\zeta_a}{v_s}\,\mathrm{div}_H\, \mathbf{v}. \tag{16}$$

Thus the absolute vorticity must decrease along streamlines of the upper flow.

The absolute vorticity can be written in the form

$$\zeta_a = 2\Omega \sin\phi + \frac{v_s}{r} + \frac{\partial v_s}{\partial r}, \tag{17}$$

where r is the radius of curvature of a streamline. If we neglect the shear, the cyclonic curvature must decrease in the direction of a streamline. If the air flow has a southerly component, the effect of latitude is negative and the decrease of curvature is greater than in cases with a northerly component.

If we could follow the development over an individua cyclone, the decrease of vorticity along upper streamlines should intensify when the upper divergence field intensifies. Since the latitude effect, which is proportional to the meridional wind component v and to the change of the Coriolis parameter with latitude, $2\Omega \cos\phi$, can easily be computed, the shape of the upper contour lines over a deepening cyclone can be used for an estimate of the upper divergence. Scherhag [56] has strongly emphasized the connection between the movement and the deepening of cyclones and the shape of the isobars or contour lines in the upper troposphere.[10]

Figure 7 shows the characteristic pressure distribution in the lower and upper troposphere for a wave cyclone and for a mature cyclone. The increased number of closed isobars in the surface layer is an expression of the increased cyclonic circulation due to low-level convergence. The deformation of the upper isobars indicates the influence of the upper-level divergence field upon the air moving through the system. In the surface layer the convergence field has a much longer time to influence the air flow than is the case in the upper troposphere where the air moves relatively quickly through the system. The warm-sector air, subjected to strong convergence in the lower and inner parts of a cyclone, leaves this region as a current with decreasing vorticity because of the upper-level divergence field. The principal ascent of the warm air masses takes place in the precipitation region marked in Fig. 7; the extension of the area of precipitation to the west along the occluded front is therefore not primarily the result of the ascent of less cold air moving in from the west or southwest.

In the final stage of occlusion the low at the upper level nearly coincides with the surface low, whereas in the beginning the axis of lowest pressure slopes to the west. Single-station analysis gives a phase displacement with height which gradually diminishes during the occlusion process. The importance of this vertical structure of cyclones was originally emphasized by von Ficker [21, 22]. The deepening of a cyclone due to the occlusion process can formally be considered as the result of an interaction between two pressure waves, an upper and a lower one, according to Defant [19]; if the upper wave moves at a higher speed, it reaches the lower wave at the time of maximum depth.

The divergent flow of the upper air above the frontal system has been verified by numerous studies of European cyclones [8, 9, 10, 40]. The characteristic upper-air flow outlined here is also in agreement with well-known

10. We refer to the very extensive list of references concerning this problem in Scherhag's textbook [56].

facts concerning the movement of cirrus clouds as, for example, shown by Pick and Bowering [49] in 1929.

Sawyer [1] has shown in an interesting study that the change of the vorticity field at different levels during the deepening of a cyclone can be used for computation of the vertical movement of the air. Instead of the real vorticity field, which is difficult to determine, Sawyer used the vorticity field computed from the pressure

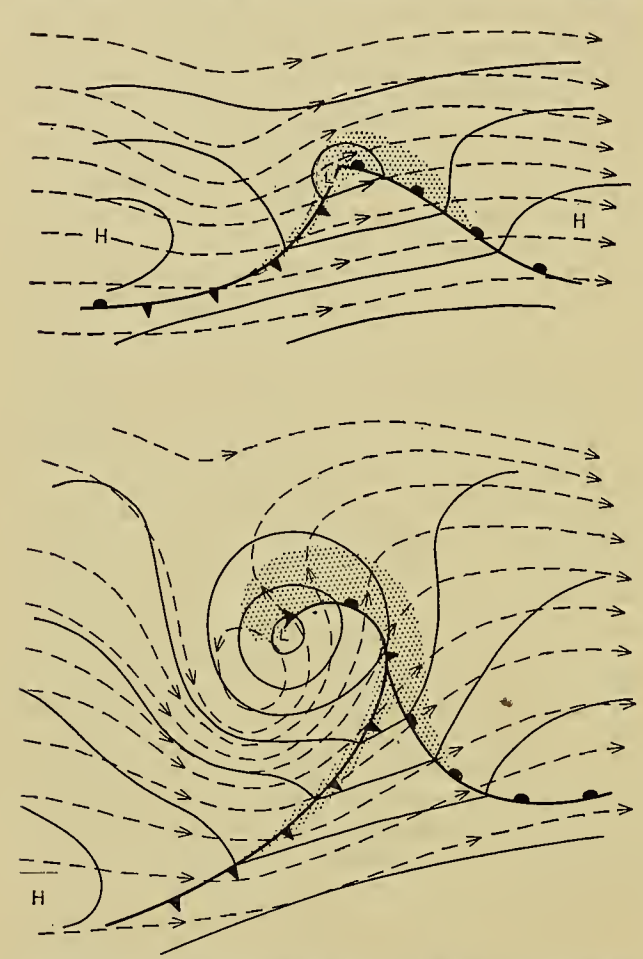


Fig. 7.—Schematic diagrams of young wave cyclone and mature occluded cyclone. Solid thin lines are sea-level isobars, dashed lines are 500-mb streamlines. Precipitation indicated by stippling.

field. Since the pressure field gives an approximate value of the wind field, the change of vorticity can be computed from the pressure changes on consecutive maps. The method is essentially the same as that used here in order to describe the formation of the characteristic upper contour lines in an occluded cyclone.

FORMATION OF HIGH-LEVEL CYCLONES AND ANTICYCLONES

It was pointed out earlier that the cold upper troughs undergo continuous changes and that they have a life history of growth and ultimate degeneration as have the surface cyclones. In some cases the upper troughs seem to be disturbances of larger scale than cyclones; in other cases, however, the dominating upper-air flow is from the west and every individual migrating upper disturbance is associated with a surface cyclone.

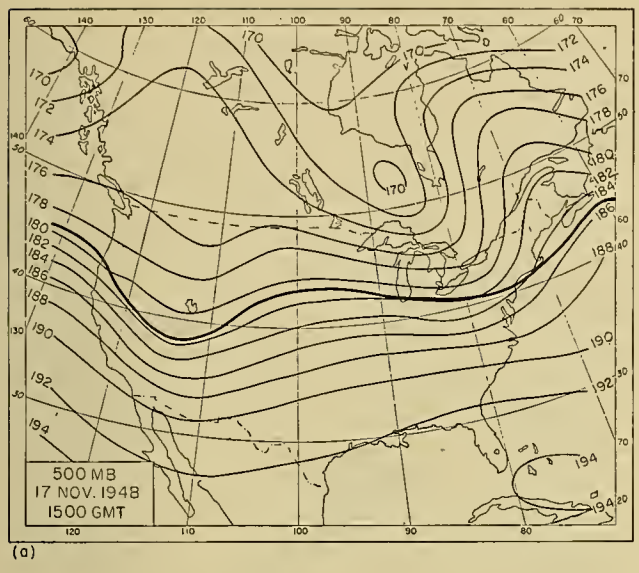
In the following paragraphs some typical cases described in the schematic chart in Fig. 4 will be discussed in more detail. As an example of the development of a strong upper-level low from a pre-existing trough we select the case of November 17–19, 1948 over North America. This situation also illustrates the characteristic structure of an occluded cyclone previously discussed. The sequence of weather maps shows that the structure characteristic of the occluded frontal cyclone can be the result of processes other than the occlusion of wave-shaped frontal perturbations of the type shown in Fig. 7.

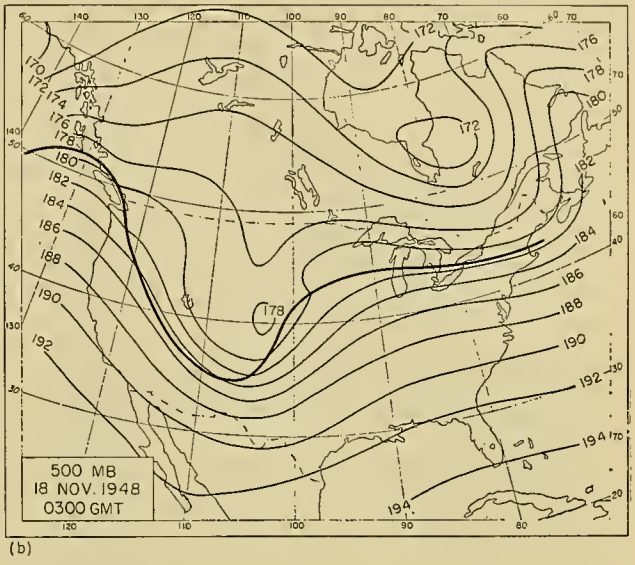
Figure 8 presents five charts for the 500-mb surface. The contour lines and the schematic front show the rapid development of a deep upper cyclone from a relatively weak trough. On the chart for November 17, 1945 the upper flow is predominantly westerly with a cold trough aloft over the western part of the United States. The cold air mass grows gradually out to the south and at the same time a closed upper circulation develops. During the process, which can be followed on the five charts with a time interval of 12 hours, the cold air mass at the 500-mb level is gradually cut off from its polar source region, and on the chart for November 19, 1500 GMT there is left a cold "drop" separated from the cold masses in the north.

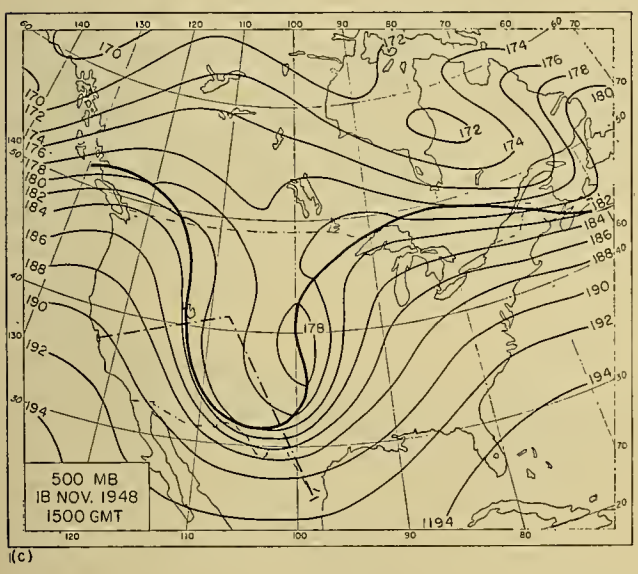
The thermal structure of the air masses can be studied in the vertical cross section (Fig. 9) along the broken line marked on the chart for November 18, 1500 GMT. Especially along the west side of the upper trough, the sloping front is very well marked. The wind component normal to the cross sections shows the characteristic concentration of the maximum wind velocity in a pronounced jet. Figure 9 thus shows that the upper trough consists of a central part filled with cold air masses of polar origin surrounded by warmer air masses. The low tropopause and high stratospheric temperature and the multiple tropopause are very characteristic of all similar situations. The fast-moving upper air in the region of the lower stratosphere is subjected to a strong subsidence on the west side of the trough and to a strong ascent on the east side (see, for example, [24]).

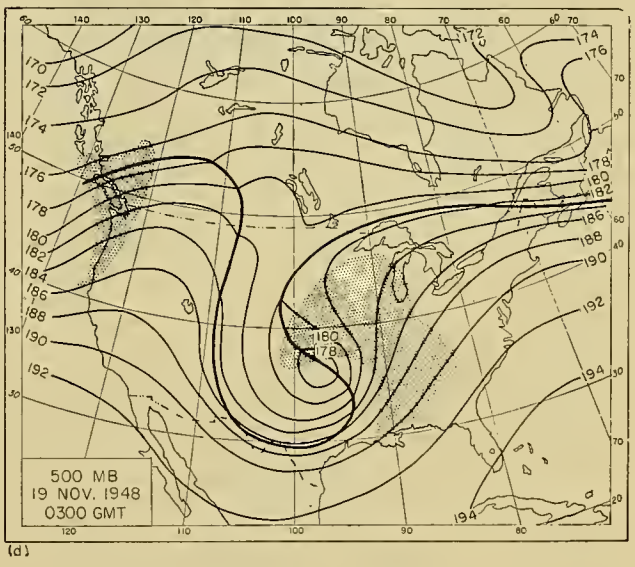
In Fig. 10 the contours of the principal front at the surface and on the 850-, 700-, and 500-mb surfaces are presented for November 18 and 19, 0300 and 1500 GMT. At the beginning of the period there are two surface fronts, one western front connected with the developing disturbance and one eastern front corresponding to the northern boundary of the moist warm Gulf air. On the chart for November 18, 1500 GMT the surface fronts have already been brought so near each other that a separation is impossible. At that time a structure corresponding to the classical picture of an occluded cyclone has already started to develop. On the next day the structure of a regularly occluded cyclone is complete, as can be seen from the surface map in Fig. 11.

In Fig. 8*d* the principal precipitation areas are marked. From the figure it can be seen that the areas of









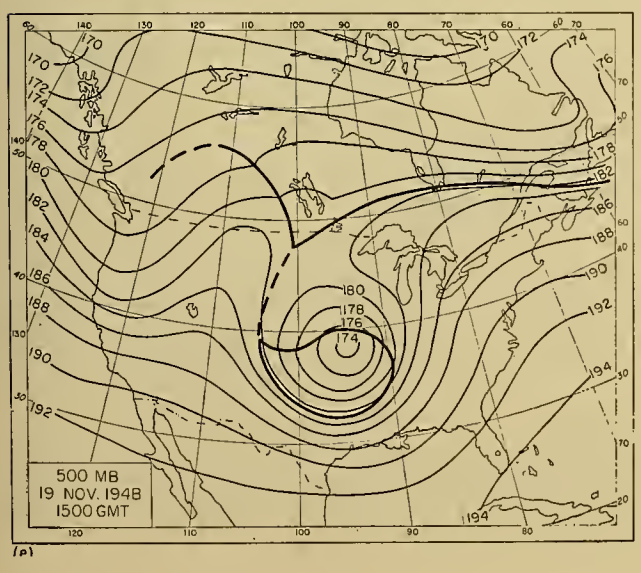


FIG. 8.—Idealized 500-mb charts at 12-hourly intervals from 1500 GMT November 17, 1948 to 1500 GMT November 19, 1948. Heavy lines, 500-mb fronts; thin lines, contours of 500-mb surface (labelled in hundreds of feet). Precipitation areas indicated by stippling in Fig. 8*d*.

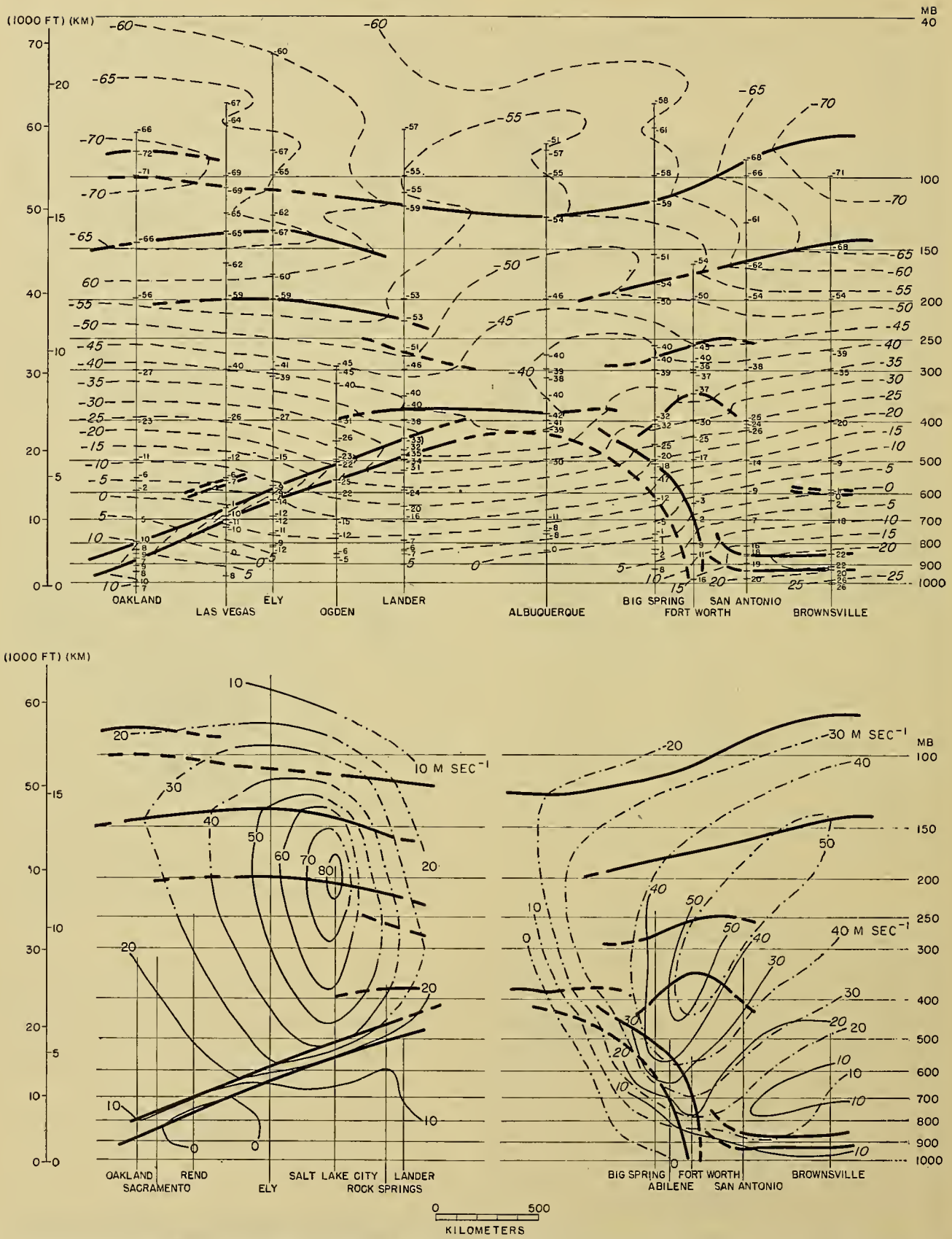


FIG. 9.—Vertical cross section through cold dome at 1500 GMT November 18, 1948, along dash-dotted line in Fig. 8c. In top figure, heavy lines are frontal boundaries and tropopauses, thin lines isotherms (degrees centigrade). Lower figures show isovels of actual wind velocity (solid thin lines, meters per second) supplemented by computed gradient wind velocities (dash-dotted lines).

precipitation approximately coincide with the areas where $\partial\zeta_a/\partial s < 0$. These areas can, according to equation (16), be found where the ascending warm air is diverging. In Fig. 10*c* the arrows indicate the flow of warm air along the frontal surface in some parts of the disturbance. Since the frontal contours also move, one cannot immediately determine the vertical component of the air flow from the map; however, the relatively slow movement of the system and the large component of the warm air flow normal to the frontal contours in the inner parts of the cyclone indicate that here the ascending component w must be large.

The charts for November 19, 1500 GMT present the fully developed upper cyclone which at that time very nearly coincides with the surface center of the "occluded" cyclone. The charts also present a situation in which the cold air at the 500-mb surface is completely cut off from the cold air in the north. This cutting-off process starts in the upper troposphere and gradually penetrates downward. The process of cutting off is not completely finished at the level of the 700-mb surface, as can be seen from Fig. 10.

The case history of November 17–19 represents a common development in the United States, as was pointed out in the previous discussion of the schematic chart in Fig. 4. The fully developed surface cyclone of November 19 has all the characteristics of an occluded polar-front cyclone in spite of the fact that it did not develop from a wave cyclone. The three-dimensional fields of temperature, pressure, and wind presented in Figs. 8–10 are, however, characteristic of every "occluded" cyclone whether it has passed through a regular process of occlusion or not. Essential for the whole development of a mature cyclone is the formation of the

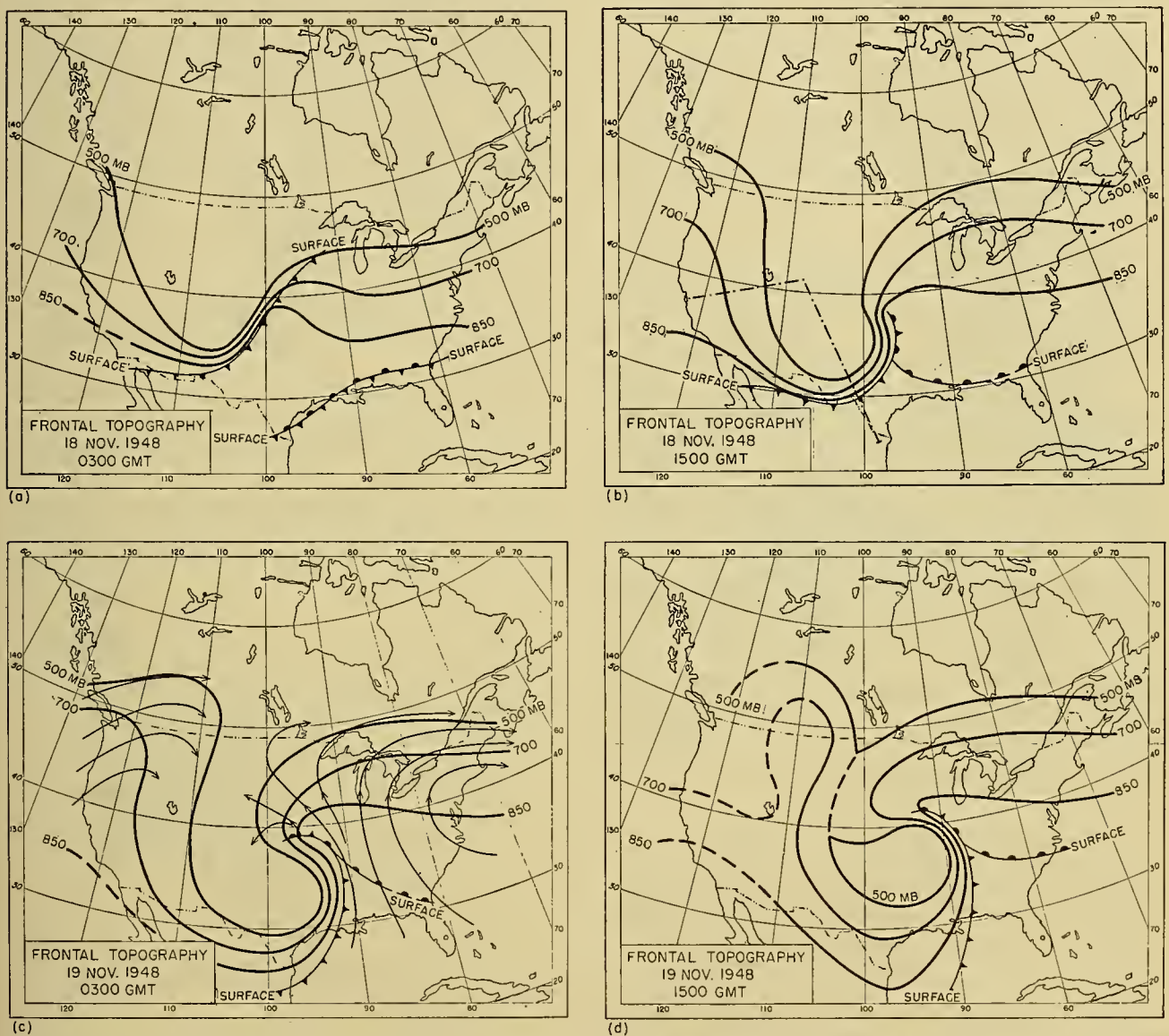


FIG. 10.—Idealized frontal contour charts, November 18–19, 1948, corresponding to charts of Fig. 8*b–e*. On Fig. 10*c*, corresponding to Fig. 8*d*, thin arrows indicate instantaneous streamlines at warm-air boundary of frontal surfaces. On Fig. 10*d* frontal contours are dashed where front is indistinct.

upper disturbance associated with the deformation of the upper front which can be followed on the charts in Fig. 8. This deformation of the upper front is associated with the formation of an upper cyclone or a very deep trough. During the process a large part of the cold air is separated from its original source region and flows as a diverging lower current very far to the south. This diverging lower cold current can be seen on the surface map for November 19 (Fig. 11) in the regions west of the surface cold front.

There is no doubt that many "occluded" cyclones on surface maps have never gone through a real process of occlusion although they show the same characteristic structure as really occluded polar-front perturbations. Obviously a well-marked surface front is not so essential for the development as was generally assumed formerly.

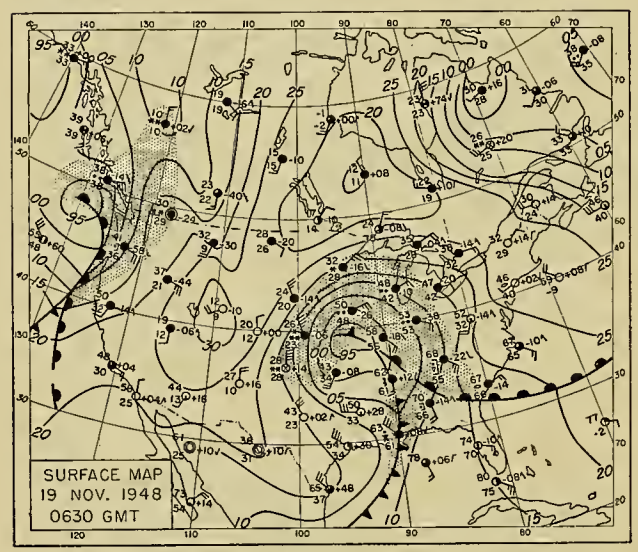


Fig. 11.—Surface map, 0630 GMT November 19, 1948, three hours later than charts of Figs. 8*d* and 10*c*.

One more point should be emphasized. By comparison of the consecutive 500-mb charts in Fig. 8 it can be seen that the area of polar air at that level decreases during the process of cyclogenesis and the development of the deep "occluded" surface cyclone. *The process of seclusion of the polar air at the 500-mb level thus corresponds to the occlusion process in lower layers.* In the upper atmosphere the warm air gains area, in the lower atmosphere the cold air gains area. This process corresponds to the scheme for release of energy of storms proposed by Margules [33] in his classical studies.[11]

As a result of the process described here, the upper cold air (for example, at the level of 500 mb) has been isolated from the main body of cold air in the north. Since the boundary layer separating the warmer air masses in the south from the polar air masses in the north also marks the zone of strongest west wind, the whole process of forming upper cyclones is associated with a meandering of the "jet stream." This meandering very often goes so far that it results in the formation of an almost symmetrical upper cold low at very low latitudes, whereas the strong west wind becomes re-established to the north of the newly formed upper cyclone. Figure 12 shows, according to Hsieh [30], the

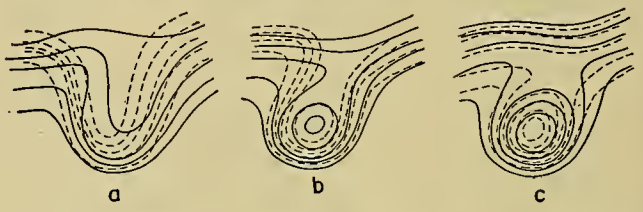


Fig. 12.—Schematic diagrams showing various stages in formation of cut-off cyclone. Solid lines are contours of 500-mb surface, dashed lines are isotherms. (*After Hsieh* [30].)

characteristic change of the pattern of isotherms and air flow in the middle troposphere during the development of a closed cold cyclonic vortex. This formation of "drops of cold air" (*Kaltlufttropfen*) has been well known since aerological observations have been extended over large areas.[12] Figure 13 shows a schematic picture of the frontal contours in such a cut-off low.

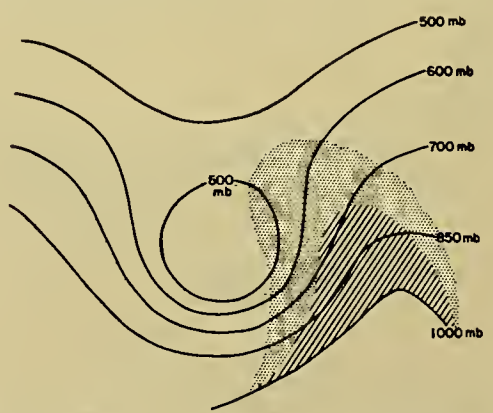


Fig. 13.—Schematic diagram showing frontal contours in cut-off low. Hatching indicates area of heavier precipitation, stippling area of lighter precipitation. (*After Hsieh* [30].)

A beautiful example of this type of high-level cyclone is the case of November 1–7, 1946 over the southwestern part of the United States. This case has been studied in some detail by Crocker [17] and Palmén [42]. Figures 14, 15, and 16 are reproduced from the latter study. Figure 16, in particular, represents an almost ideal picture of a cold symmetrical upper cyclone with its warm, low stratosphere. It should perhaps be mentioned that the tropopause marked as a continuous surface in Figs.

11. The energetics of a similar process of cyclogenesis has recently been studied by Phillips [48]. The main difficulty in applying Margules' ideas on the development of real cyclones depends upon the well-known fact, already emphasized by Schröder [57], that cyclones cannot be regarded as closed systems. Because of this and other difficulties many meteorologists are critical of Margules' theory (see for example [59]). It seems to the author, however, that the theory is essentially sound if applied correctly and with consideration for all complications.

12. See Scherhag's textbook on weather analysis and forecasting [56] for a detailed discussion of the structure and formation of such drops of cold air in Europe. Among other references Nyberg [40], and Raethjen [50] might be mentioned.

15 and 16 could probably be split into several surfaces, as in Fig. 9, by a more detailed analysis.

Raethjen [50] has recently tried to explain the formation of the high-level cyclones associated with drops

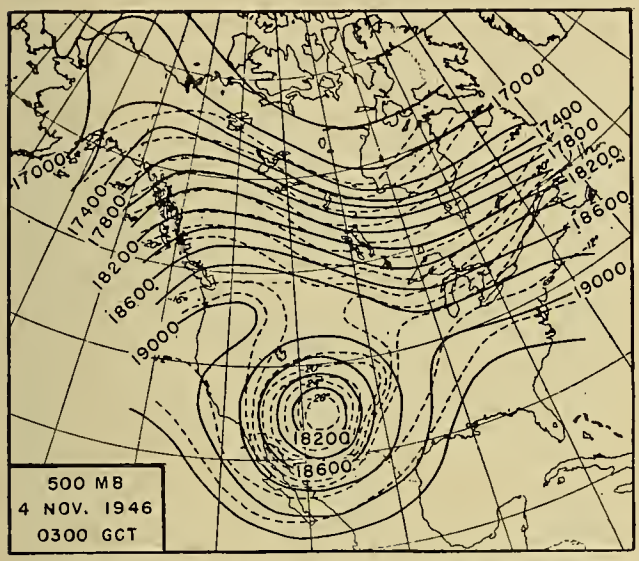


FIG. 14.—500-mb chart over North America, 0300 GMT, November 4, 1946. Solid lines are contours of 500-mb surface, dashed lines are isotherms (degrees centigrade).

of cold air as a result of isentropic exchange of air parcels conserving their absolute angular momentum. This process would obviously result in a change in circulation (increasing cyclonic vorticity in the central upper parts of the cold drop) similar to the vertical circulation outlined here. Quite different points of view have been expressed by Rossby [54] in his paper on the meridional movement of sinking cold domes. It seems obvious that the problem of the extreme meandering of the west-wind belt resulting in the formation of cold upper lows must be subjected to further investigations before the nature of the process can be regarded as completely clarified.

The meandering of the belt of upper westerlies and the deformation of the upper polar front also result in an increase of the amplitude of the warm upper ridges. In extreme cases a warm ridge can be completely separated from its warm source region in the south. The meandering then results in the formation of closed warm upper anticyclones and the belt of strong westerlies can be re-established south of it. These types of warm upper anticyclones consist mostly of cold polar air masses in the lower layers and therefore appear, in a superficial analysis, as cold polar anticyclones. The true nature of these anticyclones has been elucidated through a number of synoptic investigations. The structure of such a warm upper anticyclone has been subjected to a detailed analysis by Berggren, Bolin, and Rossby [4]. The case they studied has been used as a model for the upper anticyclone over northern Europe shown in Fig. 4. The pronounced anticyclonic vorticity in these warm upper highs can partly be explained as a result of advective transport of low absolute vorticity

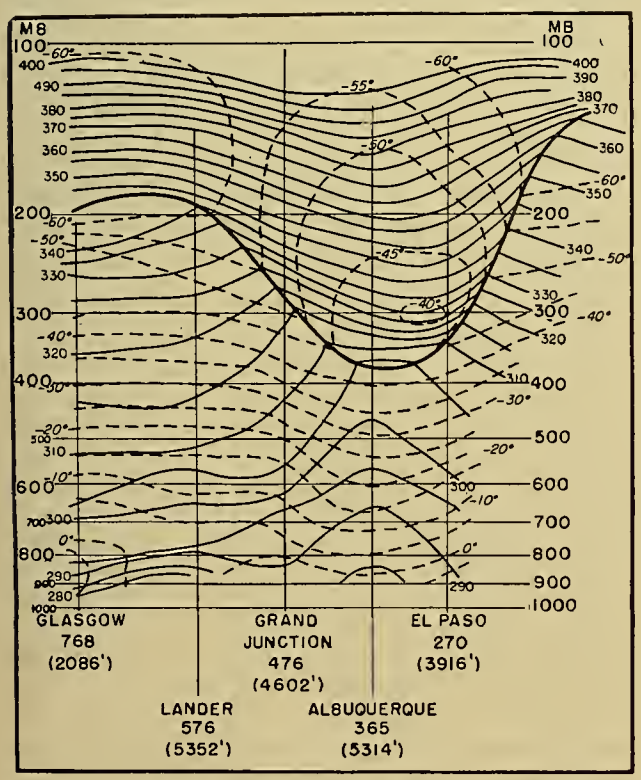


FIG. 15.—Vertical north-south cross section through center of high-level cyclone of Fig. 14. Heavy line indicates tropopause, solid thin lines isentropes (degrees absolute) and dashed lines isotherms (degrees centigrade).

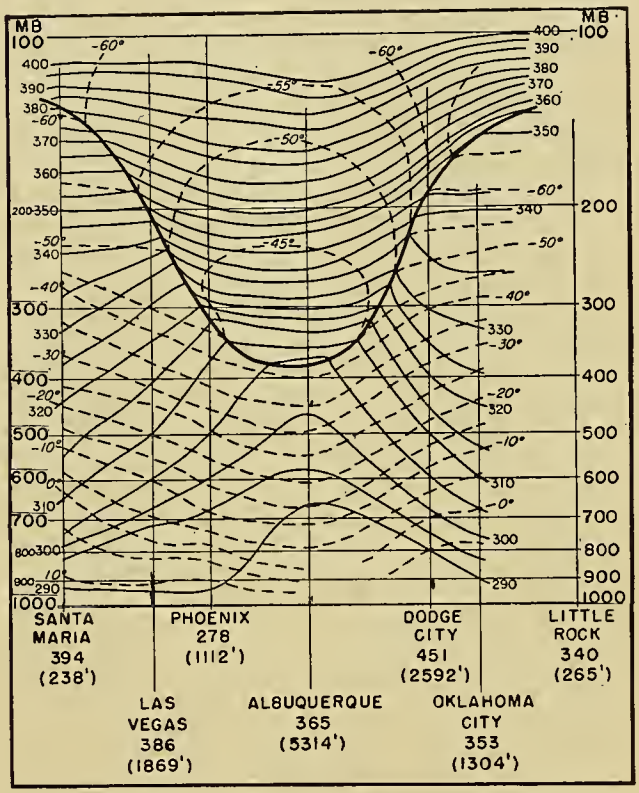


FIG. 16.—Vertical west-east cross section through center of high-level cyclone of Fig. 14. Legend as in Fig. 15.

from the south. However, the influence of the upper divergence field on the development of the relative anticyclonic vorticity should also be considered.

Both the cold upper lows and the warm upper highs can be considered a further development of the upper trough and ridge described here earlier. The cold air in the upper cyclones represents the uppermost part of a sinking cold dome. The warm high-level anticyclones consist essentially of a body of lifted tropical air. These processes are closely related to the processes connected with the formation of cyclones of different types.

SOME GENERAL REMARKS ON THE CYCLONE PROBLEM

There is no question that our knowledge of the structure and life history of extratropical cyclones has greatly improved as a result of the increased network of aerological stations which has been established during the last 10–15 years and especially as a result of the general introduction of radiosonde observations into the daily weather service. The extension of upper-air observations on a hemispheric scale has undoubtedly shown that there exists, not only in the surface layers but also in the middle troposphere, a more or less distinct boundary layer separating air masses of polar origin from those of tropical or subtropical origin. This boundary has the character of a true front only over limited regions; in other regions it appears as a belt of strong meridional solenoid field. Careful analyses of different types of synoptic situations also indicate that this boundary layer of the free atmosphere can be considered a direct continuation of the surface polar fronts. The polar front thus appears as a sloping boundary layer which extends from the surface to the upper part of the troposphere. This is in agreement with the ideas introduced for the first time in clear form by the Norwegian school of meteorologists.

Although the existence of a more or less distinct polar-front surface is thus an empirical fact, the explanation of the formation of this boundary layer is not quite clear. It is obvious that the formation and maintenance of a polar front must be connected with the general atmospheric circulation. However, there still is no complete theory of frontogenesis which considers the entire dynamics of the general circulation. On the other hand, it is undeniable that extratropical cyclones frequently and predominantly form at the polar front although there are other birthplaces for cyclonic depressions. The question of the role of extratropical cyclones in the complex problem of the general circulation is therefore of primary importance not only in the cyclone theory but also for a deeper understanding of the problem of the general atmospheric circulation.

In an earlier section it was pointed out that the polar-front zone in the middle latitudes is not only the zone of strongest solenoid concentration but also the zone of strongest concentration of kinetic energy. Thus the frontogenetic processes operating in the atmosphere on a hemispheric scale form not only the fronts or frontal layers but also the characteristic, strong concentration of the west-wind belt into an upper wind maximum, the jet stream. Frontogenesis hence means at the same time an increase of both solenoidal and kinetic energy in the polar-front zone.

On the other hand, the vertical circulation characteristic of cyclone development also seems to be a "direct" one resulting in an increase of kinetic energy. According to Starr [60], a system is able to produce horizontal kinetic energy if horizontal divergence is associated with relatively high pressure and convergence with relatively low pressure. In the previous discussion of the life history of a cyclone, it was pointed out that the development of a cyclone is characterized by low-level convergence and upper-level divergence. But if we consider not only the cyclone but the total region of an atmospheric disturbance—cyclone + anticyclone—the horizontal divergence (in the lower parts of the anticyclone and the upper parts of the cyclone) actually occurs with higher pressure than does the convergence (in the lower parts of a surface cyclone and the upper parts of a surface anticyclone). This circulation therefore means a transformation of pre-existing potential energy into kinetic energy.

The combined process of frontogenesis and cyclogenesis can probably be described in the following manner. During the time of *frontogenesis* the "available" potential energy[13] of the atmosphere is gradually increased by nonadiabatic processes, primarily radiation, and through a process of concentration of the horizontal temperature contrasts into a relatively narrow zone. During this process, however, the kinetic energy also increases (the surface wind speed remaining relatively constant). The stage of frontogenesis is therefore also a stage with increase of zonal index, especially at upper levels. During the time of *cyclogenesis* the strong zonal flow has a tendency to break down; "available" potential energy is now transformed into kinetic energy. The mechanism of the breakdown of the zonal current into a more irregular form of movement could be attributed to some kind of instability of the zonal movement which is not known in detail.[14] The fundamental cyclone problem would be solved in principle if one could find the causes and mechanism of this instability.

Any theoretical solution of the cyclone problem is still very far from a solution of the cyclone problem presented by nature. Meteorologists are still in disagreement about many fundamental aspects of the cyclone problem. While some meteorologists assume that the kinetic energy of extratropical disturbances represents only a concentration of the pre-existing kinetic energy of the west-wind belt, others regard the disturbances as active cells in the production of kinetic energy. The vertical circulation associated with the growth of the disturbances in the westerlies and particularly with the occlusion process indicates, as already pointed out, that the latter viewpoint is in better agreement with

13. The "available" potential energy is meant to include only the energy available from a redistribution of the mass of the system such that the center of gravity is lowered.

14. Compare, for example, Charney [13], Eady [20], and Godske [25].

our experience. On the other hand, there is probably no doubt that the period of cyclogenesis must be preceded by a building-up of the kinetic energy of the upper west-wind belt and that thus certain types of disturbances also appear as concentrations of already existing kinetic energy.

The principle of dynamic instability may perhaps result in new efforts to solve the problem of the instability of the west-wind belt. A considerable amount of literature on that problem already exists [62]. The aerological analyses of weather situations with distinct fronts has revealed the existence in the atmosphere of zones with a width of 100–500 km (or sometimes even more) in which the criterion of dynamic or inertial instability is fulfilled. The fact that these zones actually exist suggests the need for further investigation of this type of instability. Another approach to a solution of the problem of instability has been proposed by Rossby [54] in his discussion of the meridional displacement of cold air masses.

If the complexity of the cyclone problem is considered, it does not seem likely that any satisfactory theoretical solution can be achieved in the near future. One is also forced to this conclusion by the fact that "extratropical cyclones" obviously do not represent any well-defined single atmospheric phenomenon but more likely a whole group of phenomena. Therefore perhaps no single explanation is sufficient.

The complexity of the atmospheric phenomena associated with extratropical cyclones does not mean, however, that no general features characterize disturbances in the westerlies. Some of the most characteristic ones have been presented in this article. Further detailed synoptic investigations of selected types of disturbances will certainly improve our knowledge and gradually result in a better understanding of the dynamics of the atmosphere. In the opinion of the writer, a detailed and careful synoptic investigation of typical weather situations on a very large scale—approaching hemispheric dimensions—is the only method which can furnish us with the facts which are necessary both for a better understanding and for further theoretical study.

One of the greatest difficulties in synoptic meteorology is that the concept of causality is extremely difficult to explore. To give some examples: Horizontal acceleration appears as the small difference between two large terms, acceleration due to the horizontal pressure field and acceleration due to the earth's rotation. Similarly, the vertical acceleration is the small difference between the opposing accelerations due to the vertical pressure field and to gravity. Again, all pressure changes are small differences between two or three large terms, as can be seen from the tendency equation. Finally, the thermal wind equation represents a balanced condition between a circulation due to the vertical solenoid field and one due to the earth's rotation. Because of inevitable errors in all meteorological observations it is extremely difficult to get a sufficiently exact determination of quantities needed for computing the nonbalanced conditions in an atmospheric situation. Through comparison of successive situations some clues concerning the nonbalanced parts of the movement can be achieved. The time difference between consecutive upper-air observations, especially between radiosonde observations (now twelve hours in most regions), is too large to permit a satisfactory determination of the time derivatives. Special synoptic investigations with several stations operating on time intervals of 2–3 hours would therefore be extremely valuable.

REFERENCES

1. Bannan, J. K., Graham, R. C., Sawyer, J. S., and Sheppard, P. A., "Large Scale Vertical Motion in the Atmosphere. Recent Research at Central Forecasting Office, Dunstable." *Quart. J. R. meteor. Soc.*, 75: 185–188 (1949).
2. Bergeron, T., "Über die dreidimensional verknüpfende Wetteranalyse." *Geofys. Publ.*, Vol. 5, No. 6 (1928).
3. —— "Richtlinien einer dynamischen Klimatologie." *Meteor. Z.*, 47: 246–262 (1930).
4. Berggren, R., Bolin, B., and Rossby, C.-G., "An Aerological Study of Zonal Motion, Its Perturbations and Break-Down." *Tellus*, Vol. 1, No. 2, pp. 14–37 (1949).
5. Bjerknes, J., "Exploration de quelques perturbations atmosphériques à l'aide de sondages rapprochés dans le temps." *Geofys. Publ.*, Vol. 9, No. 9 (1932).
6. —— "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54: 462–466 (1937).
7. —— and Solberg, H., "Life Cycle of Cyclones and Polar Front Theory of Atmospheric Circulation." *Geofys. Publ.*, Vol. 3, No. 1 (1922).
8. Bjerknes, J., and Palmén, E., "Investigations of Selected European Cyclones by Means of Serial Ascents. Case 4: February 15–17, 1935." *Geofys. Publ.*, Vol. 12, No. 2 (1937).
9. —— "Aerologische Analyse einer Warmfrontfläche." *Beitr. Phys. frei. Atmos.*, 25: 115–129 (1939).
10. Bjerknes, J., u. a., "Synoptisch-aerologische Untersuchung der Wetterlage während der Internationalen Tage vom 13 bis 18 Dezember 1937." *Veröff. geophys. Inst. Univ. Lpz.*, (2) Bd. 12, Nr. 1 (1939).
11. Bjerknes, V., u. a., *Physikalische Hydrodynamik.* Berlin, J. Springer, 1933.
12. Brunt, D., "Some Problems of Modern Meteorology: I. The Present Position of Theories of the Origin of Cyclonic Depressions." *Quart. J. R. meteor. Soc.*, 56: 345–350 (1930).
13. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4: 135–162 (1947).
14. —— and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
15. Chaudhury, A. M., *Some Dynamic and Aerological Aspects of Meteorology over Indo-Pakistan.* Ph.D. Thesis, 65 pp., Dept. Meteor., Univ. of Chicago, 1949.
16. Cressman, G. P., "On the Forecasting of Long Waves in the Upper Westerlies." *J. Meteor.*, 5: 44–57 (1948).
17. Crocker, A. M., "Synoptic Applications of the Frontal Contour Chart: The Motion of Selected Lows, 5–7 November 1946." *Quart. J. R. meteor. Soc.*, 75: 57–70 (1949).
18. —— Godson, W. L., and Penner, C. M., "Frontal Contour Charts." *J. Meteor.*, 4: 95–99 (1947).
19. Defant, A., "Primäre und sekundäre, freie und erzwungene Druckwellen in der Atmosphäre." *S. B. Akad. Wiss. Wien*, Abt. IIa, 135: 357–377 (1926).
20. Eady, E. T., "Long Waves and Cyclone Waves." *Tellus*, Vol. 1, No. 3, pp. 33–52 (1949).
21. Ficker, H. v., "Beziehungen zwischen Änderungen des

Luftdruckes und der Temperatur in den unteren Schichten der Troposphäre (Zusammensetzung der Depressionen)." *S. B. Akad. Wiss. Wien*, Abt. IIa, 129: 763–810 (1920).

22. —— "Die Änderung des Wetters in den verschiedenen Entwicklungsstadien einer Depression." *S. B. Akad. Wiss. Wien*, Abt. IIa, 131: 383–415 (1922).
23. —— "Der Sturm in Norddeutschland am 4 Juli 1928." *S. B. berl. math. Ges.*, 22: 290–326 (1929).
24. Fleagle, R. G., "The Fields of Temperature, Pressure, and Three-Dimensional Motion in Selected Weather Situations." *J. Meteor.*, 4: 165–185 (1947).
25. Godske, C. L., "Zur Theorie der Bildung aussertropischer Zyklonen." *Meteor. Z.*, 53: 445–449 (1936).
26. Helmholtz, H. v., "Über atmosphärische Bewegungen." *Meteor. Z.*, 5: 329–340 (1888).
27. Hess, S. L., "Some New Mean Meridional Cross Sections through the Atmosphere." *J. Meteor.*, 5: 293–300 (1948).
28. Høiland, E., "On the Stability of the Circular Vortex." *Avh. norske VidenskAkad.*, Vol. 1, No. 1 (1941).
29. Houghton, H. G., and Austin, J. M., "A Study of Non-Geostrophic Flow with Applications to the Mechanism of Pressure Changes." *J. Meteor.*, 3: 57–77 (1946).
30. Hsieh Yi-ping, "An Investigation of a Selected Cold Vortex over North America." *J. Meteor.*, 6: 401–410 (1949).
31. Kleinschmidt, E., "Stabilitätstheorie des geostrophischen Windfeldes." *Ann. Hydrogr., Berl.*, 69: 305–325 (1941).
32. Loewe, F., and Radok, U., "A Meridional Aerological Cross Section in the Southwest Pacific." *J. Meteor.*, 7: 58–65 (1950).
33. Margules, M., "Über die Energie der Stürme." *Jb. Zent-Anst. Meteor. Wien* (1903).
34. —— "Über Temperaturschichtung in stationär bewegter und ruhender Luft." *Hann-Band, Meteor. Z.*, SS. 243–254 (1906).
35. McIntyre, D. P., "On the Air-Mass Temperature Distribution in the Middle and High Troposphere in Winter." *J. Meteor.*, 7: 101–107 (1950).
36. Miller, J. E., *A Study of Vertical Motions in the Atmosphere.* Mimeogr., 68 pp., New York University, 1946.
37. —— "Studies of Large Scale Vertical Motions of the Atmosphere." *Meteor. Pap., N. Y. Univ.*, Vol. 1, No. 1, 49 pp. (1948).
38. Mintz, Y., "On the Kinematics and Thermodynamics of the General Circulation of the Atmosphere in the Higher Latitudes." *Trans. Amer. geophys. Un.*, 28: 539–544 (1947).
39. Namias, J., and Clapp, P. F., "Confluence Theory of the High Tropospheric Jet Stream." *J. Meteor.*, 6: 330–336 (1949).
40. Nyberg, A., "Synoptical-Aerological Investigation of Weather Conditions in Europe, 17–24 April 1939." *Medd. meteor.-hydr. Anst. Stockh.*, No. 48, 122 pp. (1945).
41. Palmén, E., "On the Distribution of Temperature and Wind in the Upper Westerlies." *J. Meteor.*, 5: 20–27 (1948).
42. —— "On the Origin and Structure of High-Level Cyclones South of the Maximum Westerlies." *Tellus*, Vol. 1, No. 1, pp. 22–31 (1949).
43. —— and Newton, C. W., "A Study of the Mean Wind and Temperature Distribution in the Vicinity of the Polar Front in Winter." *J. Meteor.*, 5: 220–226 (1948).
44. Palmén, E., and Nagler, K. M., "An Analysis of the Wind and Temperature Distribution in the Free Atmosphere over North America in a Case of Approximately Westerly Flow." *J. Meteor.*, 5: 58–64 (1948).
45. —— "The Formation and Structure of a Large-Scale Disturbance in the Westerlies." *J. Meteor.*, 6: 227–242 (1949).
46. Petterssen, S., "Contribution to the Theory of Frontogenesis." *Geofys. Publ.*, Vol. 11, No. 6 (1936).
47. —— *Weather Analysis and Forecasting.* New York, McGraw, 1940.
48. Phillips, N. A., "The Work Done on the Surrounding Atmosphere by Subsiding Cold Air Masses." *J. Meteor.*, 6: 193–199 (1949).
49. Pick, W. H., and Bowering, D. F., "Cirrus Movement and the Advance of Depressions." *Quart. J. R. meteor. Soc.*, 55: 71–72 (1929).
50. Raethjen, P., "Zyklogenetische Probleme." *Arch. Meteor. Geophys. Biokl.*, (A) 1: 295–346 (1949).
51. Riehl, H., "On the Formation of West Atlantic Hurricanes." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 27, 67 pp. (1948).
52. Rossby, C.-G., "Planetary Flow Patterns in the Atmosphere." *Quart. J. R. meteor. Soc.*, 66 (Supp.): 68–87 (1939).
53. —— "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2: 187–204 (1945).
54. —— "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6: 163–180 (1949).
55. —— and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2: 38–55 (1939).
56. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose.* Berlin, J. Springer, 1948.
57. Schröder, R., "Die Regeneration einer Zyklone über Nord- und Ostsee (Analyse der Wetterepoche: 29 September bis 3 Oktober 1912)." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 4: 49–118 (1929).
58. Solberg, H., "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones." *P. V. Météor. Un. géod. géophys. int.*, Edimbourg, 1936. II, pp. 66–82 (1939).
59. Spar, J., "Synoptic Studies of the Potential Energy in Cyclones." *J. Meteor.*, 7: 48–53 (1950).
60. Starr, V. P., "On the Production of Kinetic Energy in the Atmosphere." *J. Meteor.*, 5: 193–196 (1948).
61. Van Mieghem, J., "Perturbations d'un courant atmosphérique permanent zonal." *Inst. R. météor. Belg., Mém.*, 18: 1–35 (1945).
62. —— "Interprétations énergétiques du critère d'instabilité de Kleinschmidt." *Bull. Acad. Belg. Cl. Sci.*, (5) 31: 345–352 (1945).
63. Willett, H. C., *Descriptive Meteorology.* New York, Academic Press, 1944.
64. Yeh, T.-c., "On Energy Dispersion in the Atmosphere." *J. Meteor.*, 6: 1–16 (1949).

ANTICYCLONES

By H. WEXLER

U. S. Weather Bureau, Washington, D. C.

It is the purpose of this article to summarize briefly what is known of the origin, structure, and transformation of anticyclones and to describe their role in the general circulation and some aspects of their control of weather and climate.

It is natural for the division of subject matter for this Compendium to be made in terms of cyclones, anticyclones, general circulation, and so on. However, this breakdown, while very convenient for the editor and the reader, imposes certain difficulties on the author since it is quite impossible to discuss the anticyclone as a separate entity, with respect to either its origin or its role in the general circulation. For that reason, although much of the discussion here will deal directly with anticyclones, there will be occasional and unavoidable digressions into broader problems.

The writing of this article was made much easier by reference to the excellent review of anticyclones by C. E. P. Brooks [4] in *Some Problems of Modern Meteorology*, the famous British progenitor of this Compendium.

DEFINITION AND THERMAL STRUCTURE OF ANTICYCLONES

An anticyclone is a large atmospheric eddy which rotates clockwise in the Northern Hemisphere about a center at which barometric pressure is higher than that of the surrounding air. Since barometric pressure is a very close measure of the mass of overlying air, an anticyclone is characterized by an excess of air aloft. Thus, over an anticyclone either the effective height of the atmosphere is greater or the air is on the average denser over the same height. Since there exists no evidence to indicate a greater height of the atmosphere, it was once thought that the air over an anticyclone must be colder and denser at all levels than the air over a cyclone. However, from a study of mountain observations in Europe and western United States, Hann in 1876 [14] showed that up to the height of the mountains (3 to 4 km) anticyclones averaged warmer than cyclones except in a shallow surface layer. Furthermore, Hann stated the abnormal warmth and dryness was caused by subsidence. From recent aerological data more has been learned of the general thermal structure of anticyclones as distinguished from that of cyclones. For example, values compiled by W. H. Dines [7] and Palmén [22], based on soundings in England, showed that cyclones are characterized by a cold troposphere, a low (8.5 km),[1] warm (−50C) tropopause, and a warm stratosphere, while anticyclones have a warm troposphere (over a cold, shallow surface layer), a high (11.5 km), cold (−65C) tropopause, and a cold stratosphere.

In 1908, Hanzlík [15] demonstrated the existence of two types of European anticyclones, the shallow cold (polar) anticyclone moving rapidly, usually in the rear of a cyclone, and the deep warm anticyclone moving very slowly or not at all. Quite often the shallow cold anticyclone changes into a deep warm anticyclone, and in so doing slows down and sometimes becomes stationary.

So far as is known there is no statistical study available on the vertical temperature structure of the polar and warm type of anticyclones although it is likely that Palmén's statistics on anticyclones refer mostly to the warm type of anticyclone, since the true polar anticyclone is rare in England. As a rough generalization it may be said that in North America the polar anticyclone is characterized by a troposphere colder than its environment, especially in the lower portion, and a low (5 to 8 km), warm (−50C to −65C) tropopause and warm, lower stratosphere, while the warm anticyclone may have a thin, cold layer at the surface, a warm troposphere, a high (12 to 17 km), cold (−65C to −80C) tropopause, and a cold, lower stratosphere. Quite often, however, anticyclones exist which have the deep surface layer of cold air characteristic of polar anticyclones and the high cold tropopause of warm anticyclones. Such anticyclones, combining the high-pressure producing qualities of both anticyclonic types, are usually extremely intense and would probably never be observed at low latitudes because of absence of cold tropospheric air. For example, in the nineteen unusually intense anticyclones in the North Atlantic and Eurasian regions listed by Scherhag [30], where central pressures ranged from 1047 to 1079 mb, the lowest latitude of location was 50°N.

ORIGIN OF ANTICYCLONES AND THEIR ROLE IN THE GENERAL CIRCULATION

Polar Anticyclone. The polar anticyclone is created by cooling of the surface layer of air, which loses its heat to an underlying cold surface by radiation [35] or by eddy conductivity. The lower temperature of the surface itself is caused by the nocturnal radiational cooling of snow and ice fields, such as those in polar regions, or by bodies of large thermal inertia (heat capacity), such as oceans and large lakes. The cooling and vertical shrinking of the surface layer of air depresses the isobaric surfaces aloft, creating an intensifying "polar cyclone" which causes an inflow of air across the isobars, thus creating a larger barometric pressure at the surface. This process was studied by

1. Values are taken from Palmén.

Wexler [36], using the Brunt-Douglas isallobaric concept, and later by Schmidt [31].

Warm Anticyclone. A satisfactory explanation of the warm, deep anticyclone has not yet been given. The attempted explanations are usually along three lines: radiative, advective, and dynamic. The reasoning based on radiation goes something like this: Since the troposphere over the warm-type anticyclone is warmer than the surroundings, the increased pressure must be caused by colder, denser air in the upper troposphere and lower stratosphere. The argument is then made that this cooling is caused by favorable radiative conditions over a restricted area. Those arguing advectively believe that the cold air results from northward advection of the cold equatorial upper troposphere and lower stratosphere. Various dynamic reasons for warm anticyclogenesis have been proposed, based on wave mechanics of the westerlies, lateral frictional drag, motion of polar domes, etc. Each of these proposed explanations will be discussed in turn.

Radiative Theory. The explanation of anticyclogenesis based on radiative cooling of the stratospheric air [18] appears to be the least convincing. Gowan [13] in his investigation of nocturnal cooling of the ozonosphere (where water vapor, ozone, and carbon dioxide are assumed to be the principal emitting and absorbing gases) finds the temperature decrease in eight hours from sunset varies from less than 0.1C at 15 km to 1C at 30 km, and thus concludes, "It seems certain that the cooling of the lower stratosphere, perhaps up to 30 km, is not governed by radiation." If one looks above 30 km (30-mb pressure), it is difficult to ascribe any significant anticyclogenetic effects to possible horizontal differences in the rates of radiative cooling of air in a layer whose contribution to sea-level pressure is so small.

Advective Theory. The advective theory of stratospheric cooling and anticyclogenesis appears superficially to be very attractive. As Brunt [5] states, the close apparent analogy between the polar and cyclonic stratospheres on the one hand and the equatorial and warm anticyclonic stratospheres on the other, tempts one to explain the anticyclogenesis as a consequence of solid currents moving from south to north. But if the entire atmospheric column at 50° latitude should be replaced by that at 10° latitude, then it can be shown from Wagner's aerological averages [34] that the net change of sea-level pressure would be practically zero. If, however, differential advection were established such that the mass of the column at 50°N between the surface and 8 km remained unchanged, say by the presence of westerly winds, while the air above 9 km came from 10° latitude with southerly winds, then the sea-level pressure would increase by 25 mb, an amount more than sufficient to account for most cases of anticyclogenesis. Thus it is theoretically possible for anticyclogenesis to be explained by high-level advection over a long trajectory from the south. But the problem is more complicated than the simple arithmetic exercise would indicate. Aerological experience indicates that the extreme type of vertical shear in the flow pattern postulated seldom if ever occurs, but that the full depth of the westerlies usually partakes in its meanderings (as would be expected in a barotropic atmosphere). Thus, for example, if the southerly advection from 10°N to 50°N took place above the top of a shallow (2-km) inert layer of surface air, the increase in pressure would only be 5 mb in summer and 11 mb in winter, amounts which are not sufficient to account for many cases of anticyclogenesis.

The advective transport of warm and cold air is intimately connected with the dynamics of the westerly flow pattern. If the westerlies are zonal in character, there would be no north-south advection across the belt of westerlies. True enough, there may be advective transport of a shallow (2-km) surface layer of polar air southward underneath the westerlies; this is associated with a certain type of dynamic anticyclogenesis discussed by Simmers [33] and Wexler [37]. But it is when the zonal westerlies break down into large-scale waves of great amplitude (meandering) that the most pronounced cases of warm anticyclones are found. The amplitude of these waves may become so large that the ridges break down into large-scale anticyclonic vortices to the north and the troughs change into cyclonic vortices to the south. Southerly advection accompanies this process, bringing northward over a shallow surface layer of polar air the warm, lower tropospheric air, and the cold upper tropospheric and lower stratospheric air characteristic of tropical latitudes. *However, advection is not the primary cause of the anticyclogenetic process, but rather both it and the initiation of the anticyclogenesis are the result of a major change in the character of the flow pattern in the westerlies.* It follows that the primary problem becomes one of dynamics although it is acknowledged that once initiated, the secondary advective effects may contribute materially to the anticyclogenesis.

Dynamic Theory. The dynamical explanations of warm anticyclones may be divided into two main types: First, an explanation based on the large-scale changes in the westerly flow pattern, as discussed briefly above; and second, once the "background" flow pattern has become established, an explanation based on cross-isobaric flow of air resulting from nongradient winds. Let us discuss briefly each of the proposed explanations.

In the past two decades intensive research has been performed on what might be called the "wave mechanics" of the westerlies. The pioneering work of V. Bjerknes and collaborators [3] and J. Bjerknes [2], was given further impetus by Rossby and collaborators [28] and was aided by greatly improved aerological coverage over most of the Northern Hemisphere. At the present time the recognition and explanation of the complex behavior of the westerlies have come to be accepted as the central problem of meteorology. Numerous articles in this Compendium deal with various phases of this problem in a much broader manner than can be given here. But in any discussion of anticyclonic origin, development, movement, dissipation, etc., it is necessary at first to depart from the narrow confines of the area usually covered by an anticyclone and obtain a broader

view of the behavior of the westerlies and their influence on anticyclones both of the polar and warm types.

Most of this paragraph is based on Rossby and Willett's discussion of the circulation changes accompanying a complete index cycle [29]. If the westerlies are zonal over the hemisphere, corresponding to "high index," then the anticyclones are located to the south and are large in area, but not abnormally intense, with axes oriented west-east. As the index decreases, the westerlies move southward and become less zonal, permitting the formation of stronger polar anticyclones to the north. With further decrease of the zonal index to its minimum value, the westerlies develop strong ridges and troughs, with cutting-off of warm anticyclones in the higher latitudes and cold cyclones in the lower latitudes. This is the period of maximum dynamic anticyclogenesis of polar anticyclones to the north. When the index begins to increase again, the westerly waves decrease their amplitude and become longer, the higher-latitude anticyclones dissipate, and the "cycle" is ready to begin over again.

This simplified account of the relationship of the westerlies and anticyclones is one of *association* only, and does not pretend to isolate causes and effects. What is quite clear and outstanding, however, is that the phenomena must be considered on a hemisphere-wide scale and not as a local development closed off from the rest of the atmosphere. It is true that there are important local influences pertaining primarily to the character of the earth's surface, some of which will be discussed later, but these must be considered as secondary effects superimposed on the primary pattern of the general circulation.

There is definite evidence that the beginning of the mechanism that transforms a high-index, essentially zonal flow pattern into a low-index, essentially meridional flow pattern is the appearance of a "blocking action" in some portion of the westerlies which effectively obstructs the normally eastward motion of waves in the upper westerlies, or their sea-level counterpart in the form of anticyclones and cyclones. In the excellent aerological study of a case of the breakdown of zonal westerlies, Berggren, Bolin, and Rossby [1] showed how a nearly zonal flow pattern in the 500-mb westerlies, in which were imbedded wave cyclones, moving with uniform speed across the North Atlantic, encountered "blocking" in the eastern Atlantic which caused the westerly waves to fold up like an accordion as their wave length decreased and amplitudes increased. The increase in amplitudes finally resulted in the appearance of "cut-off" cold lows in the south and warm highs in the north. The authors showed that the appearance of blocking is somewhat similar to the "hydraulic jump" which develops in open-channel flow when the depth of the water current drops below a critical depth, and where a certain amount of the initial energy of flow is transformed into turbulent or eddy energy. If the analogy with atmospheric "blocking" is accepted, then perhaps a certain supercritical speed of the westerly jet results in the sudden breakdown of a zonal flow into the large-scale anticyclonic and cyclonic eddies. The exact manner under which this remarkable change in flow pattern occurs is unknown, nor is the process known whereby the blocking, once formed, proceeds upstream for considerable distances—in many cases as a recognizable pressure rise at high latitudes capable of being followed westward for more than 180° of longitude. The importance of blocking in forecasting was first pointed out by Garriott [12] and has received further attention recently by long-range forecasters [9, 21].

The immediate origin of kinetic energy of atmospheric large-scale motions is a knotty problem that has not been fully solved. Rossby and Willett, in their description of the index cycle cited above, explain that the return of the high index, or speed-up of the westerlies, is caused by the re-establishment of the normal, north-south thermal gradient caused by radiative cooling in higher latitudes and radiative heating in lower latitudes. But it is not clear how the new sources of atmospheric energy, in the form of internal heat energy or potential energy, are "tapped" to provide the increased kinetic energy.

Let us turn now to the second type of dynamical explanation of warm anticyclones, based on horizontal mass convergence arising from cross-isobaric flow associated with supergradient winds. Many ingenious solutions have been proposed to explain supergradient winds. One popular explanation is that based on the removal of an air parcel from its high momentum environment to an environment of lower momentum where the winds are gradient. The foreign parcel, maintaining its higher momentum at least momentarily, will, by virtue of the greater Coriolis force acting on it, be flung across isobars towards higher pressure. This explanation was first proposed by Helmholtz [16] in studying the stability of the circular vortex.

Recently many investigators, while agreeing in their use of this basic reasoning in explaining pressure changes, have differed widely in their explanations as to *how* the air parcels are removed from an environment of high momentum to one of low momentum. Rossby [26] proposed that current systems possessing curved lateral velocity profiles would create frictionally driven eddies in isentropic surfaces, and that these eddies would transport momentum laterally thus producing unbalanced flow. In the free atmosphere the lateral frictional stresses within isentropic surfaces would be of greater importance than the vertical frictional stresses arising as a result of vertical wind shear since, in the latter case, the normal thermal stability of the atmosphere would resist vertical motions. Rossby found from a simplified model that the percentual increase in total pressure drop across the current, resulting from the diffusion of momentum, would amount to 8 per cent. The significance of reasoning of this type is enhanced by the recent discovery of the narrow, fast, westerly "jet-stream" flanked by exceedingly strong shears and curvatures of the lateral velocity profile (*e.g.*, see Palmén and Nagler [23]).

Another explanation based on changes in environ-

ment has been proposed by Priestley [24] as an outgrowth of earlier work by Durst and Sutcliffe [8] on the role of vertical currents in producing nongradient winds. Given an increase of wind with height, an upward current will create subgradient winds and a downward current, supergradient winds. Priestley used this rule to explain anticyclogenesis where the downward vertical motions were initiated by (1) frictional outflow from the surface anticyclone or (2) radiative cooling of the tropospheric air over a large area. But, as Priestley showed, neither (1) because of lack of sufficiently large vertical wind shear in warm anticyclones, nor (2) because of lack of strong enough radiative cooling over a sufficiently large area, are by themselves adequate to explain anticyclogenesis as observed. Priestley believes that the two processes must operate simultaneously to produce the observed pressure rises. However, a severe restriction on Priestley's model is that an anticyclone with surface frictional outflow must already be in existence before the nonradiatively caused descent can occur. According to Priestley's reasoning, then, it follows that surface features, such as topography or thermal stability of the air, which diminish the surface frictional outflow of the anticyclone, weaken the downward currents, thus causing a lack of supergradient winds and retarding anticyclogenesis. This result is not in agreement with the high frequency of anticyclones observed over physiographic depressions, such as the Great Basin in the western United States or over cold water, such as the Great Lakes area in summer [32, p. 172]. Regarding the radiatively caused descent, Priestley showed that for a reasonable model, a cooling of the free air of 2C per day over an area 2500 km in diameter is necessary to maintain the anticyclone against normal surface frictional outflow. This cooling value seems too high since Priestley quoted Elsasser [10] with respect only to the emissive cooling of the free atmosphere by long-wave radiation, and did not mention the heating by absorption of solar radiation, which amounts to ½C per day. Also, according to Elsasser [10, p. 95], since water vapor is the principal emitter, the emissive cooling itself decreases very rapidly with moisture content, and this would mean that the dry air above an anticyclone would cool even less than Priestley's assumed value.

In a search for the presence of available energy in the atmosphere which might be used to create unbalanced currents, a sudden cyclogenesis upstream and the resulting acceleration and "banking" of the westerlies to the south has been suggested [21, p. 63]; but this explanation is merely substituting the unknown solution of the cyclogenetic problem for that of the anticyclogenetic problem. A more promising and not unrelated lead is that of the propagation of energy with the group velocity of a train of waves in the westerlies where air parcels move with constant absolute vorticity. This concept, removed from the restraints of sinusoidal wave motion first assumed by Rossby, forms the basis of the Charney-Eliassen numerical prediction technique of westerly waves [6].

The following is a summary of the dynamic causes of anticyclones which appeals most to the writer. It is based on the combined effects of wave mechanics of the westerlies, lateral mixing, radiative cooling of the free atmosphere, thermal stability of the atmosphere (especially that of the surface layer), and the character of the underlying earth's surface. Lateral frictional stresses will cause supergradient winds south of the axis of the westerlies. This will create a northerly component of flow which will vanish far to the south. The convergence of air to the south of the zonal westerlies will result in the formation of an anticyclonic *ridge*. Waves in the westerlies will fracture the ridge into anticyclonic *cells*. Lateral frictional stresses which continue to operate on the sinusoidal westerlies will further increase the convergence into the anticyclonic cells. Factors which block the downward flow of the converging air in the cells and its return to the westerlies in the surface frictional layer will favor anticyclogenesis. These optimum factors are (1) greater thermal stability of the free air combined with low rates of radiative cooling of the air (as in dry air) to hinder the nonadiabatic descent of the air, (2) presence of a very stable surface layer of air, or "shielding layer," which will prevent the descending air from coming under influence of the surface frictional stresses, and thus inhibit air transport north and south across isobars, and (3) presence of a large physiographic depression such as the Great Basin of the western mountains in the United States which "traps" the descending air and interferes with its motion away from the area. Of all these auxiliary conditions, the presence of a shielding layer seems to be most significant. The formation of such a shielding layer is favored by local cooling of air over large areas such as cold water surfaces, the undercutting of the atmosphere by a subsiding and increasingly stable wedge of polar continental air, and by radiative cooling and air drainage in large physiographic depressions.

An important dynamic consequence of the horizontal convergence of the air responsible for anticyclogenesis would be an increasing cyclonic circulation if one applied the conservation of absolute angular momentum to the contraction of rings of air in the absence of external forces. Thus, if convergence is an important factor, one might expect to observe a cyclonic core embedded in the center of an anticyclone aloft. This has actually been observed by Simmers [33, p. 88], but appears to be observed rarely both because of lack of a sufficient density of upper-wind observations and the tendency of the peripheral anticyclonic circulation to obliterate by friction the weak cyclonic circulation at the core.

It should be emphasized that, given the proper conditions, the lateral frictional stresses south of the axis of the westerlies are exerted on all the air flowing through the region in question. The supergradient winds thus created cause a northerly component of flow, probably too small to be observed directly. The resulting horizontal convergence will be relieved by vertical motions downward in lower levels and upward in the upper troposphere. These latter motions will create a

higher, colder upper troposphere and tropopause. The presence of this "cold cap" above surface anticyclones was pointed out by Mügge [18], who showed that advection from the south was not the sole cause. An excellent analysis of a case of anticyclogenesis, showing the vertical structure, vertical motions, and the cold cap, has been given by Fleagle [11]. Fleagle stresses the fact that as one follows the motion of the anticyclone, the upper cold cap is not composed of the same air particles, which is a further indication of the existence of dynamic processes in anticyclogenesis.

MOTION AND TRANSFORMATION OF ANTICYCLONES

The polar anticyclone in its source region is characterized by a large surface inversion and a nearly isothermal layer above, extending usually to no more than 3 km above the surface which marks the top of the true polar continental air [35]. When the anticyclone moves away from its source region and travels over a warm surface, both the heating from below and the onset of mechanically induced turbulence wipe out the surface inversion, and may, in fact, create a steep and even superadiabatic lapse rate in the surface layer. Above this layer of steep lapse rate is usually found an inversion which increases in thickness and magnitude as the air in the isothermal layer subsides. The top of the inversion marks the top of the true polar continental air, while the air above, despite its subsidence, maintains a greater lapse rate. The spectacular nature of the "subsidence inversion," separating as it does the cool, moist, cloudy air below from the warm, clear, dry air aloft, has been studied intensively by Namias [19] and Hewson [17]. The more pronounced cases of subsidence are believed to accompany the transformation of the polar anticyclone into the warm or dynamic anticyclone, and may indeed be considered as one of the by-products of such a transformation.

Forecasting the motion of polar anticyclones away from their source regions is a problem of considerable interest to meteorologists. Often one polar anticyclone follows another in a regular procession from the source region and at other times intense large anticyclones remain stationary for many days. From observations it has been noted that a strong belt of zonal westerlies equatorward of the source region tends to act as a "barrier" preventing appreciable motion southward of polar anticyclonic offshoots. But if waves develop in the westerlies, the troughs act as "spillways" for the release of polar anticyclones.

Another approach to the problem was recently presented by Rossby [27] who studied the conditions necessary for the motion, away from the geographical pole, of barotropic vortices embedded in a motionless barotropic atmosphere of great depth. Taking into account the variation of the Coriolis force with latitude, he found that if vortices are located, or are displaced, away from the geographic pole, the cyclonic vortices will move poleward while the anticyclonic vortices will move equatorward with increasing speed (friction neglected). The motion of the polar anticyclonic vortex will be accompanied by a flattening of the dome and a conversion of the loss in potential and in rotational kinetic energy into kinetic energy of translation of the dome. Some of this translational energy is imparted to the environment which thus acquires motion. Rossby showed that, as the anticyclonic vortex moves southward (in the Northern Hemisphere) and sets the environment into motion, the anticyclonic center (center of the streamlines) is displaced westward and increasingly so with increased speed southward of the vortex. This process represents a dynamic movement of the center of the original polar anticyclone westward from its original position under the highest point of the polar dome. This conclusion is in excellent agreement with the fact long noted by synoptic meteorologists that the top of a southward moving polar dome is found about midway between the surface cyclone center and the anticyc one center to the west. It is not known whether this type of "dynamic anticyclone" leads to higher pressure than that originally found in the center of the polar anticyclone. At the time the center of the anticyclone is shifted westward, a crescent "moon-shaped" cyclonic center appears on the eastern edge of the polar dome. Considering the circular boundary of the original polar dome as the "front," Rossby showed that the southward displacement of the dome is accompanied by frontolysis and a spreading out of the cold air along the western edge, and frontogenesis and steepening of the front along the eastern edge of the cold dome. Both of these latter conclusions are in excellent agreement with observations.

In applying Rossby's results to the problem of release of polar anticyclones initially located away from the geographical pole, it must be remembered that the thermally produced anticyclone should theoretically be accompanied by a polar cyclone aloft at the source region [36], thus indicating the baroclinicity of the polar atmosphere. But if one assumes that Rossby's results, derived for a barotropic atmosphere, hold for this case, there would be a tendency for the surface anticyclone to move equatorward and the upper cyclone to move poleward. This "splitting-off" of the surface anticyclone from its upper cyclone is again quite reasonable in the light of aerological experience. However, as Rossby points out, if the concomitant sinking of the polar dome occurs in an actual atmosphere of normal stability, the resulting convergence aloft should create a cyclonic vortex at upper levels. This dynamically created cyclonic vortex may accompany the cold air dome equatorward, but sooner or later it will have to turn poleward in accordance with the increased poleward force acting on it. Thus, it is likely that, both at the source region and later on in the trajectory of the polar anticyclone, forces operate in such a way as to remove cyclones aloft, and pave the way for upper-level anticyclogenesis.

The likelihood of transformation of a polar anticyclone into a dynamic anticyclone appears to be a function primarily of its position relative to the wave pattern of the westerlies. If the polar anticyclone is associated closely with the motion of a pronounced

trough aloft, it will probably not change into a dynamic anticyclone until the trough fills in response to a change in the wave pattern; this may involve consideration of forces acting over the entire belt of westerlies. When the upper trough fills, leaving the polar anticyclone well to the south of the westerlies, surface conditions become important insofar as they maintain or strengthen the shielding layer which, as explained previously, will permit the accumulation of the subsiding air aloft, leading to anticyclogenesis.

There remain problems not explained by Rossby's theory. For example, a deep combined polar and dynamic anticyclone at the source region might be expected to move more rapidly equatorward than a shallow polar anticyclone possessing its "thermal" cyclone aloft. This is not in accord with synoptic experience and, in fact, the deep anticyclone may remain stationary for such long periods that it appears on a mean monthly map. Such a situation occurred in February 1947 over northern Greenland, as described in the mean by Namias [20]. In this case the deep anticyclonic vortex appeared to be bounded on the south by cyclonic vortices. Applying Rossby's theory we might interpret the stagnancy of the vortex pattern as an equilibrium between forces tending to drive the anticyclone vortex southward and the cyclonic vortices northward.

The ultimate fate of anticyclones which survive as individual entities in their progress to south and east seems to be absorption in and strengthening of the subtropic anticyclones. The subtropic anticyclones themselves are examples of dynamically caused anticyclones located south of the westerlies over areas where both shielding layers (trade inversion) and low, surface frictional outflow favor location of the anticyclones [37].

ANTICYCLONES AND WEATHER

Too often, in the minds of the meteorologist interested in weather forecasting, the anticyclone is considered merely as something that occupies the space on the weather maps between one front and the next. To be sure there is always the problem of forecasting the minimum temperature and the formation and dissipation of stratus and fog, but on the whole, anticyclonic weather is a period of relaxation for the forecaster. This apparent lack of association of spectacular weather with anticyclones probably accounts for the far greater emphasis on study of cyclones. However, under certain conditions the anticyclone does exert a strong, though perhaps subtle, control on weather phenomena, and it is the purpose of this section to describe a few examples.

Summer Showers in the United States. The anticyclonic control of summer shower distribution in the United States has been demonstrated by Reed [25] and Wexler and Namias [38]. In the warm season the westerlies are displaced northward near the United States-Canadian border thus allowing frictionally driven anticyclonic eddies to cover much of the United States. These eddies have preferential average positions —one center located over the central Great Plains and the other off the southeastern Atlantic Coast [37]. These preferential locations indicate that something more than purely dynamic effects of the westerlies are involved in the maintenance of these eddies. The "anchoring" of the summer anticyclonic eddies in the United States is believed to be the result of the very strong horizontal solenoidal field created by the cold air above the eastern Pacific Ocean and the heated air above the elevated plateau region of the western United States, acting in such a manner as to form a pronounced trough in the westerlies located in a mean position off the West Coast. This thermally produced trough creates a "resonance" trough in the Mississippi Valley, the exact position changing according to the varying strength of the westerlies and the position of the West Coast trough. Other troughs are found farther downstream, but in the absence of a long series of aerological observations their average locations must be deduced from other evidence. Between the troughs of the westerlies are the anticyclonic cells composed of two main currents spiraling clockwise—one a dry current from the north and the other a moist current from the south. Summer showers occur more frequently under the deep moist tongues. The preferential average locations of the positions of the anticyclonic eddies and their moist and dry tongues have profound effects on the summer climate in the United States. For example, the dry tongue curving around the West Coast trough accounts for the dry summers in southern California and western Arizona, the moist tongue curving anticyclonically from the Gulf of Mexico around the western anticyclone accounts for the summer maximum in shower activity in eastern Arizona, New Mexico, and portions of Colorado. The dry tongue of this same eddy curving anticyclonically down the Mississippi Valley into the West Gulf and eastern Texas accounts for the fact that this region has less than 50 per cent of the summer precipitation enjoyed by the eastern Gulf states which are under the influence of the moist tongue of the eastern anticyclone. Similar anticyclonic control of summer shower activity must exist in other regions south of the westerlies, and perhaps enough data now exist to permit location of the mean position of the centers of the eddies for the rest of the Northern Hemisphere. The eddy positions and sizes vary from year to year; this has a pronounced effect on the summer rainfall distribution.

The "Indian Summer" Anticyclones in the Eastern United States. During the autumn, and especially in October, stagnant anticyclones appear frequently over the eastern United States and account for calm, mild, hazy days and cool nights. The haze results from the inability of the atmosphere to disperse combustion products and other surface material. The large thermal stability of the atmosphere and the strong solar heating result in large ranges of diurnal temperature; insolation, however, in many cases does not succeed in "burning-off" entirely the inversion created by subsidence within the anticyclone. In some extreme cases where pockets of cold air form in hilly country or river valleys, not even the shallow nocturnal inversion may be completely destroyed by solar heating. This is par-

ticularly true if a thick fog reflects much of the incident solar radiation. In industrial areas located in river or valley locations, pollution products help to create dense, peristent fogs which not only protect the inversion from "burning-off" but serve as a carrier for pollution products confined to the surface layer by the inversion. This in essence is the meteorological background of the disastrous Donora, Pennsylvania, "smog" of October 25–31, 1948 [32] and the similar case in Liège, Belgium, in 1930. Two earlier cases of Donora smog occurred in 1923 and 1938, each time in October. A suspected case occurred in April 1945. Wexler [32] analyzed the necessary conditions for a deep persistent anticyclone to occur over the eastern United States, and found that autumn (in particular October) fulfilled these conditions best, with winter next.

"Back-Door" Cold Fronts. In late spring and summer along the North Atlantic Coast of the United States some of the most outstanding forecast failures occur when a seemingly persistent heat wave is broken unexpectedly by a sudden push of a shallow layer of polar Atlantic air southwestward along the coast. It is no exaggeration to say that this type of weather process brings the greatest relief to the heat-harassed populace and the bitterest self-recrimination to the forecasters involved. Overnight drops in temperature from 98F to 65F and in dew-point temperature from 82F to 62F are by no means uncommon in Washington, D. C.; it would seem offhand that such spectacular weather changes must be accompanied by equally spectacular and well-marked patterns on the weather map, but such unfortunately is not the case. The preceding conditions are usually those in which a cold front has pushed into northeastern United States from Canada, stalled in the vicinity of Boston or New York and then, showing every sign of retreating northeastward as a warm front, gathers sufficient energy to push southward along the coast in a narrow tongue. The cause is usually a sudden but slight anticyclogenesis in the cold air north of the front, which "energizes" the surface cold air into motion. The anticyclogenesis itself seems to have its origin in the warm air above the front and not in any sudden accumulation of cold air below. This tantalizing and important problem has not had the study it deserves.

EXAMPLE OF AN UNUSUAL ANTICYCLONIC THERMAL STRUCTURE

In the first section some "typical" values of the principal upper-air thermal features of anticyclones were presented. It is altogether probable, however, that other more bizarre combinations of upper-air structure may contribute to the excess of pressure at the surface, known as the anticyclone. For example, the writer, in attempting to find illustrations of "typical" soundings of polar and warm anticyclones, studied the anticyclone of December 23–26, 1949, which established some record high pressures in New England. In this case a polar anticyclone left its source region in the Canadian Northwest on December 23 with a sea-level pressure of 1033 mb and typically cold air in the troposphere; at the same time, an older polar anticyclone, which had previously pushed southward to the southwestern United States, proceeded to move northeastward with an initial sea-level pressure of 1028 mb. The anticyclones were steered by their respective currents at 500 mb and approached each other on a collision course, intensifying slightly as they did so, and finally amalgamating with an explosively sudden increase of pressure of 17 mb in 24 hours from the 24th to the 25th, the maximum pressure of 1054 mb being reached in northern Maine. In Fig. 1 it is seen that the sounding at Fort Smith, N.W.T., (665 ft elevation), taken at 0300Z on December 23 when the polar anticyclone of 1033 mb was nearly centered at the station, exhibits the strong surface inversion and isothermal layer above, characteristic of polar continental air, but possesses a high (8.7 km), moderately cold (−60C) tropopause; above the tropopause inversion a steep lapse rate is found to the top of the sounding, apparently indicating the probable presence of another higher, and perhaps colder, tropopause above. In Fig. 1 is also shown a sounding taken sixty hours later at Caribou, Maine (628 ft elevation), when the center of the amalgamated anticyclone of central pressure of 1054 mb was close to the station. This sounding shows a typically modified polar continental air mass in low levels; that is, a steep lapse rate in a thin surface layer and above this, to 2.5 km, marked stability and dryness characteristic of subsidence. Above, a steep

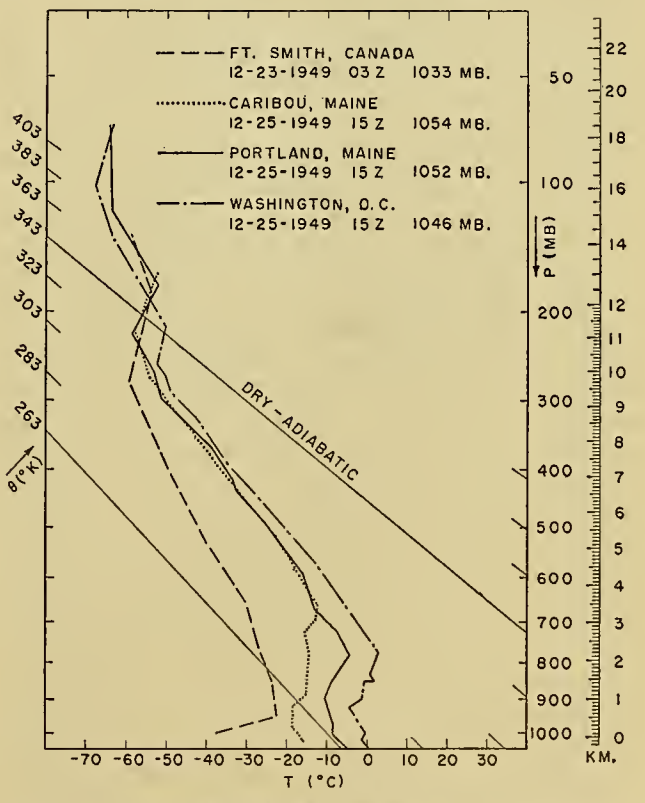


Fig. 1.—Soundings taken near the center of the anticyclone of December 23–26, 1949. Tabulated pressures are reduced to sea level.

lapse rate is found which terminates at a tropopause at 11.2 km, and, in view of the high surface pressure, a surprisingly warm tropopause (−58C). This sounding unfortunately ends at 13.0 km (−52C). In order to obtain an indication of temperatures above the top of the Caribou sounding, recourse was made to the closest sounding which penetrated significantly above 13 km, and was still under the anticyclonic influence. This turned out to be the sounding for Portland, Maine (225 miles south-southwest of Caribou), taken at 1500 Z, December 25; this station has a sea-level pressure of 1052 mb. The Portland sounding shows the presence of two well-marked tropopauses, the lower at 11.1 km, temperature −59C; this is undoubtedly the same tropopause as found over Caribou since the potential temperatures at the bases of the inversions are nearly the same, 330K and 332K, respectively. The base of the upper tropopause at Portland is found at 15.1 km, −64C, and 387K potential temperature. It therefore appears reasonable that the Caribou sounding taken near the center of the 1054-mb anticyclone would also have possessed an upper, high (15 km), cold (−65C) tropopause characteristic of warm anticyclones.

The presence of the relatively high, cold tropopause of the polar anticyclone at its source region, although unusual, is not so surprising, but the presence of a double tropopause in the warm anticyclone was unexpected.

The high potential temperature (387K) of the upper tropopause is worthy of special note since it is even higher than the potential temperature of the equatorial tropopause, 380K, as given in *Physikalische Hydrodynamik*, [3, p. 671]. The Swan Island (17.4°N, 83.9°W) sounding for 1500Z, December 25, has a well-marked tropopause inversion at 17 km, a temperature of −78.2C, and a potential temperature of 388K. In the Washington, D. C., soundings made at twelve hours before and twelve hours after 1500Z of December 25, the potential temperatures at the bases of the exceedingly well-marked upper tropopause inversions are even higher, 395K, 403K, and 408K, respectively. The Washington sounding for 1500Z, December 25, is also shown in Fig. 1. The remarkably high value of the tropopause potential temperatures observed at Washington, D. C., and Portland, Maine, could not have been achieved by adiabatic lifting of the prevailing tropopause at middle latitudes whose potential temperatures of 320K to 360K [3] are so much lower than those observed in the case under discussion. Nor does there appear to be present any radiative process which would cause the formation of an upper inversion over such a large area. It appears certain that in this situation there must have been advection northward of the tropopause from equatorial latitudes to at least the latitude of Portland, Maine, (43.7°N). The complete aerological discussion of this interesting case of anticyclogenesis is outside the scope of this article, but it should certainly be studied in detail to see how each of the large-scale factors—mechanics of the westerly waves and advection from equatorial latitudes—contributed to the anticyclogenesis.

SOME UNSOLVED PROBLEMS

The central problem concerning anticyclones is the explanation of the warm, deep "dynamic" anticyclone. The "piling-up" of air which is associated with this type of anticyclone is of course the counterpart to the removal of air which characterizes the cyclone. Complete explanations of both of these fundamental phenomena are still lacking. The deep anticyclone may be created out of transformation of a shallow polar anticyclone, or it may arise in its own right. Although some promising clues have been uncovered involving configuration of the westerlies and surface conditions, their influence on the cause and location of deep anticyclones is not fully known. That dynamic factors are important is indicated by semipermanent anticyclones located in the subtropics. That surface conditions are important is indicated by certain preferential locations of deep anticyclones. The role of advection in the upper troposphere in instituting or strengthening anticyclogenesis requires clarification. Whether changes in the height and temperature of the tropopause precede, accompany, or follow anticyclogenesis deserves further study. The existence and location of the high-level anticyclones south of the westerlies should be examined, especially in relation to tropical rainfall patterns.

REFERENCES

1. Berggren, R., Bolin, B., and Rossby, C.-G., "An Aerological Study of Zonal Motion, Its Perturbations and Break-Down." *Tellus*, Vol. 1, No. 2, pp. 14–37 (1949).
2. Bjerknes, J., "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54: 462–466 (1937).
3. Bjerknes, V., and others, *Physikalische Hydrodynamik*. Berlin, J. Springer, 1933. (See pp. 739–797)
4. Brooks, C. E. P., "Some Problems of Modern Meteorology. No. 9—The Origin of Anticyclones." *Quart. J. R. meteor. Soc.*, 58: 379–387 (1932). (Later published by the Royal Meteorological Society in *Some Problems of Modern Meteorology*. London, 1934.)
5. Brunt, D., *Physical and Dynamical Meteorology*, 2nd ed. Cambridge, University Press, 1944. (See p. 378)
6. Charney, J. G., and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
7. Dines, W. H., "Cyclones and Anticyclones." *J. Scot. meteor. Soc.*, (3) 16: 304–312 (1914).
8. Durst, C. S., and Sutcliffe, R. C., "The Importance of Vertical Motion in the Development of Tropical Revolving Storms." *Quart. J. R. meteor. Soc.*, 64:75–84 (1938).
9. Elliott, R. D., and Smith, T. B., "A Study of the Effects of Large Blocking Highs on the General Circulation in the Northern Hemisphere." *J. Meteor.*, 6:67–85 (1949).
10. Elsasser, W. M., "Heat Transfer by Infrared Radiation in the Atmosphere." *Harv. meteor. Studies*, No. 6 (1942).
11. Fleagle, R. G., "The Fields of Temperature, Pressure, and Three-Dimensional Motion in Selected Weather Situations." *J. Meteor.*, 4: 165–185 (1947).
12. Garriott, E. B., *Long-Range Weather Forecasts*. W. B. Bull. No. 35, U. S. Weather Bureau, Washington, D. C., 1904. (See pp. 61–62)
13. Gowan, E. H., "Night Cooling of the Ozonosphere." *Proc. roy. Soc.*, (A) 190: 227–231 (1947).

14. Hann, J. v., "Über das Luftdruck-Maximum vom 23. Jänner bis 3. Februar 1876 nebst Bemerkungen über die Luftdruck-Maxima im Allgemeinen." *Z. öst. Ges. Meteor.*, 11: 129–135 (1876).
15. Hanzlík, S., "Die räumliche Verteilung der meteorologischen Elemente in den Antizyklonen." *Denkschr. Akad. Wiss. Wien.*, 84: 163–256 (1909).
16. Helmholtz, H. v., "Über atmosphärische Bewegungen, I." *S. B. preuss. Akad. Wiss.*, SS. 647–666 (1888). (Trans. by C. Abbé in "The Mechanics of the Earth's Atmosphere." *Smithson. misc. Coll.*, No. 843 (1891). (See pp. 78–93))
17. Hewson, E. W., "The Application of Wet-Bulb Potential Temperature to Air Mass Analysis." *Quart. J. R. meteor. Soc.*, 62· 387–420 (1936). II, "Ascent of Air at Warm Fronts." *Ibid.*, 63: 7–30 (1937). III, "Rainfall in Depressions." *Ibid.*, 63: 323–337 (1937). IV. *Ibid.*, 64: 407-418 (1938).
18. Mügge, R., "Über warme Hochdruckgebiete und ihre Rolle in atmosphärischen Wärmehaushalt." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 3: 239–266 (1927).
19. Namias, J., "Subsidence Within the Atmosphere." *Harv. meteor. Studies*, No. 2 (1934).
20. —— "Characteristics of the General Circulation over the Northern Hemisphere During the Abnormal Winter 1946–47." *Mon. Wea. Rev. Wash.*, 75: 145-152 (1947).
21. —— and Clapp, P. F., "Studies of the Motion and Development of Long Waves in the Westerlies." *J. Meteor.*, 1: 57–77 (1944).
22. Palmén, E., "Aerologische Untersuchungen der atmosphärischen Störungen." *Mitt. meteor. Inst. Univ. Helsingf.*, No. 25, 65 pp. (1933).
23. —— and Nagler, K. M., "An Analysis of the Wind and Temperature Distribution in the Free Atmosphere over North America in a Case of Approximately Westerly Flow." *J. Meteor.*, 5: 58–64 (1948).
24. Priestley, C. H. B., "Atmospheric Pressure Changes: The Importance of Deviations from the Balanced (Gradient) Wind." *Aust. J. sci. Res.* (A) 1: 41–57 (1948).
25. Reed, T. R., "The North American High-Level Anticyclone." *Mon. Wea. Rev. Wash.*, 61: 321–325 (1933).
26. Rossby, C.-G., "On the Mutual Adjustment of Pressure and Velocity Distributions in Certain Simple Current Systems, I and II." *J. mar. Res.*, 1: 15–28 (1937), 239–263 (1938).
27. —— "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6: 163–180 (1949).
28. —— and Collaborators, "Relation Between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2: 38–55 (1939).
29. Rossby, C.-G., and Willett, H. C., "The Circulation of the Upper Troposphere and Lower Stratosphere." *Science*, 108: 643–652 (1948).
30. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose.* Berlin, J. Springer, 1948. (See p. 126)
31. Schmidt, F. H., "On the Causes of Pressure Variations at the Ground." *Meded. ned. meteor. Inst.*, Ser. B, Deel 1, Nr. 4 (1946).
32. Schrenk, H. H., and others, "Air Pollution in Donora, Pa." *Publ. Hlth. Bull. Wash.*, No. 306 (1949).
33. Simmers, R. G., "Isentropic Analysis of a Case of Anticyclogenesis." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 1 (1938). "Fluid Mechanics Applied to the Study of Atmospheric Circulations. Part. I—A Study of Flow Patterns with the Aid of Isentropic Analysis." *Ibid.*, pp. 72–125.
34. Wagner, A., "Klimatologie der freien Atmosphäre," Bd. 1, Teil F, 70 SS. *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr. Berlin, Gebr. Bornträger, 1931.
35. Wexler, H., "Cooling in the Lower Atmosphere and the Structure of Polar Continental Air." *Mon. Wea. Rev. Wash.*, 64: 122–136 (1936).
36. —— "Formation of Polar Anticyclones." *Mon. Wea. Rev. Wash.*, 65: 229–236 (1937).
37. —— "Some Aspects of Dynamic Anticyclogenesis." *Dept. Meteor. Univ. Chicago, Misc. Rep.* No. 8, 28 pp. (1943).
38. —— and Namias, J., "Mean Monthly Isentropic Charts and Their Relation to Departures of Summer Rainfall." *Trans. Amer. geophys. Un.*, (19th Annual Meeting) Pt. I: 164–170 (1938).

MECHANISM OF PRESSURE CHANGE

By JAMES M. AUSTIN

Massachusetts Institute of Technology

The problem of weather forecasting has been intimately related to the problem of the prediction of the pressure field and, therefore, to the mechanism of pressure change. For this reason meteorologists have sought an explanation of pressure change. However, despite the efforts of research workers during the past century it appears that there is not yet a completely satisfactory explanation of this phenomenon. In view of the uncertain state of knowledge it is the intention of the author to review briefly the more prominent theories of the pressure-change mechanism.

It is important to note that this discussion refers to pressure changes and not to pressure systems. A pressure system and its associated wind and temperature fields may be considered to be the result of a complex series of events following pressure changes. It is difficult, therefore, to investigate pressure changes through a study of the characteristics of pressure centers. For example, a center of a migratory pressure system is intermediate between a pressure rise and fall. This distinction between pressure change and pressure distribution is emphasized because it is believed that too much attention has been focused on meteorological data above the centers of pressure systems.

Historical Note

During the latter part of the nineteenth century Ferrel [8] developed a comprehensive theory of atmospheric processes based upon his own conclusions and the ideas of the earlier meteorologists. Ferrel explained the development of a cyclone as a consequence of localized heating which causes an outflow at high levels and, therefore, a surface pressure fall. The importance of low-level convergence, high-level divergence, and the earth's rotation in the development of the wind field around the pressure fall was well recognized. The heating was attributed principally to the release of latent heat with the condensation of water vapor in conditionally unstable air.

Several objections against this thermal theory were raised by Hann [14] and others. One group of objections which was directed against the latent heat as the cause of the warming could be met by allowing for other processes which would produce localized heating. However, the more serious objections were based upon temperature observations at mountain stations which frequently showed high temperatures over surface anticyclones and low temperatures over surface cyclones. These observations were considered to contradict a heating or cooling hypothesis and, therefore, there was a search for a more acceptable explanation of pressure change. The group of theories which attempt to explain pressure changes without localized heating or cooling as a primary cause are frequently referred to as *dynamic* explanations. An early example is given by Hann [14] who postulated that cyclones originate as a consequence of the damming and acceleration of the high-level air currents. A defect of many such *dynamic* theories is the failure of the authors to give specific details of the mechanism.

Before proceeding further with other theories it is important to investigate to what extent the empirical evidence on pressures and temperatures violates the thermal theory. The temperature data which were used to discredit Ferrel's hypothesis were obtained in the vicinity of the centers of mature slow-moving pressure systems. With such systems the temperature field has been greatly modified by vertical motion. Consequently these data on the temperature distribution in one layer of the atmosphere do not invalidate an argument that pressure falls and rises as a result of localized warming and cooling, respectively.

A survey of the early literature then suggests the following conclusions:

1. A thermal explanation of pressure change was unjustly criticized by reference to temperature observations taken above the centers of mature pressure systems.
2. If the thermal theory is valid there must be a more adequate explanation of the localized heating.
3. The observations which were believed to contradict a thermal theory greatly stimulated research on the so-called "dynamic" theories.

Pressure-Change Theories

Insofar as the pressure at a point is the weight of the air column above that point, it is convenient to discuss pressure changes by means of the hydrostatic equation

$$p_h = \int_h^\infty \rho g \, dz. \tag{1}$$

It follows from this equation that, if the variations of gravity g with height are ignored, the pressure p_h changes whenever there is a net change in the mass of air above h. This change of mass may be accomplished by the heating or cooling of the air without a change in the height of the free surface or by a change in the height of the free surface without the heating or cooling. Consider the simple analogy of a tank of water. If the water is cooled everywhere to the same extent there is a drop in the height of every pressure surface above the base of the tank. Now consider the case where one-half of the tank is cooled. Water flows into the cooled region in response to the horizontal pressure gradients created between the two sections by the cool-

ing and it follows that the pressure at the base of the cooled region rises, in view of the additional mass, while the pressure at the base of the other section falls, as a result of a lowering of the free surface. In a similar manner, it is necessary to heat or cool the atmosphere in one region relative to the remainder of the atmosphere in order to produce pressure changes. The pressure change depends upon the space variation of the heating or cooling and not on the magnitude and sign of the local temperature change. For example, pressure can fall in a region of cooling provided that the cooling is stronger elsewhere. This mechanism of pressure change may be referred to as a *thermal* theory. Various explanations have been offered for the source of local heating and cooling, such as the release of latent heat, the advection of warm or cold air, and the nonadiabatic processes. More information on the thermal theory will be presented in a later section.

An alternate method of producing a change in pressure, which involves no consideration of thermal processes, requires some outside agency to exert a force on the air. For example, a pressure drop may be produced in the center of a container of water by causing the container to rotate so that the drag at the walls sets the fluid in motion. Another illustration of this process is the pressure rise created at the bank of a stream where the water piles up at a bend. The question now arises as to whether pressure changes can occur in the atmosphere by an analogous mechanical process. The formation of high pressure on the windward side and of low pressure on the leeward side of a mountain range may be considered a mechanical process. However, there appear to be no other agencies which can produce significant pressure changes in the atmosphere by mechanical means.

The role of lateral mixing processes in the development of pressure changes has been the object of various studies. Rossby [26] analyzed the pressure changes which accompany the lateral diffusion of momentum in a straight current. It was concluded that the diffusion process could give rise to the formation of a low-pressure trough to the left and a high-pressure ridge to the right of the current. The pressure changes arise from a change in the height of the free surface. This study of the mutual adjustment of pressure and velocity distributions evidently was undertaken in connection with the interpretation of the so-called *dynamic* pressure systems, that is, the warm highs and cold lows. Such an explanation of pressure changes is open to criticism insofar as it leaves unexplained the manner in which the strong current was created in view of the continual operation of diffusion processes. Lateral mixing may be an adequate explanation of some pressure changes but to complete the picture it is necessary to explain the development of the situation which precedes the operation of the lateral mixing process. In his discussion of Jeffreys' theory [19] of atmospheric circulation, Whipple [30] offers a somewhat similar argument to explain the pressure distribution. The suggestion is made that the strong west winds aloft induce strong west winds below through turbulence and that air is flung outwards and creates anticyclonic belts. Whipple points out that the details of the mechanism are not clear.

There have also been a number of explanations of pressure changes which have been based upon the principle that the wind field changes and, therefore, the pressure field changes as a consequence of the mutual adjustment of pressure and wind distributions. The use of constant absolute vorticity trajectories [12] to predict a 10,000-ft pressure pattern is an example of this principle. In this case the question may be raised as to the applicability of the simplified vorticity equation. The final wind field is deduced on the basis of many assumptions concerning the behavior of individual air particles. There appears to be no sound reason why the pressure distribution should change so that the air particles can follow the assumed trajectories. Other explanations of pressure changes as a consequence of wind changes are to be found in meteorological literature. The statements are frequently vague without an explanation of how the pressure changes so that it is difficult to discuss the arguments. However, these explanations usually suffer from a defect like the assumption of a simplified air trajectory or an unexplained change in wind speed. This group of pressure-change theories is often referred to as a dynamic explanation of pressure changes.

A comprehensive discussion of dynamic and other theories of cyclogenesis has been presented by Raethjen [24]. The discussion is pertinent to the problem of the mechanism of pressure change since it considers the formation of a pressure minimum. Raethjen emphasizes the problem of the development of cyclonic and anticyclonic vorticity and includes an analysis of such dynamic theories of pressure change as the concept that the pressure field of a cyclone forms from its wind field. The significance of lateral mixing and friction as processes which influence the vorticity distribution is discussed in some detail by Raethjen.

The various approaches to the pressure-change mechanism may now be reviewed. The *thermal* theory appeals in view of its simple physical picture but presents the problem of how the local temperature changes are produced. The *dynamic* theories which consider that the pressure-change field is a result of a changed wind field are not altogether consistent with the fact that wind changes arise from the accelerations which accompany local pressure changes. Of course, it is possible to determine the average pressure force acting on a particle if the velocity of the particle is known at two different times. However, it does not follow that the change in wind velocity can be considered the cause of pressure changes. Rather it appears more logical to view the changes in the wind field as a consequence of pressure changes. Nevertheless the wind field, in particular the vorticity distribution, is a vital feature of a pressure system. Consequently any theory of pressure change, such as a thermal theory, must include an explanation of the observed changes in the wind field.

Pressure Tendency Equations

In recent years the problem of pressure changes has often been approached from the standpoint of analyzing the components of various equations for the pressure variation $\partial p/\partial t$. Bjerknes [2] discussed the development of pressure changes by use of the following equation for the pressure change at a level h.

$$\left(\frac{\partial p}{\partial t}\right)_h = -\int_h^\infty g\left(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y}\right)dz - \int_h^\infty g\rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)dz + g(\rho w)_h, \tag{2}$$

where u, v, and w are the x-, y-, and z-components of the wind. Equation (2) states that the pressure change is determined by the horizontal advection and divergence above the level h and by the vertical advection at h. Sufficient evidence has been accumulated with the advective integral [15] to show that it is of the same order of magnitude as the pressure tendency. On the other hand the divergence integral and the vertical motion term are usually one order of magnitude greater than the pressure tendency. It follows then that the contribution of the last two terms of equation (2) is the small difference between two large quantities. This difference is expressed analytically when the tendency equation is written in the form given by Houghton and Austin [18]:

$$\left(\frac{\partial p}{\partial t}\right)_h = -\int_h^\infty g\left(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y}\right)dz + \int_h^\infty gw\left(\frac{d\rho}{dz} - \frac{\partial \rho}{\partial z}\right)dz. \tag{3}$$

This tendency equation shows that the field of horizontal divergence and vertical motion gives rise to a pressure change depending upon the manner in which the motion changes the density distribution above the level h.

Alternate forms of the tendency equation have been derived in various ways. The reader is referred to a review given by Godson [13]. Another example of a tendency equation which does not involve the integration to infinity is given by Matthewman [20]. In recent years, with the availability of upper-air data, there have been attempts to compute the values of the factors which give rise to pressure changes. For example, Fleagle [11] has given values of the components of equation (4) at various levels in the atmosphere.

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t} = \frac{1-K}{p}\frac{\partial p}{\partial t} + \frac{1}{\theta}\mathbf{V}_h\cdot\nabla_h\theta + \frac{w}{\theta}\frac{\partial \theta}{\partial z} - \frac{1}{\theta}\frac{d\theta}{dt}. \tag{4}$$

$K = R/c_p$, R is the gas constant for dry air, c_p is the specific heat at constant pressure, θ is the potential temperature, $\mathbf{V}_h$ is the horizontal velocity vector, and $\nabla_h\theta$ is the horizontal gradient of potential temperature. The significance of these terms is apparent when it is noted that

$$\left(\frac{\partial p}{\partial t}\right)_0 = -\int_0^\infty \frac{1}{\rho}\frac{\partial \rho}{\partial t}\,dp. \tag{5}$$

Fleagle has also presented values of the horizontal mass divergence.

The observational evidence of Fleagle and others indicates that all the components of the tendency equation are important, that all layers of the atmosphere contribute to the various integrals, and that the net value of an integral is often the small difference between large values of opposite signs. For this reason it is not possible to compute the pressure tendency accurately from the components of a tendency equation.

The utility of these tendency equations will now be considered. A relationship as expressed by equations (2)–(5) gives an insight into the mechanism of pressure changes only as long as there is a clear understanding of how the components of the equations change. All the tendency equations involve the three-dimensional field of motion and it can readily be shown that, for the computation of the pressure variation, the velocity must be known to an accuracy of the order of 1 cm sec^{-1}. Consequently a tendency equation is helpful in understanding pressure changes provided that the mechanism of wind change is known. This involves a knowledge of the manner in which accelerational fields develop in the atmosphere, and undoubtedly the local pressure changes themselves are of major importance in producing accelerations. From such considerations as the above it would appear that the tendency equations do not contribute significantly to an understanding of the mechanism of pressure changes. Computations of the magnitudes of the various terms do indicate, however, the importance of the various processes which accompany pressure changes. These computations show what is happening in the atmosphere while the pressure changes. For this reason it is necessary that the empirical work be extended as it is probable that the vertical distribution of such quantities as the horizontal divergence differs from one part of an isallobaric system to another. Finally it might be noted that these tendency equations have no prognostic value insofar as it is not possible to predict the values of the various integrals of the tendency equations.

A number of theoretical attempts have been made to analyze the pressure-change mechanism by means of a tendency equation and an assumed wind field. For example, Bjerknes and Holmboe [3] discuss the distribution of horizontal convergence and divergence at various levels in the atmosphere when it is assumed that the wind is gradient at trough lines and wedge lines. Priestley [22] also analyzes pressure changes on the assumption of gradient flow and stresses the importance of the path curvature on the control of atmospheric pressure. Schmidt [27] discusses the components of a tendency equation through the introduction of an approximation concerning the acceleration of the wind. These approaches are subject to criticism on account of the assumptions concerning the wind field. Houghton and Austin [18] and Priestley [23] discuss aspects of the

effect of the deviations from the gradient wind upon the surface pressure field. It seems evident that only limited conclusions concerning the pressure-change process can be drawn from an analysis which specifies a particular type of wind field. Nevertheless this type of approach gives valuable information on the importance of the various parts of the wind field and, therefore, aids in establishing an internally consistent picture of the field of motion with a field of pressure change.

Vorticity Studies

The density changes which lead to a change in mass in an atmospheric column invariably involve the wind field as illustrated by equation (2). A transformation of the horizontal equations of motion leads, with a few simplifying assumptions, to a vorticity equation

$$\mathrm{div}_2 \mathbf{V} = -\frac{1}{\zeta + \lambda}\frac{d}{dt}(\zeta + \lambda), \qquad (6)$$

where $\mathbf{V}$, ζ, and λ are the horizontal velocity vector, the vertical component of vorticity, and the Coriolis parameter, respectively. Although equation (6) does not directly refer to pressure change, it plays an important role in pressure-change research because it expresses certain relationships which must be satisfied by the wind field. For example, if a hypothesis specifies horizontal divergence in the upper troposphere, then vorticity changes consistent with equation (6) should occur in this region.

Consequently, such a vorticity equation and the pressure tendency equations together provide a means for testing the validity of a theory of pressure change. However, like the pressure tendency equations, equation (6) does not lead directly to an explanation of the mechanism of pressure change.

An ever-present difficulty in all wind studies is the lack of a dense network of reliable wind observations, particularly above 10,000 ft. Consequently, most studies of vorticity change have been based upon some assumed wind field such as a geostrophic wind field [28]. In all likelihood, the vorticity of the geostrophic wind differs from the vorticity of the actual wind and this difference may well be large in a region of active pressure change. Since equation (6) is a powerful tool for testing pressure change theories it seems desirable that additional research be conducted on the vorticity field and its changes.

Thermal Theory of Pressure Change

Several aspects of a thermal explanation of pressure change will now be considered. This elaboration of a thermal approach is undertaken because the theory has the advantage of a simple physical picture and because it appears to be as well developed as other theories of pressure change.

Nonadiabatic Temperature Changes. The significant aspects of the thermal theory of pressure change may be demonstrated by means of a simple model. Consider a mass of air above the earth's surface, *BCNM* in Fig. 1*a*, and let the air be heated uniformly from 1000 mb to 600 mb. As a consequence of the heating all the pressure surfaces above *BC* are raised. This expansion is not accompanied by adiabatic temperature changes insofar as there is no change in pressure on individual air particles. The heating then creates a pressure gradient to accelerate air out of the column above *BC*. The vertical displacement of each pressure surface is given by

$$\frac{\delta Z}{Z} = \frac{\delta T}{T}, \qquad (7)$$

where Z is height of the surface and T is the mean temperature from the earth's surface to the level Z. It follows that δZ increases from 0 for the 1000-mb surface to a maximum at 600 mb and that all pressure surfaces above 600 mb are raised by the same amount as the 600-mb surface. The new horizontal pressure field then creates a maximum mass outflow about 600 mb. In response to this outflow the pressure falls near the base of the column *BC* and upward accelerations are created

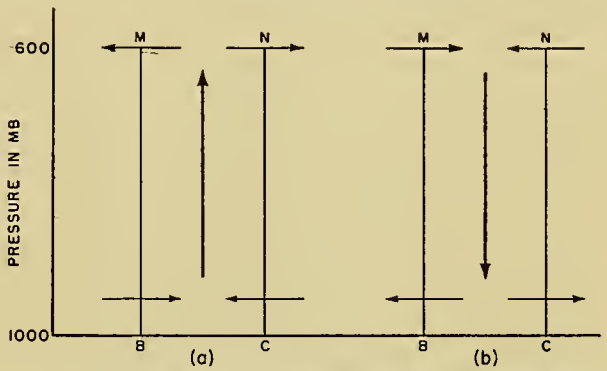


FIG. 1.—Two atmospheric models with (*a*) heating and (*b*) cooling in the region *BCMN*. The arrows indicate the fields of horizontal divergence and vertical motion.

within *BCNM*. As a consequence of these pressure falls there are inward horizontal accelerations near the base of the column. Even though this analysis has treated the problem in a stepwise manner, it is recognized that the process is a continuous one.

The heating then produces a pressure fall at *BC* and a pressure rise in the environment. One modifying influence is the adiabatic temperature changes which accompany the vertical displacements after the establishment of the accelerational field. These adiabatic changes modify the original temperature difference between the heated column and its environment, and the pressure difference at any given time depends upon the net contribution of the heating and cooling processes to the establishment of a temperature difference between the column *BC* and its environment. The wind field at the same time may be determined from the accelerational field created by the heating, from the Coriolis acceleration, and from the effect of friction. The well-known low-level convergence and high-level divergence in the vicinity of centers of pressure fall are readily deduced.

Similar reasoning could be applied to the converse problem of the effect of local cooling. It is apparent from the previous discussion that cooling gives a pressure rise at *BC*, a pressure fall in the environment, and a divergence field like that of Fig. 1*b*.

It is now necessary to investigate the magnitudes of the changes produced by heating and cooling. If the air column *BCNM* were heated 5C relative to the environment, the magnitude of the sea-level pressure change would be somewhat less than the change which would arise if the height of the 600-mb surface should remain invariant and the air column below 600 mb were replaced by a 5C-warmer column, that is, less than 9 mb. The reasons for the smaller change are the raising of the 600-mb surface and the distribution of the outflowing air over a finite region. The extent of the surroundings also determines the sea-level pressure rise in the surroundings. It does appear, therefore, that heating and cooling can produce significant pressure changes and that nonadiabatic processes may be important for the development of pressure fields.

A brief survey of various types of nonadiabatic processes will now be presented. The atmosphere is continually gaining and losing heat by radiation. However, radiation studies indicate that the net loss of heat by the tropospheric air is small and that the space variation of this loss is slight. It appears that the cooling is greater in regions of high specific humidity and consequently direct radiative processes in the troposphere may produce weak pressure fields. The space variation of radiation absorption in the stratosphere, such as in the ozone region, should give rise to high-level accelerational and pressure-change fields (see Haurwitz [16]) However, it is not obvious that these fields have a pronounced effect on the sea-level circulation. A significant difference between low- and high-level heating is the absence of a fixed boundary at the base of the heated layer in the case of the high-level heating. Before it may be concluded that such heating can produce substantial sea-level changes it is necessary to investigate the possible significance of adiabatic processes in the troposphere below the level of heating. Adiabatic cooling can account for a definite pressure fall at high levels and an insignificant sea-level change. Hence the answer to the question of the effect of radiation absorption in the stratosphere upon low-level changes must await the results of further theoretical and empirical studies.

Nonadiabatic temperature changes also arise from the flux of heat between the atmosphere and the earth's surface. Because this heating or cooling is not uniform over the entire globe it should be expected that this thermal process would be accompanied by accelerational fields and pressure changes. The significance of this factor may be illustrated by a few simple examples.

1. The development of low pressure in a region of strong heating is often observed on weather charts. These cyclones are usually referred to as thermal lows.

2. The sea- and land-breeze circulations are good examples of the creation of a small-scale pressure-change field through surface heating and cooling.

3. The mean pressure maps of the Northern Hemisphere demonstrate that the average sea-level pressure over the hemisphere is approximately 4 mb lower in July than in January. This appreciable pressure change between summer and winter is consistent with the heating of one hemisphere relative to the other.

4. Perhaps the most striking pressure changes are those which occur in middle latitudes between land and water areas from summer to winter. In winter, air is being cooled over the land and air from the land is being heated over the water. The converse holds during the summer season. This distribution of heating and cooling requires a pressure rise over the land from summer to winter and a pressure fall over the ocean. From mean pressure and temperature charts it has been possible to determine approximately the pressure and temperature differences. The data in Table I clearly indicate that substantial pressure changes accompany the heating and cooling.

Table I. Seasonal Variation in Average Surface-Air Temperature and Average Sea-Level Pressure

	Lat. 40°N		Lat. 50°N	
	Land	Ocean	Land	Ocean
January				
Pressure (mb)	1023	1015	1022	1006
Temperature (°F)	36	48	10	37
July				
Pressure (mb)	1011	1020	1011	1016
Temperature (°F)	82	67	69	55

5. Wexler [29] has explained the development of a polar anticyclone through cooling and its associated isallobaric convergence. The analysis of the pressure change which accompanies the heating or cooling has been extended by Schmidt [27]. Schmidt assumes a certain distribution of the heating and computes the pressure changes from a consideration of radiative density variations, isallobaric divergence, and the vertical motion of the air. The theoretical formula is tested by a study of the seasonal pressure variations over the vast continents. The theoretical estimates are in good agreement with the observed pressure changes from summer to winter.

6. Estimates [4] have been made of the rate of addition of heat to cold polar air as it leaves the mid-Atlantic coast and moves out over the ocean. This heating can be very pronounced and it would appear that this process alone could produce substantial pressure changes since there is a space variation in the heating. It should be noted, however, that it is difficult to check the importance of the heating directly since a strong advection of cold air is usually associated with a pressure rise (see below). This type of strong heating occurs near the Atlantic coast of North America and the Pacific coast of Asia.

The survey suggests that nonadiabatic heating or cooling of air, through contact with the earth's surface, may produce significant pressure changes. The horizontal mixing of different air masses and the cooling or heating of air by falling precipitation are other examples of nonadiabatic processes which could give rise to pressure changes.

The conclusion which is drawn from this review of the effect of nonadiabatic processes is that they are of a sufficient magnitude to influence the daily pressure

patterns and to be detected in the field of pressure changes. For the day-to-day pressure variations, except for the diurnal change, it would appear that the space variation of the transport of air over colder and warmer surfaces is the most important of the nonadiabatic processes. A satisfactory explanation of the development, intensification, and motion of the fields of acceleration and pressure change must be based, in part, upon the effect of nonadiabatic temperature changes.

Horizontal Advection. Local changes in temperature in one region relative to another may arise through the horizontal advection of warm or cold air. The concept of thermal advection as a cause of pressure changes has been discussed by many meteorologists since Ferrel [8] advanced a thermal theory of pressure changes. For example, Exner [7], Henry [17], Defant [5], and McDonald [21] have illustrated relationships between temperature advection at the surface and the simultaneous pressure change. Austin [1] has shown that regions of maximum advection of warm air near the earth's surface are accompanied by pressure falls at sea level while pressure rises occur in regions of maximum advection of cold air. It was further demonstrated that advection at high levels did not appear to be associated with significant pressure changes at sea level.

Even though prominent relationships have been established between the advection of cold and warm air in the lower troposphere and the occurrence of pressure rises and falls, it is still necessary to explain the mechanism of the pressure change. Perhaps the process may be visualized as the horizontal motion of warm air into a region of cold air giving an outflow at high levels and, therefore, a pressure fall just as the direct heating in Fig. 1 gives rise to a pressure fall. The converse is the transport of cold air into a warm region with high-level inflow and a pressure rise at sea level. Such an explanation replaces the direct heating or cooling of the thermal model in Fig. 1 with the localized heating and cooling which arises from horizontal transport. The process has been described by Douglas [6] as "a convectional overturning between adjacent cold and warm masses." The similarities between the nonadiabatic and advection hypotheses may be summarized as follows:

1. Both types of heating and cooling give rise to sea-level pressure changes when the temperature change extends upward from the earth's surface.

2. High-level temperature changes produced by nonadiabatic processes or indicated by geostrophic advection are not necessarily associated with prominent sea-level pressure changes.

3. The accelerational fields which arise from nonadiabatic processes appear to be similar to those which accompany the differential advection of warm or cold air.

4. Both types of heating or cooling give rise to pressure-change fields only as long as there is a space variation in the heating and cooling.

Two important differences may be stated briefly as follows:

1. The nonadiabatic process gives rise to the pressure change and accelerational field from an initially barotropic state. The advective process requires a particular field of motion relative to the temperature field.

2. The nonadiabatic heat source is stationary whereas the advective heating is a moving source. This distinction is important in problems concerning the pattern of horizontal divergence and vertical motion with moving pressure-change systems.

At this stage it is desirable to consider those aspects of atmospheric pressure change which can be attributed to the differential advection of temperature at low levels.

The major part of a pressure fall or rise in a concentrated region of large pressure changes at sea level can be explained by the advection of warm or cold air in the lower atmosphere. Also many features of the propagation of pressure-change centers may be attributed to this advection process. For example, consider a simple model, as in Fig. 2. It will be assumed that the essential

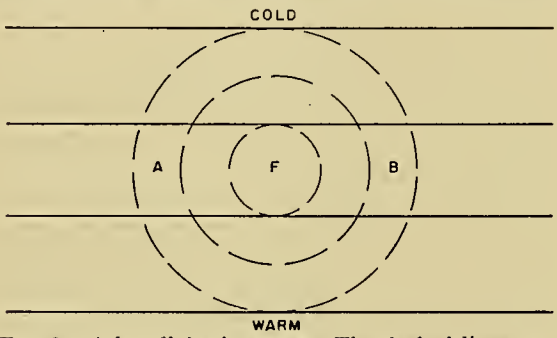


FIG. 2.—A katallobaric system. The dashed lines are katallobars and the solid lines are isotherms.

features of the horizontal advection can be judged from the geostrophic flow even though it is recognized that the motion cannot be geostrophic. The katallobaric system is associated with a strong advection of warm air and consequently it should be expected to move in the direction of the flow, that is, in the direction of the geostrophic wind. However, the isallobars show that the geostrophic field is changing so as to intensify the warm-air advection at B and to produce cold-air advection at A. This changing field of horizontal motion itself gives rise to a motion of the katallobaric center in the direction AB. Hence it should follow that the actual motion of the pressure-change center is in a direction somewhere between the direction AB and the existing geostrophic direction at F. Synoptic studies appear to confirm this conclusion. On the other hand this differential advection, as judged by geostrophic flow, cannot account for the development of new isallobaric systems or for the intensification or weakening of existing isallobaric centers.

One aspect of the motion which requires further investigation in connection with the development of pressure changes at sea level is the role played by ground friction. It seems probable that the retarding effect of friction influences the low-level inflow and outflow which follow the creation of isallobaric fields. This retarding force may affect the ultimate pressure change, and consequently surface friction requires further consideration.

High-Level Processes

A European school of meteorologists proposed a modified thermal theory [9] of pressure change in which the primary cause is thought to be temperature advection in the stratosphere with secondary thermal effects in the troposphere. There is no doubt that there are strong advective fields about the tropopause and, therefore, it seems logical to pay attention to the influence of stratospheric advection on pressure changes. A comprehensive analysis of the large changes which take place near the tropopause has been presented by Fleagle [10]. However, as Fleagle mentions, these changes are not sufficient evidence to conclude that the "seat" of the pressure change resides in the stratosphere. The evidence which has been presented on the relationships between sea-level pressure change and high-level advection do not support a conclusion that the cause of the sea-level change is to be found in advection in the stratosphere.

The strong changes which occur in the upper atmosphere may be considered as arising from vertical motion fields which accompany the sea-level pressure changes and which arise from such low-level processes as heating or cooling. A nonuniform field of vertical motion creates strong advective fields in the lower stratosphere as a result of the abrupt change in the vertical temperature gradient near the tropopause. Hence it may be argued that the high-level advective fields are the result rather than the cause of sea-level pressure changes. Similar explanations have been presented by many meteorologists such as the discussion given by Douglas [6] on "Some Facts and Theories about the Upper Atmosphere." It should also be noted in connection with the high-level processes that Reed [25] has been successful in relating the prominent ozone variations to the fields of vertical motion which are known to accompany the sea-level pressure system.

It would appear then that there is no definite evidence to support a hypothesis that stratospheric advection may be considered a cause of sea-level pressure change. However the evidence does not contradict hypotheses that other high-level processes may produce low-level pressure changes. For example, the observed variations of the temperature field across the tropopause may lead one to suspect that instability in this general region may give rise to significant low-level changes. It would then appear desirable to investigate, analytically and empirically, processes other than advection which might take place in the vicinity of the tropopause and contribute to sea-level changes. As in the case of the high-level heating discussed previously, it will be necessary to consider the effect of tropospheric processes upon the net sea-level change which arises from some high-level process.

High-Level Pressure Changes

The motion and changes in intensity of the weather systems are accompanied by pressure changes at all levels in the atmosphere. The various constant-level charts present a series of horizontal slices through the atmosphere whose over-all behavior is reflected by the sea-level chart.

So far the discussion has concentrated mainly upon the problem of the pressure change at sea level. This approach is justified on the basis of the presence of a fixed boundary which should simplify the problem. When high-level pressure changes are considered it has to be recognized that air can move vertically through the level at which the pressure variation is being analyzed. One method of viewing a high-level pressure change is to explain the change on the basis of the pressure variation at the earth's surface and the change in the temperature or density of the air column from the surface to the level in question. This type of approach is adequate in view of the status of pressure-change theories.

Empirical Evidence

Because of the questionable status of the theories of pressure change it is desirable to investigate those empirical facts which have to be explained by a pressure-change theory. Figure 3 presents a schematic cross section through an area of pressure rise and fall and includes some pertinent information on the changes in the temperature field, the vertical velocities, and the vertical accelerations which accompany the pressure changes. The idealized picture is based upon observational evidence obtained by the pressure-change project at the Massachusetts Institute of Technology and upon the empirical data presented by Fleagle [10, 11]. Whether Fig. 3 is a good approximation to a cross section through an actual weather system is open to question. Since upper-air observations are usually taken at twelve-hour intervals, much of the empirical information on the instantaneous field of vertical velocities has to be deduced from twelve-hour changes. For this reason it is to be expected that a computed field may give a somewhat erroneous impression of the magnitude and space variation of the true field of vertical velocities.

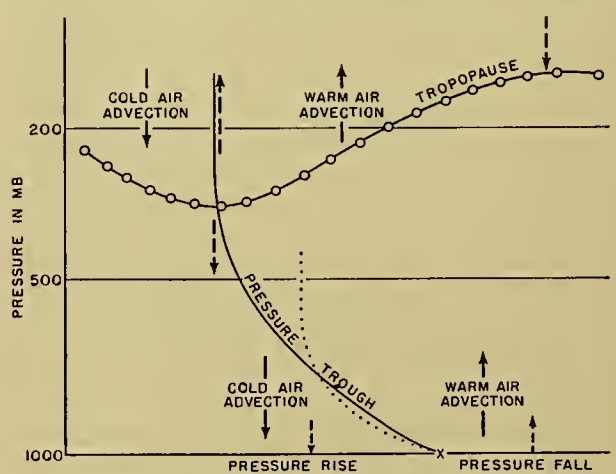


Fig. 3.—A vertical cross section through an idealized pressure trough. The solid and broken lines indicate the directions of the vertical velocities and vertical accelerations, respectively. The dotted line indicates the region of strong horizontal temperature gradient.

The vertical accelerations are the most questionable of the observational data. However, it is believed that the comparatively slow displacement of the high-level field of vertical velocities as compared with the large horizontal velocities makes it possible to estimate the magnitude of the vertical accelerations in the upper atmosphere. The vertical accelerations in Fig. 3 were estimated from such considerations. This idealized model serves to illustrate those features of pressure changes which have been discussed earlier. Two other features of this diagram deserve attention.

1. The pressure trough and the zone of maximum temperature contrast are approximately coincident in the lower atmosphere, consistent with the general concept of a frontal surface. However, above the middle troposphere this coincidence is no longer present. The pressure troughs and ridges in the higher atmosphere are near regions of maxima and minima in the temperature field. The pressure systems move along with the low-level temperature field at a speed which is not vastly different from the horizontal wind speeds. Above the lower troposphere, however, the prominent features of the temperature and pressure field move much slower than the horizontal wind speeds at the respective levels. This behavior of the higher atmosphere may suggest that the primary cause of pressure changes is to be found in the lower troposphere. However, before such a conclusion can be reached it is necessary to explain the origin of the vertical accelerations at high levels.

2. A developing pressure system is usually associated with an intensification of the high-level pressure trough and the associated temperature maximum in the stratosphere. Figure 3 shows that this maximum arises from descending motion upstream from the point of maximum temperature. Consequently a developing pressure system must be accompanied by downward accelerations many hundreds of miles upstream from the stratospheric temperature maximum. This evidence on the high-level accelerational field indicates that the development of a pressure system is accompanied by significant changes over a wide area.

Hence it appears that the problem of pressure changes cannot be localized to processes within the immediate vicinity of an isallobaric system.

Conclusions

It has been shown that there have been various approaches to the problem of the mechanism of pressure change. At present the physical picture is incomplete and, therefore, no definite conclusion can be reached as to the most desirable line of approach. It is evident that a theory must explain an increase or decrease of mass in an air column. In this review the emphasis has been placed upon the thermal explanation of pressure changes since it appears to satisfy many of the observed facts concerning the temperature field. Moreover, a thermal approach is in accord with the general concept that heating and cooling are the primary causes of atmospheric motion. However, it is apparent that the thermal theory has not been fully developed. Many problems remain:

1. Theoretical analyses of the development of divergence and vertical-motion fields with pressure changes appear to be inadequate.

2. The empirical picture of divergence and vertical motion is well established for stationary heat sources; however, it is evident that there are gaps in our knowledge concerning the manner in which divergence and vertical-motion fields develop and move with migratory pressure systems. Much of the empirical information is based upon twelve-hour changes and it is possible that such data give an incorrect impression of the motion in areas of pressure change.

3. The direct influence of nonadiabatic temperature changes on pressure changes requires further consideration.

4. The details of horizontal advection require more investigation in order to ascertain whether advection should be considered a cause or an effect of pressure change.

5. The complexity of the temperature field in the vicinity of the tropopause warrants further consideration of the stability of this field.

6. The influence of surface friction must be further investigated.

7. It is necessary to develop a thermal model which gives a pressure change in a quiescent region, followed by a moving pressure-change field with its associated changes at various levels, and then the end stage of a cold cyclone or warm anticyclone.

Finally, it is important to note that the empirical data indicate that the explanation of pressure changes cannot be localized to processes taking place over a small area. This concept of the extent of the field of influence applies to all analyses of the problem of pressure change. For this reason there is some advantage in specifying a certain unbalanced condition as an initial state in a region where the pressure is changing and later investigating how such a particular initial state might have originated. Such a procedure, which has been followed in some dynamic approaches, might well be adopted in a thermal or convectional approach.

Even though the pressure-tendency equations do not appear to lead to fundamental explanations, they play an important role in research on pressure change. A preliminary step toward an explanation of pressure change is an accurate description of all features of the wind, pressure, and temperature fields in the vicinity of pressure changes. The pressure tendency equations make it possible to check the internal consistency of such pictures. In connection with this important aspect of pressure-change research the author would like to suggest the desirability of utilizing to a greater degree other relationships which may be derived from the equations of motion. For example, vorticity-divergence relationships should also be helpful in determining the reality of an empirical model of a stationary or moving pressure-change system.

This breakdown of approaches to the pressure-change mechanism into the two categories of *thermal* and *dynamic* is perhaps artificial. The classification is consistent with the conventional use of the adjectives thermal and

dynamic. However, it is apparent that *thermal* hypotheses are concerned with the laws governing the air in motion and that a thermal approach must not overlook the necessity of explaining the particular changes in the wind field which accompany pressure changes. Likewise *dynamic* hypotheses recognize the thermal origin of atmospheric motion and a dynamic approach cannot disregard the observed temperature field. The ultimate goal of a complete explanation of pressure change involves the description of processes which satisfy all the laws governing the behavior of the atmosphere. In an attempt to advance toward this goal it may be convenient to reason from one or more of the fundamental laws. Hence all approaches which have physical bases, whether they be called dynamic or thermal, supplement rather than compete with one another. The verified conclusions of one approach should aid the extension of a different line of attack.

REFERENCES

1. Austin, J. M., "Temperature Advection and Pressure Changes." *J. Meteor.*, 6: 358–360 (1949).
2. Bjerknes, J., "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54: 462–466 (1937).
3. —— and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
4. Burke, C. J., "Transformation of Polar Continental Air to Polar Maritime Air." *J. Meteor.*, 2: 94–113 (1945).
5. Defant, A., *Wetter und Wettervorhersage*. Zweite Auflage, Leipzig, F. Deuticke, 1926.
6. Douglas, C. K. M., "Some Facts and Theories about the Upper Atmosphere." *Quart. J. R. meteor. Soc.*, 61: 53–71 (1935).
7. Exner, F., "Grundzüge einer Theorie der synoptischen Luftdruckveränderungen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 115: 1171–1246 (1906).
8. Ferrel, W., *A Popular Treatise on the Winds*. New York, Wiley, 1889.
9. Ficker, H. v., "Beziehungen zwischen Änderungen des Luftdruckes und der Temperatur in den unteren Schichten der Troposphäre (Zusammensetzung der Depressionen)." *S. B. Akad. Wiss. Wien*, Abt. IIa, 129: 763–810 (1920).
10. Fleagle, R. G., "The Fields of Temperature, Pressure and Three-Dimensional Motion in Selected Weather Situations." *J. Meteor.*, 4: 165–185 (1947).
11. —— "Quantitative Analysis of Factors Influencing Pressure Change." *J. Meteor.*, 5: 281–292 (1948).
12. Fultz, D., "Upper Air Trajectories and Weather Forecasting." *Dept. Meteor. Univ. Chicago, Misc. Rep.* No. 19 (1945).
13. Godson, W. L., "A New Tendency Equation and Its Application to the Analysis of Surface Pressure Changes." *J. Meteor.*, 5: 227–235 (1948).
14. Hann, J. v., *Lehrbuch der Meteorologie*. Leipzig, Tauchnitz, 1901.
15. Haurwitz, B., and Collaborators, "Advection of Air and the Forecasting of Pressure Changes." *J. Meteor.*, 2: 83–93 (1945).
16. Haurwitz, B., "Relations between Solar Activity and the Lower Atmosphere." *Trans. Amer. geophys. Un.*, 27: 161–163 (1946).
17. Henry, A. J., and Collaborators, *Weather Forecasting in the United States*. Washington, D. C., U. S. Govt. Printing Office, 1916.
18. Houghton, H. G., and Austin, J. M., "A Study of Nongeostrophic Flow with Applications to the Mechanism of Pressure Changes." *J. Meteor.*, 3: 57–77 (1946).
19. Jeffreys, H., "On the Dynamics of Geostrophic Winds." *Quart. J. R. meteor. Soc.*, 52: 85–104 (1926).
20. Matthewman, A. G., "A Discussion of the Pressure-Tendencies Associated with Gradient and Horizontal Geostrophic Flow. A Formula for the Variation with Height of the Vertical Velocity." *Phil. Mag.*, 37: 706–716 (1946).
21. McDonald, W. F., "Notes on a Method of Analysis of the Data of Forecasting." Unpublished (1929).
22. Priestley, C. H. B., "Dynamical Control of Atmospheric Pressure." *Quart. J. R. meteor. Soc.*, 73: 65–84 (1947).
23. —— "Atmospheric Pressure Changes: The Importance of Deviations from the Balanced (Gradient) Wind." *Aust. J. sci. Res.*, 1: 41–57 (1948).
24. Raethjen, P., "Zyklogenetische Probleme." *Arch. Meteor. Geophys. Biokl.*, (A) 1: 295–346 (1949).
25. Reed, R. J., *The Effects of Atmospheric Circulation on Ozone Distribution and Variations*. Sc.D. Thesis, Mass. Inst. Tech., 1949.
26. Rossby, C.-G., "On the Mutual Adjustment of Pressure and Velocity Distributions in Certain Simple Current Systems." *J. mar. Res.*, 1: 15–28 (1937).
27. Schmidt, F. H., "On the Causes of Pressure Variations at the Ground." *Meded. ned. meteor. Inst.*, (B), Deel 1, Nr. 4 (1946).
28. Sutcliffe, R. C., "A Contribution to the Problem of Development." *Quart. J. R. meteor. Soc.*, 73: 370–383 (1947).
29. Wexler, H., "Formation of Polar Anticyclones." *Mon. Wea. Rev. Wash.*, 65: 229–236 (1937).
30. Whipple, F. J. W., "A Note on Dr. Jeffrey's Theory of Atmospheric Circulation." *Quart. J. R. meteor. Soc.*, 52: 332–333 (1926).

LARGE-SCALE VERTICAL VELOCITY AND DIVERGENCE

By H. A. PANOFSKY

New York University

Influence of Scale on Divergence and Vertical Velocity

The vertical velocity is the only component of the air velocity vector not generally recorded. Yet it produces important effects of turbulence and is basic to the formation of precipitation. On a small horizontal scale it has been measured directly in thunderstorm and turbulence research, but the average vertical velocity over large areas (10^{14} cm^2 or more) must be estimated indirectly.

The horizontal velocity divergence, div V (called simply *divergence* in this article), is related to the vertical velocity by the equation of continuity. According to this equation, vertical divergence $\partial w/\partial z$ is almost exactly equal in magnitude and opposite in sign to horizontal divergence, provided that the fractional change of density is small. The distribution of horizontal divergence therefore prescribes the distribution of $\partial w/\partial z$. If the vertical velocity w is known at any one level, the distribution of horizontal divergence determines the distribution of the vertical velocity itself. Generally, the vertical velocity can be assumed to be known at the ground. Because of this close relationship between vertical velocity and horizontal divergence, the two quantities are treated together in this article.

The magnitudes of divergence and vertical velocities are influenced to an extreme degree by the size of the areas over which they are averaged. This can be seen most easily for the divergence. Let x and y be two horizontal Cartesian coordinates and u and v the velocity components in the x and y directions, respectively. Then the divergence is given by

$$\text{div V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}. \tag{1}$$

If this quantity is averaged over a square of length L, the result is

$$\overline{\text{div V}} = \frac{(\bar{u}_3 - \bar{u}_1) + (\bar{v}_4 - \bar{v}_2)}{L}, \tag{2}$$

where $\bar{u}_3$ is the mean value of u on side 3 of the square, and the other velocity terms have corresponding interpretations (see Fig. 1).

The magnitudes of $\bar{u}$ and $\bar{v}$ are almost independent of scale, and are of the order of 10 m sec^{-1}. Even differences of the horizontal velocities depend relatively little on scale; horizontal velocity differences of the order of 10 m sec^{-1} appear over 100 m as well as over 1000 km.

On all scales there is a tendency for $\bar{u}_3 - \bar{u}_1$ to be opposite in sign and almost equal in magnitude to $\bar{v}_4 - \bar{v}_2$ (*cf.* [5]). It turns out that the numerator of equation (2) is of the order of 1 m sec^{-1}, regardless of scale. Thus the order of magnitude of the divergence is $1/L$ sec^{-1} if L is measured in meters.

On the scale involved in the thunderstorm studies[1] L is of the order of 3000 m, so that actually div V should be near 3×10^{-4} sec^{-1}. The observed values are usually somewhat higher because thunderstorms occur at times of unusual vertical motion. On this scale, 3×10^{-4} sec^{-1} might be regarded as a "normal" order of magnitude of divergence.

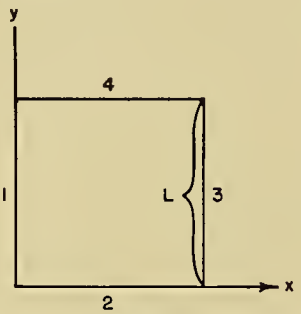


FIG. 1.—Illustration for definition of mean divergence in square area.

On the "weather-map scale" (10^5 m $< L < 10^6$ m) the divergence is between 10^{-5} sec^{-1} and 10^{-6} sec^{-1}. The remainder of this article will be particularly concerned with divergence of this order of magnitude.

On this weather-map scale we may express the equation of continuity in the form (assuming the surface horizontal):

$$w_h = -\frac{\bar{\rho}}{\rho_h} h \overline{\text{div V}}, \tag{3}$$

where w is vertical velocity, h is an arbitrary level, and $\bar{\rho}$ and $\overline{\text{div V}}$ represent density and divergence, respectively, averaged over height. If h is 3000 m and if there is no noticeable difference between the magnitudes of div V and $\overline{\text{div V}}$, the vertical velocity at weather-map scale should be between 0.3 cm sec^{-1} and 3 cm sec^{-1}. Again, vertical velocities much larger than these values appear with smaller scales. Byers finds vertical velocities averaging about 10 m sec^{-1} in thunderstorms.

Direct Measurement of Divergence

Since the divergence is defined in terms of horizontal winds, and since winds can be measured with reasonable accuracy, it is possible in principle to determine the divergence directly from equation (2). Generally, the pseudo-Cartesian coordinates of meteorology are employed in this procedure with x toward the east and y

1. Consult "Thunderstorms" by H. R. Byers, pp. 681–693 in this Compendium.

toward the north. In that case equation (1) has to be modified to

$$\text{div } \mathbf{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} - \frac{v}{R} \tan \phi, \qquad (4)$$

where R is the radius of the earth and ϕ is the latitude.

Usually, separate maps of u and v are constructed and analyzed, and the partial derivatives are approximated by ratios of finite differences. The differences of velocity components are taken over distances of the order of 100 miles. This method frequently leads to small, irregular patterns [13], the reality of which is doubtful.

A superior technique is based on equation (2). Here it is necessary to find averaged wind components along each side of a square. If the square is so small that only about one wind observation falls along each side of the square, the average wind along each side will be but poorly determined, particularly since the winds are strongly affected by local eddies. In practice, values of divergence can be trusted only if they have been determined from squares with sides 500 km long or longer. Even then, the measurements lead to doubtful results unless the wind coverage is complete in an area somewhat larger than the square.

Another technique of measuring horizontal divergence is based on the expression of mean divergence in "natural" coordinates instead of in Cartesian coordinates:

$$\overline{\text{div } \mathbf{V}} = \frac{\bar{V}_2 S_2 - \bar{V}_1 S_1}{A},$$

where S_1 and S_2 are the distances between two streamlines, measured along orthogonals to the streamlines; $\bar{V}_2$ and $\bar{V}_1$ are the average velocities of outflow and inflow across S_1 and S_2; and A is the area enclosed by the two bounding streamlines, S_1 and S_2 (Fig. 2).

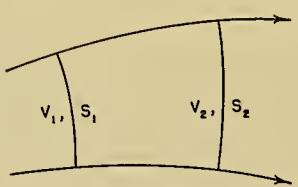


FIG. 2.—Illustration for definition of mean divergence in natural coordinates.

Mantis (see [16]) showed that both the "component technique" (Cartesian coordinates) and the streamline technique (natural coordinates) are about equally inaccurate; even for large areas and good wind coverage the sign of the divergence can be trusted only for relatively large observed magnitudes, for example, greater than 5×10^{-6} sec^{-1}. Recently, an objective method proposed by Bellamy [2] has been used extensively. This procedure is based on a determination of divergence from observations at the corners of a triangle. (For another, much more tedious, objective method see [21].)

Determination of Vertical Velocities

Direct Measurements of Vertical Velocities. Vertical velocities can be measured directly when they are larger than 10 cm sec^{-1}. Some of these methods of measurement are described in Byers' article on thunderstorms.[1] In addition, instantaneous vertical velocities near the ground have been measured by many investigators of turbulence.

Large-scale vertical velocities (horizontal area $> 10^{14}$ cm^2) must be determined indirectly. The two principal methods which have been suggested for estimating these small vertical velocities involve computations (1) from horizontal velocities and the equation of continuity (kinematic method), and (2) from the effect of vertical velocities on temperature, potential temperature, or wet-bulb potential temperature (adiabatic method).

The Kinematic Method of Determining Vertical Velocities. On the scale under discussion here, the vertical velocity at level h can be found from the equation of continuity in the form

$$w_h = -\frac{1}{\rho_h} \int_s^h \rho \text{ div } \mathbf{V} \, dz + w_s \frac{\rho_s}{\rho_h}. \qquad (5)^2$$

Here s stands for surface. The integral can be approximated by a sum, and w_s can be estimated from the horizontal wind field and the slope of the terrain.

A more elegant form of this method is obtained if equation (5) is transformed to

$$w_h = -\frac{\bar{\rho}}{\rho_h} \text{ div} \int_s^h \mathbf{V} \, dz. \qquad (6)$$

The vector $\int_s^h \mathbf{V} \, dz$ is the "resultant vector" which is equal to the horizontal distance from the observer to the position of the pilot balloon at level h, multiplied by the speed of ascent of the balloon. This distance is available directly from the original pilot-balloon runs. The divergence of the resultant vector is computed by one of the techniques described above. The kinematic method yields instantaneous values of vertical velocities, usually averaged over considerable areas. A dense network of actual wind observations is needed, so vertical velocities have been computed by this method only up to 10,000 ft and in regions of good weather. Radar and radio direction-finder pilot balloons should increase the range and applicability of the method.

The Adiabatic Method of Computing Vertical Velocities. The adiabatic method, which is completely independent of the kinematic method, is based on the assumption that rising and sinking air will cool or warm at a known rate. Then a measured temperature change will permit computation of the vertical velocity.

We start from the mathematical identity

$$\frac{dT}{dt} = \frac{\partial T}{\partial t} + \mathbf{V} \cdot \nabla_H T + w \frac{\partial T}{\partial z},$$

where T is temperature and ∇_H is the vector differential

2. Equation (3) is essentially the same as equation (5), but applies only for horizontal ground.

operator applied in the horizontal direction. If temperature changes are adiabatic, dT/dt is given by $-w\gamma_{ad}$ where γ_{ad} is the adiabatic lapse rate. Then

$$w = -\frac{\frac{\partial T}{\partial t} + \mathbf{V}\cdot\nabla_H T}{\gamma_{ad} - \gamma} = -\frac{\frac{\delta T}{\delta t}}{\gamma_{ad} - \gamma}, \qquad (7)$$

where $\delta T/\delta t$ is the change of temperature along a horizontal trajectory and γ is the existing lapse rate.

Air trajectories constructed from observations of either geostrophic or observed winds are not very accurate due to the large time lapse between observations, and the eddy fluctuations of the wind. The inaccuracy of the air trajectories presumably causes errors in the vertical velocities determined by the adiabatic method. Therefore the results could be greatly improved if constant-level balloons [26] were in regular use. Other errors arise from nonadiabatic temperature changes which cannot readily be evaluated. The adiabatic method can be applied in many different forms [10, 11, 14, 20, 23]. Basically, however, all these procedures are subject to similar assumptions and similar errors.

Vertical velocities computed by the adiabatic method are average values over a considerable length of time (usually twelve hours) and over a considerable distance (the length of the trajectory). They are relatively inaccurate at low levels due to nonadiabatic temperature changes there. Furthermore, when the atmospheric stability is nearly neutral, the method does not lead to dependable results.

Unlike the kinematic method, the adiabatic method is not restricted to observed winds; geostrophic winds may be used instead. Since geostrophic winds can be computed even in areas of bad weather, and at high levels, the adiabatic method is useful for drawing daily charts of vertical velocities covering a given region. Moreover, the method has been applied successfully to levels as high as 16 km.

Figure 3, taken from the paper by Panofsky [20] shows a comparison between vertical velocities obtained by the two methods. Both methods yield values of the order of 1 cm sec^{-1}. The figure shows considerable scattering which might be ascribed to inaccuracy of trajectories, small eddies in the wind field, nonadiabatic temperature changes, etc.

Several other methods have been suggested for the determination of vertical velocities. For example, the rainfall intensity is proportional to the mean vertical velocity in a saturated layer [1, 12, 17]. Hence rainfall intensities could be converted into average vertical velocities. Such vertical velocities, however, have some essentially different properties from those computed by the other two methods. For example, if convection produces equal amounts of upward and downward motion, considerable precipitation may fall. Yet the kinematic method would yield a value of zero for the mean vertical motion. This example is important, since occurrence of moderate rainfall without considerable convective activity is not common. Bannon [1] applied this technique and obtained vertical velocities greater than 10 cm sec^{-1}. This large order of magnitude is probably due to the effect of turbulence, or possibly to a somewhat smaller horizontal scale.

Another method of computing vertical velocities is based on determination of divergence from the vorticity equation and integration of the divergence. This technique has been applied by Sawyer [25] to a limited degree. His results agree qualitatively with those obtained by the other methods.

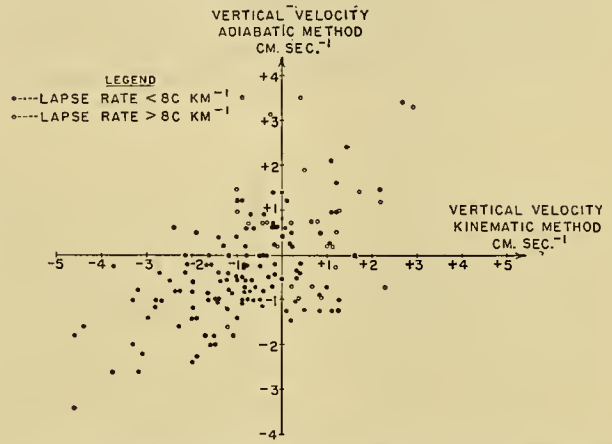


Fig. 3.—Vertical velocity by adiabatic method plotted as function of vertical velocity by kinematic method.

Distribution of Large-Scale Vertical Motion and Divergence

Considerable information is available regarding the distribution of vertical motion in the United States east of the Rocky Mountains at 10,000 ft (700 mb) and to a somewhat smaller extent at 5000 ft [15, 16, 17]. In three selected weather situations, computations were carried out up to 16 km [7]. Similar studies of more limited vertical extent were made in Europe by Hewson [11] and Petterssen [23].

Figure 4 shows a schematic picture of the distribution of divergence and vertical velocity in relation to the pressure distribution. Generally, above 5000 ft the wedge and trough lines coincide with the lines of zero vertical motion, with downward motion east of the wedge line and upward motion east of the trough lines. In other words, above 5000 ft upward vertical motion is associated with geostrophic winds (and observed winds) which have components from the south, and downward vertical motions with winds which have components from the north. Moreover, since pressure systems normally move from west to east, local pressure changes are negatively correlated with vertical velocities. The degree of correlation between vertical velocity on the one hand and meridional velocity and pressure tendency on the other varies from level to level. From 10,000 ft to the tropopause the correlation coefficients vary in magnitude in the range from 0.58 to 0.72.

In the stratosphere, the vertical velocities are generally smaller than they are in the troposphere and are of the same sign, resulting in strong vertical convergence (horizontal divergence) near the tropopause in regions of upward motion. Also, in regions of upward

motion, the equation of continuity requires horizontal convergence at the surface (Fig. 4). Quantitative computations [8] show that divergence in and above the tropopause almost exactly compensates the surface convergence. The same is true for surface divergence and stratospheric convergence. The observations are too crude to permit the exact location of a level of "nondivergence" in the upper troposphere.

For several months, synoptic vertical velocity charts were constructed at New York University at 10,000 ft or 700 mb along with the regular analysis of standard maps. Aside from the correlation between vertical velocities and meridional flow at the same levels, the following qualitative relations between the features of the surface charts and the vertical velocity charts were noted:

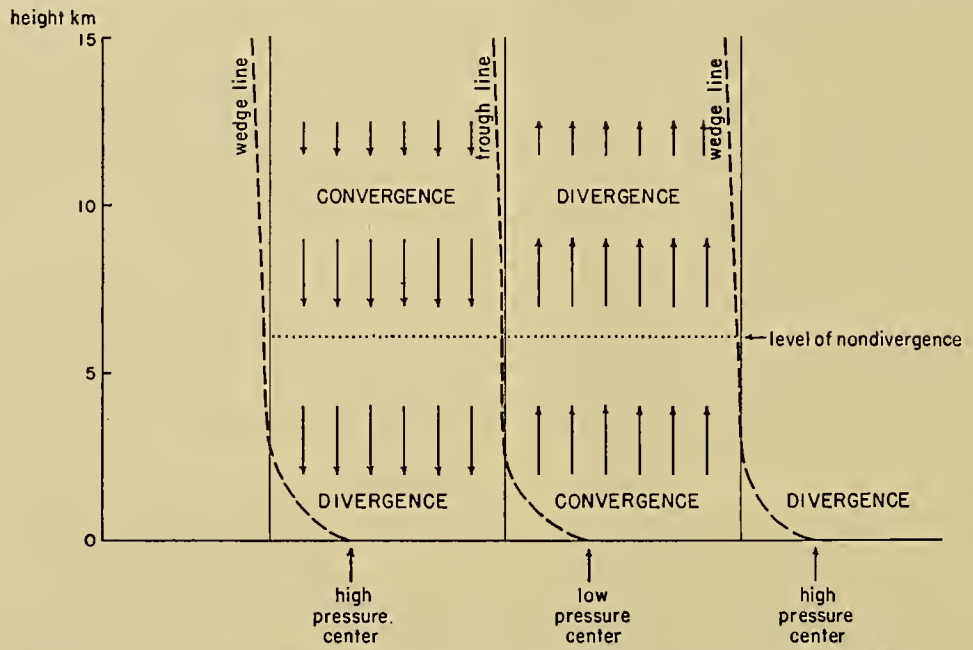


Fig. 4.—Schematic distribution of divergence and vertical velocity in an east-west cross section. Full drawn vertical lines are lines of zero vertical velocity and divergence.

1. Upward motion in cloud and precipitation areas.
2. Upward motion above low-pressure centers and fronts.
3. Downward motion above polar high-pressure regions.
4. Upward motion to the west of wedge lines in strong southerly flow.

In agreement with these results, Petterssen [23] found anticyclonic curvature of surface isobars generally associated with subsidence.

Theoretical Explanations of the Distribution of Vertical Velocity and Divergence

Bjerknes and Holmboe [3] arrived at a distribution of divergence for finite motions under the assumption of the gradient-wind equation. Essentially, there are two contributions to the divergence of the gradient wind: (1) the variation of the Coriolis parameter produces convergence in the northward flow, and (2) the curvature term in the gradient wind equation produces faster motions at the wedge than at the trough lines, hence, for sinusoidal isobars, there is convergence with flow from the north. The curvature term depends on the square of the wind speed, the latitude term on the first power. Thus the curvature term is relatively more important at high than at low levels. Quantitative computations show that near the surface the latitude term determines the distribution of divergence, and that above a critical level the curvature term takes the upper hand. This theory, again, leads to qualitative agreement with Fig. 4. Quantitatively, however, the Bjerknes-Holmboe theory predicts smaller divergence and vertical motion than are actually observed. Part of the discrepancy may be due to the fact that the Bjerknes-Holmboe theory applies to a slightly larger scale than the observations.

Charney [4] arrived at a similar distribution of vertical velocity and divergence from the complete meteorological equations; his work, however, was done by the perturbation method, so it does not permit a quantitative comparison between theoretical and observed vertical velocity and divergence.

Surface friction, which is neglected in these studies, may account for the discrepancy between theory and observations. A simplified treatment based on constant eddy viscosity [16] shows a relationship between $\nabla^2 p$, surface divergence, and vertical velocity above the friction layer; pressure minima are associated with convergence and upward vertical motion.

A simple derivation with no assumption regarding the distribution of eddy viscosity with height shows

that w_H, the vertical velocity at the top of the friction layer, is given by

$$w_H = \frac{1}{\rho f} \operatorname{curl}_v \tau_0,$$

where τ_0 is the surface stress acting on the boundary between air and ground in the direction of the wind, ρ the density, subscript v denotes the vertical component, and f is the Coriolis parameter. It is difficult to evaluate the surface stress quantitatively over land; but since the surface stress is in the direction of the wind and the wind blows counterclockwise about low-pressure centers, this formula also indicates upward motion above pressure minima. Since nonfrictional theories predict upward motion in the southerly flow above low-pressure centers, friction has the effect of increasing the absolute magnitudes predicted by these theories. Possibly a combination of the Bjerknes-Holmboe theory and the theory of friction would account for the observed distribution of divergence and vertical velocity in the troposphere.

The relation between meridional and vertical motion has been observed so far only in the United States and England, where the isotherms run essentially from east to west. It may be asked whether a similar relationship would hold when the isotherms run in some other direction.

Charney's theory is based on the assumption that the isotherms run from east to west. According to the Bjerknes-Holmboe theory the distribution of divergence depends on sinusoidal air motion superimposed on an east-west channel; upper-level streamlines behave in this way only when the isotherms run from east to west. Consequently both these theories show a relation between vertical and meridional velocities only if the isotherms are parallel to the latitude circles. In general, both theories would predict a relation between vertical motion and the horizontal wind component at right angles to the isotherms, or between vertical motion and horizontal temperature advection. Only when the isotherms run from east to west is a relation expected between vertical and meridional motion.

Effects of Divergence

Divergence and Pressure Change. One of the reasons why meteorologists have been interested in divergence is that it is related to pressure change through the Bjerknes pressure-tendency equation:

$$\left(\frac{\partial p}{\partial t}\right)_h = (g\rho w)_h - g\int_h^\infty \operatorname{div} \mathbf{V}\rho \, dz - g\int_h^\infty \mathbf{V}\cdot\nabla_H \rho \, dz, \quad (8)$$

or

$$\left(\frac{\partial p}{\partial t}\right)_s = -g\int_s^\infty \operatorname{div} \mathbf{V}\rho \, dz - g\int_s^\infty \mathbf{V}\cdot\nabla_H \rho \, dz, \quad (9)$$

where s stands for the surface which is assumed horizontal. Fleagle [8] showed that the vertical velocity term in equation (8) almost exactly compensates the divergence term, and in (9) the low-level divergence almost exactly compensates high-level divergence. Therefore neither equation can be applied to determine pressure changes from measured divergence. Moreover, most theories of divergence are not sufficiently accurate to permit estimates of pressure changes from these equations.

Divergence and Change of Vorticity. Another reason for interest in divergence is the effect it has on the individual change of absolute vorticity. The vorticity equation may be written (if we neglect solenoidal fields, friction, and terms depending on the horizontal variation of vertical velocity)

$$\frac{1}{\zeta + f}\frac{d\zeta}{dt} = -\frac{v}{R(\zeta + f)}\frac{\partial f}{\partial \phi} - \operatorname{div} \mathbf{V}, \quad (10)$$

where ζ is the vertical component of vorticity and v is the meridional velocity component.

Rossby's trajectory method was derived from the vorticity equation on the basis of negligible divergence. Some investigators attribute the systematic errors of the method in certain regions to the omission of divergence [18]. Studies at New York University seem to indicate that the divergence term is at least of the same order of magnitude as the term containing the variation of the Coriolis parameter with latitude, even for very large scales (areas of the order of 10^{16} cm^2). The omission of the divergence term is justified only near the level of nondivergence or when equation (10) is integrated vertically through the whole atmosphere.

Divergence and Change of Stability. Local changes of stability can be brought about by three factors [6]: (1) vertical divergence, (2) different horizontal temperature advection at different levels, and (3) vertical advection of lapse rate. Since vertical divergence almost equals the negative horizontal divergence, the latter may be considered as a factor producing stability changes. Fleagle [6] found that the effect of horizontal divergence on stability changes is of the same order of magnitude as that of differential advection; he also noted that accurate forecasts of local stability changes based on measured divergence and differential advection are inaccurate, probably due to the large errors in measurements of divergence.

Effects of Vertical Velocity

Vertical Advection of Velocity. Many of the meteorological equations contain vertical advection terms which have frequently been neglected. For example, the acceleration of the horizontal wind vector may be written

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + \mathbf{V}\cdot\nabla_H \mathbf{V} + w\frac{\partial \mathbf{V}}{\partial z}.$$

Charney [5] indicated that the term containing w should be one order of magnitude smaller than the other terms in the expression. This conclusion, based on a dimensional argument, is at variance with results obtained at New York University. Vertical velocities average about 1 cm sec^{-1}; the vertical shear of horizontal wind

is near 5×10^{-3} sec^{-1}. Hence, the term averages around 5×10^{-3} cm sec^{-2} in magnitude, about half as large as the other acceleration terms. Neglecting the vertical advection terms may lead to considerable errors.

Vertical Advection of Temperature and Potential Temperature. Vertical advection of temperature and potential temperature is also important, as was shown by the fact that reasonable values of w have been obtained by the effect of vertical motion on the temperature field. Again, the effect of vertical motion on local temperature changes averages about half the effect of horizontal advection [7, 19].

Vertical Motion and Pressure Changes. Since the vertical velocity has an important effect on the temperature field, and since changes in the temperature field are associated with pressure changes, Raethjen [24] first suggested a pressure-tendency equation in which the pressure tendency is expressed in terms of an integral with respect to height of local density changes. The density changes in turn are computed from horizontal advection and the effect of vertical motion. The equation can be written

$$\left(\frac{\partial p}{\partial t}\right)_{h_1} = \left(\frac{\partial p}{\partial t}\right)_{h_2} - g\int_{h_1}^{h_2} \mathbf{V}\cdot\nabla_H \rho\, dz + g\int_{h_1}^{h_2} w\rho E\, dz + g\int_{h_1}^{h_2} (1-k)\,\frac{\rho}{p}\,\frac{\partial p}{\partial t}\, dz, \tag{11}$$

where g is the acceleration of gravity, h_1 and h_2 are arbitrary levels, k is the ratio of the gas constant to the specific heat at constant pressure, and E is the stability $(\gamma_{ad} - \gamma)/T$. This equation avoids the low-level, high-level compensation of the Margules-Bjerknes tendency equation. However, the vertical velocity is not known sufficiently well to permit forecasts of pressure tendencies from (11) or a modification of it; on the contrary, Panofsky [19] used this equation to compute the vertical velocity from the pressure tendency. Later Fleagle [8] discussed the relative importance of vertical and horizontal motion on pressure changes in selected situations, and Godson [9] considered the conditions favorable for cyclogenesis, on the basis of a similar equation. These studies indicated that vertical motion has effects on local pressure changes of the same order of magnitude as horizontal motion.

Vertical Velocities and Changes in Cloudiness. Large-scale vertical velocities have their most conspicuous effect in the production of cloudiness. Since air cools when it rises, increasing cloudiness should be associated with positive vertical velocities. Panofsky and Dickey [22] and Miller and others [17] showed that middle cloudiness tends to increase along the trajectories of rising air and decrease along the trajectories of sinking air. However, vertical velocities do not seem to discriminate between overcast with or without precipitation. Figure 5, taken from Miller [15], shows this effect

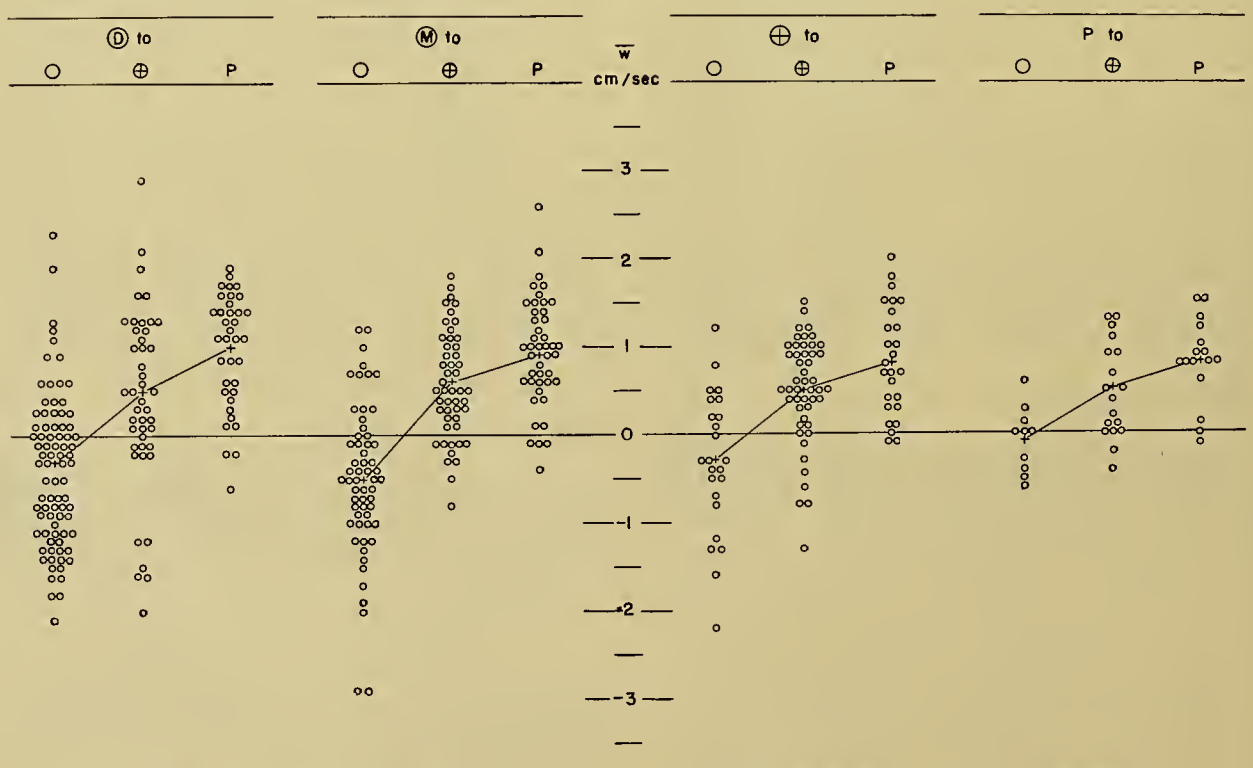


FIG. 5.—Changes of weather following 24-hr isobaric trajectory at 700 mb as function of vertical velocity. The circled D means clear with relative humidity 30 per cent or less, the circled M clear with relative humidity greater than 30 per cent, ○ clear irrespective of relative humidity, ⊕ overcast, and P precipitation. The slanting lines connect mean vertical velocities corresponding to the weather changes indicated at the heads of the columns.

graphically. According to this figure, clear weather at the end of the trajectory was usually preceded by slightly negative vertical velocities, and cloudy weather or precipitation by vertical velocities averaging about $+1$ cm sec^{-1}. These values reflect average conditions only; the figure shows instances where the final weather was clear with upward vertical velocities between 1 and 2 cm sec^{-1}, and a few cases of final precipitation with slightly negative vertical velocities. Apparently, the relation is not perfect, largely due to errors in constructing trajectories, and errors in the vertical velocities.

Vertical Velocities as a Forecast Tool

The relation between vertical velocities and changes in cloudiness suggests the possibility of applying measured vertical velocities to forecasting weather changes objectively. If the air has an upward component of motion, we might expect bad weather soon. One difficulty suggests itself immediately. If vertical velocities are computed by the adiabatic method, observations at the end of a 12-hr period are needed in order to compute average vertical velocities for that period. These vertical velocities are centered in the middle of the period and are therefore already 6 hr old at the time the analysis is started. A further lag is caused by transmission time, analysis and computation, so that the vertical velocities are at least 8 hr late when they are ready for application to "forecasting."

As Fig. 5 indicates, cloud changes are related to the average vertical motion along the air trajectory. If a 24-hr forecast is desired, a 24-hr air trajectory must be forecast and the average vertical velocity along the trajectory must be estimated from quantities at the beginning of the trajectory.

Since the forecasting method was to be objective, it was necessary to devise a method of forecasting trajectories objectively. It would require too much space here to describe the method finally applied; but it is clear that errors in the forecast trajectory will lead to inaccurate future positions of the air and hence to additional errors in the forecast.

Experiments with different variables indicated that an estimate of the mean vertical velocity along the trajectory could be made from the initial vertical velocity and the initial meridional velocity. The multiple correlation coefficient of mean vertical velocity on initial vertical velocity and meridional wind components was 0.64 [15]. Clearly, an estimate of the mean vertical velocity based on these two variables is subject to considerable error.

In practice it was possible to side-step the computation of the average vertical velocity along the trajectory. Since the average vertical velocity is a function of initial vertical and meridional velocity components, and the change of weather is a function of the average vertical velocity along the trajectory, weather changes should be directly related to the initial meridional and vertical velocity components. Charts were constructed for the first half of December 1945, which showed the final weather and cloudiness as a function of initial weather and of the vertical and meridional components of velocity. In certain regions of these charts clear sky (cloudiness 0.4 or less) was predominant; in other regions overcast and precipitation predominated. A line was drawn which separated the sections of predominantly clear sky from sections of sky mainly overcast. It was quite difficult to find dividing lines between overcast and precipitation on these charts. Such lines were drawn, however, under the assumption that precipitation should occur with large upward vertical velocities along the trajectories.

Forecasts were made from these charts for the last half of December 1945. The total percentage of correct forecasts was 69.9 per cent. The chance score was 57.3 per cent.

A more severe test is the comparison of the percentage of successful forecasts with those obtained from "low skill" forecasts. For example, if "clear" had been forecast all the time, the score would have been 69.2 per cent; if no change of local weather had been forecast, the percentage of hits would have been 59.7 per cent. The same forecasts were repeated by two graduate students who had considerable forecasting experience. These forecasters had no knowledge of the vertical velocities at the time. The two forecasters scored 63.4 per cent and 69.8 per cent correct, respectively.

Altogether, the objective method of forecasting vertical velocity did not perform badly, in spite of the many sources of error. Later studies [16] showed the method considerably less successful. Moreover, Miller showed that equally good forecasts could be made by a very similar method, if the initial vertical velocities were not known at all, but the forecasts were based solely on the initial weather and meridional velocity component. Therefore the laborious computation of vertical velocities for forecasting purposes is possibly unnecessary. Apparently, the meridional velocity component is a sufficiently good indicator of the direction of the vertical air motion. The relation between meridional motion aloft and weather is known to many meteorologists, and has been incorporated, indirectly, into other objective forecasting methods.

Some attempts were made to use vertical velocity patterns qualitatively. On daily vertical velocity charts the "unusual" features that did not conform to the simultaneous weather were noted. For example, an area of upward motion over an east coast wedge was judged significant. Such unusual features were followed sometimes, but not always, by unusual developments; for example, unusual centers of upward motion were associated with cyclogenesis. Several forecast rules based on unusual vertical velocity patterns were suggested, but none of them proved reliable. Whether this was due to errors of the vertical velocities cannot yet be decided.

Suggestions for Future Research

Since most of the studies of vertical velocity were completed the network of radio and radar wind observations has been greatly extended. As a result, de-

termination of divergence and vertical velocity by the kinematic method would seem possible in forecasting techniques. The advantage of the kinematic method is the possibility of computing vertical velocities within only a few hours after the time to which they apply. This and the rather indecisive results of the earlier forecast studies might make it desirable in the future to study forecasting techniques based on vertical velocities.

Another possibility, which has been only partially tested thus far, is the application of the vorticity equation (10) to the forecast of changes of vorticity from observed divergence.

So far, systematic work on vertical velocities has been restricted to middle latitudes in the Northern Hemisphere. It is quite possible that the observed relation between vertical and meridional velocities, for example, is not valid everywhere. Studies in other parts of the world are desirable in order to determine whether there exists a general relation between vertical and meridional motion or vertical velocity and advection. Moreover, a study of the relation between vertical velocities, divergence, and weather is desirable in other sections of the globe.

In general, a better understanding of vertical motion will presumably arise out of the work which Dr. Jule Charney and his collaborators are doing at The Institute for Advanced Study on numerical forecasting based on the physical equations.

REFERENCES

1. Bannon, J. K., "The Estimation of Large Scale Vertical Currents from the Rate of Rainfall." *Quart. J. R. meteor. Soc.*, 74: 57–66 (1948).
2. Bellamy, J. C., "Objective Calculations of Divergence, Vertical Velocity and Vorticity." *Bull. Amer. meteor. Soc.*, 30: 45–49 (1949).
3. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
4. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4: 135–162 (1947).
5. —— "On the Scale of Atmospheric Motions." *Geofys. Publ.*, Vol. 17, No. 2 (1948).
6. Fleagle, R. G., "A Study of the Effects of Divergence and Advection on Lapse Rate." *J. Meteor.*, 3: 9–13 (1946).
7. —— "The Fields of Temperature, Pressure, and Three-Dimensional Motion in Selected Weather Situations." *J. Meteor.*, 4: 165–185 (1947).
8. —— "Quantitative Analysis of Factors Influencing Pressure Change." *J. Meteor.*, 5: 281–292 (1948).
9. Godson, W. L., "A New Tendency Equation and Its Application to the Analysis of Surface Pressure Changes." *J. Meteor.*, 5: 227–235 (1948).
10. Graham, R. C., "The Estimation of Vertical Motion in the Atmosphere." *Quart. J. R. meteor. Soc.*, 73: 407–417 (1947).
11. Hewson, E. W., "The Application of Wet-Bulb Potential Temperature to Air Mass Analysis." *Quart. J. R. meteor. Soc.*, 62: 387–420 (1936); 63: 7–29 (1937); 64: 407–418 (1938).
12. Holmboe, J., Forsythe, G. E., and Gustin, W., *Dynamic Meteorology*. New York, Wiley, 1945. (See p. 143)
13. Houghton, H. G., and Austin, J. M., "A Study of Nongeostrophic Flow with Applications to the Mechanism of Pressure Changes." *J. Meteor.*, 3: 57–77 (1946).
14. Longley, R. W., "Subsidence and Ascent of Air as Determined by Means of the Wet-Bulb Potential Temperature." *Quart. J. R. meteor. Soc.*, 68: 263–276 (1942).
15. Miller, J. E., *Application of Vertical Velocities to Objective Weather Forecasting*. Prog. Rep., Mimeogr., College of Engineering, New York University, 1946.
16. —— "Studies of Large Scale Vertical Motions of the Atmosphere." *Meteor. Pap., N. Y. Univ.*, No. 1, 48 pp. (1948).
17. —— and others, *A Study of Vertical Motion in the Atmosphere*. Prog. Rep., Miemogr., College of Engineering, New York University, 1946.
18. Namias, J., and Clapp, P. F., "Normal Fields of Convergence and Divergence at the 10,000-Foot Level." *J. Meteor.*, 3: 14–22 (1946).
19. Panofsky, H. A., "The Effect of Vertical Motion on Local Temperature and Pressure Tendencies." *Bull. Amer. meteor. Soc.*, 25: 271–275 (1944).
20. —— "Methods of Computing Vertical Motion in the Atmosphere." *J. Meteor.*, 3: 45–49 (1946).
21. —— "Objective Weather-Map Analysis." *J. Meteor.*, 6: 386–392 (1949).
22. —— and Dickey, W. W., "Vertical Motion and Changes in Cloudiness." *Bull. Amer. meteor. Soc.*, 27: 312–313 (1946).
23. Petterssen, S., and others, "An Investigation of Subsidence in the Free Atmosphere." *Quart. J. R. meteor. Soc.*, 73: 43–64 (1947).
24. Raethjen, P., "Advektive und konvektive, stationäre und gegenläufige Druckänderungen." *Meteor. Z.*, 56: 133–142 (1939).
25. Sawyer, J. S., "Recent Research at Central Forecasting Office, Dunstable," in a discussion, "Large Scale Vertical Motion in the Atmosphere." *Quart. J. R. meteor. Soc.*, 75: 185–188 (1949).
26. Spilhaus, A. F., Schneider, C. S., and Moore, C. B., "Controlled-Altitude Free Balloons." *J. Meteor.*, 5: 130–137 (1948).

THE INSTABILITY LINE

By J. R. FULKS

U. S. Weather Bureau, Washington, D. C.

Introduction and General Features

The synoptic meteorologist frequently encounters nonfrontal squall lines, particularly in the warm sectors of certain types of extratropical cyclones. These nonfrontal squall lines, now designated by the International Meteorological Organization as *instability lines*, are relatively frequent in the United States east of the Rockies and may be severe in character. Some are accompanied by tornadoes and presumably extreme vertical instability through a relatively deep layer of the atmosphere.

A well-developed instability line is marked by squalls or thunderstorms along a line that is usually several hundred miles in length, and in the typical case fifty to three hundred miles ahead of a surface cold front. Or, where scattered showers or general rains are already occurring, the instability line represents a sharp intensification of squall conditions and rainfall, lasting usually for about an hour at any one location. Since these conditions may occur, for the most part, between regular synoptic reports, the best evidence is often in the reported time, character, and amount of rainfall, or in frequent intermediate reports. The pattern of reported thunderstorms is usually also an indication of the existence and location of an instability line, but in some situations scattered thunderstorms occur before and after its passage. By convention the line is taken to be the leading edge of the band, usually thirty to fifty miles in width, of greatest convective activity. In more complicated cases, there may be several bands of squalls. Aircraft encounter marked and sometimes dangerous turbulence in flying through an instability line.

The term *instability line* is also applied to the incipient condition, when synoptic factors indicate that a nonfrontal squall line is forming, and may be applied to the line of instability in the dissipating stage when squalls have ceased.

Alternately, the term *instability line* is defined as a pseudo-cold front for which the cold air is presumably produced by precipitation falling from a line of accompanying showers. While the pseudo-cold front marking the leading edge of the outflowing rain-cooled air from instability showers is observed in most nonfrontal squall lines, it appears to be normally a secondary effect, other factors probably being more important in forming and maintaining the line of squalls. However, there are some cases where thundershowers develop over a localized region of, say, 10,000 or 20,000 square miles forming a surface layer of cool air which flows outward in all directions, new thundershowers then tending to develop along the resulting pseudo-cold front where the wind direction in the warm air is favorable to lifting over the cooler surface air.

Unlike a true front, the instability line is transitory in character, usually developing to maximum intensity within a period of twelve hours or less and then dissipating within about the same period of time. Although by definition it is nonfrontal in character, it may cross a front. Often the line extends from the warm sector of a low northward across the warm front and is associated with a line of squalls in the northeastern quadrant of the low.

Harrison and Orendorff [5] in 1941 studied the "precoldfrontal squall line," especially from the standpoint of airline operations. In addition to describing the squall line and associated conditions, they examined physical factors possibly contributing toward its development. Without attempting to reach definite conclusions they presented data which suggested a pseudo-cold front of rain-cooled air moving against a rainless zone in the same air mass, and convergence within a potentially unstable air mass, as being at least possible causative factors in most of the cases studied. One of the significant facts they reported was that activity along the cold front decreases during the active stage of the squall line but regenerates as the squall line dissipates. Their study was confined to the Cheyenne–New York airway (between April 1939 and May 1940) and probably was not fully representative of squall-line conditions elsewhere; for example, they reported no sustained temperature change through the squall line aloft. The heights of upper levels considered were not fully specified; presumably they were flight levels and thus often below heights at which cooling is frequently observed in the general region of the instability line.

Lloyd [6], in his studies of tornadoes, attributed at least the tornado-producing squall-line condition to an upper cold front. Advection of colder air aloft and the resulting decrease of temperature, as would be required by existence of an upper cold front, is observed in many cases. But cooling aloft also takes place, in some instances, ahead of the squall condition, indicating that the degree of instability near the leading edge of the cooling aloft is not always sufficient for convective activity. Back from the leading edge, temperatures aloft may be lower, and surface temperatures higher, so that at some point the critical lapse rate necessary for convection may be realized. The exact distance between the squall condition and the leading edge of the colder air aloft is difficult to determine by observation, but appears to vary from a few miles (possibly less) to hundreds of miles. Where the leading edge of advection of colder air aloft is very close to the squall condition, and there is evidence of a discontinuity in density as required for a front aloft, the condition is not properly classified as "nonfrontal." In many instances, however, the existence of a discontinuity aloft along or near the

squall condition is difficult or impossible to determine from synoptic data and the only practical recourse for the synoptic meteorologist is to follow the instability line.

Synoptic Aspects and Some Associated Physical Factors

Figure 1 represents diagrammatically a typical well-marked instability line in the warm sector of an extra-

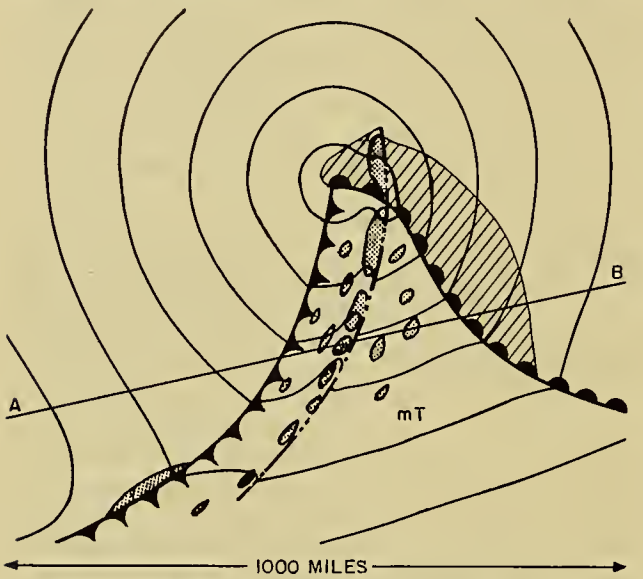


FIG. 1.—Model of cyclone with instability line in warm sector. Shading indicates areas of active squalls, and hatching shows warm-front precipitation.

tropical cyclone. Squalls and thunderstorms (shaded area) are shown as a nearly solid band along the instability line and at scattered points elsewhere in the warm sector and within the area of warm-front precipitation (hatched). In this type of low, and at this stage of development, precipitation is rarely observed behind the portion of the cold front following the instability line.

Figure 2 is a vertical cross section through the line *AB* of Fig. 1. Thin solid lines represent isotherms of

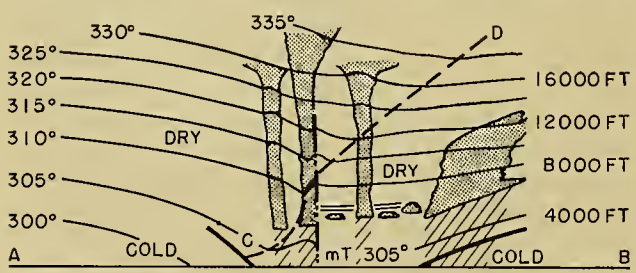


FIG. 2.—Vertical cross section through cyclone with warm sector instability line. The section is taken along the line *AB* of Fig. 1 and is on the same horizontal scale; the vertical scale is indicated roughly in feet on the right-hand side. Thin quasi-horizontal lines are potential temperature, labeled in °K on the left. The dashed line *CD* is the axis of warm air, temperatures decreasing or changing but little horizontally to the right, and decreasing to the left of *CD* (except near the ground where temperature decreases sharply with onset of instability line squalls). Shading indicates cloud masses, and hatched lines show active precipitation; the double horizontal line indicates a stable layer.

potential temperature. The line *CD* is the center of a warm tongue which extends roughly at right angles to the given vertical section. Thus, any constant level or constant pressure surface cutting the vertical section will have a warm tongue, the axis of which intersects the line *CD*. Generally the horizontal temperature gradient to the right of *CD* is less than to the left of *CD*. As the system moves from left to right (normally eastward) cooling aloft over any fixed point will begin as the line *CD* passes, and since the cooling takes place first at the higher levels, there will be a progressive decrease in vertical stability until the arrival of the instability line. Often, at least in the central and eastern United States, there is a stable layer or inversion at approximately 1½ km above the surface and to the east of the instability line; this inversion disappears with the arrival of the instability line. It is probable that this stable layer, when it exists, plays an important part in the mechanism of the instability line, because it provides a "cap" under which the temperature and moisture content may increase until, combined with any cooling aloft, the vertical instability is sufficient for vertical convection. This effect is made possible by slower eastward movement of the surface air as compared to the system as a whole, and is enhanced by conditions usually favorable for low-level advection northward of moisture and warmer air, as pointed out by Means [7]. Apparent warm advection is in some cases, however, indicative of lifting rather than low-level warming.

The cooling aloft, as illustrated in Fig. 2, is in part the result of horizontal advection, as may readily be verified by inspection of constant-level or constant-pressure charts. The typical condition is one of cooling by advection aloft over a region where there is warming by advection at lower levels. The warming by advection at low levels is generally greater in magnitude than the cooling aloft, but cooling aloft is significant in that it permits vertical convection to extend to higher levels than would otherwise be possible, thus becoming an important factor in the intensity of resulting convective activity. It seems certain, however, that factors other than advection are also important in the cooling which takes place aloft over the warm sectors of many lows. An early example of cooling aloft ahead of a surface cold front, accompanied by warm-sector precipitation was presented by J. Bjerknes [1] in 1930. This study was based on a detailed examination of frequent soundings at Uccle, Belgium, during the period December 26–28, 1928.

The cold front as shown in Fig. 2 does not extend far above the surface. When the instability line is well developed, there is usually little or no evidence of the existence of the associated cold front at more than two or three thousand feet above the ground. The position of the cold front at the ground is usually quite well marked by a wind shift, pressure trough, and change of moisture content (dryer on the cold air side). Usually there is little, if any, temperature change immediately across the front, but there is a horizontal temperature gradient toward colder air to the rear, suggesting that,

in many cases at least, there is no true discontinuity of density but rather a discontinuity in the gradient of density. In synoptic practice it is customary, however, to continue to carry and designate this line as a cold front, because at a later stage in the history of the low the instability line tends to dissipate and the "front" again takes on the characteristics of a true cold front.

Along the instability line, when it is well developed, there is marked cyclonic wind shear at low levels, associated with a pressure trough. At higher levels, as in extratropical cyclones generally, there is a pressure trough which roughly parallels the surface cold front. In many cases of well-marked instability lines, the pressure trough aloft coincides with the instability line. In some cases it is back of the surface cold front, as is normal in lows which do not have an instability line.

As with shower conditions generally, surface temperatures usually fall with the arrival of the instability line, and this fact, combined with the cyclonic wind shear, suggests a surface frontal structure. However, the temperature often rises after passage of the instability line and remains high until after passage of the cold front; moisture content also remains high until after passage of the cold front.

The fact that cyclonic wind shear is observed at the surface along a well-developed instability line provides one clue to its location on the surface synoptic chart. Or, when a line of squalls is not yet in existence, the appearance of a line of cyclonic wind shear within a warm moist air mass, particularly in the warm sector of a low, may indicate the existence of an incipient instability line. The instability line is either in a pressure trough, or if parallel to the isobars, is along a discontinuity in pressure gradient. The wind shear in the horizontal is similar to that along a front, but unlike a front, the wind discontinuity is more nearly vertical except along the outrush of cooler air at very low levels ahead of individual squalls.

With the density of reporting stations in the United States, the position of the line of shear can ordinarily be located on the synoptic chart only within a range of fifty or one hundred miles, using only reported pressures and winds. A more exact location of the instability line requires detailed examination of individual reports, particularly for the time of onset of squalls associated with the line and for the time of the wind shift.

Since the instability line is associated with a horizontal discontinuity in pressure gradient moving with the line, it might be expected to exhibit a tendency field similar to any pressure trough, that is, falling pressure ahead and rising pressure behind, or falling ahead and falling less rapidly behind, etc. When there is a sharp pressure trough, a tendency field of this type can be detected, but in general the three-hour tendencies represent a confused pattern that may be misleading if not considered with caution. When the line is accompanied by severe squalls, there may be rapidly falling pressure ahead of the line, a sharp rise with the onset of squalls, then a rapid fall within an hour or so, all of which may take place within the three-hour period for which the tendencies are reported. Also, the severity of the weather along the line tends to vary from station to station, and there are corresponding irregularities in the reported tendencies. The tendencies are, however, a useful guide if viewed with caution and if allowances are made for known squall activity at individual stations. Their use as a guide to location of the instability line should be looked upon as subordinate to the more direct location of the pressure trough or line of wind shear, or the observed onset of squalls at individual stations.

Upper winds, when available, are useful in locating the line of wind shear at the gradient level if reports are not too far apart. Surface winds are similarly helpful if allowance is made for local topographical effects and for the fact that isolated squalls may occur ahead of or behind the instability line.

There is no clear guide to the speed of movement of the instability line other than its past history and the known speed of movement of a low with which it may be associated. The pre-coldfrontal squall line moves slightly faster than the cold front which it precedes. The movement of the instability line is at about the same speed as the axis of the warm tongue aloft, which is normally along or ahead of the instability line and moves with a speed somewhat less than that of the wind speed aloft normal to the axis. The movement of the axis of warm air aloft is not particularly useful for forecasting the movement of the instability line once the line is in existence, but for longer-period forecasting its movement is some guide to the time and location of possible future squall-line developments.

Over the western Great Plains and the Mississippi Valley, a cold front aloft or the advection of colder air aloft often has its inception as a surface cold front over the western mountain region. The surface front, marking the arrival of fresh maritime polar (mP) air from the Pacific, moves across the Plateau region. The shallower portion of the cold air, near the leading edge, becomes heated by the warmer ground over which it moves, and also by compression as it flows downward along the east slope. The heating effect is less within the deeper portion of the cold air so that a horizontal temperature gradient is established within the cold air. The leading edge of the cold air while over the Plateau region remains colder than the surface air ahead of the front, but as the front moves down the east slope of the Plateau it encounters maritime tropical (mT) air which overlies the western plains region and which may intersect the east slope at an elevation of from three to five thousand feet above sea level. Typically, under these conditions, there is a stable layer near the top of the mT air, and the mT air below this stable layer has a potential temperature lower than that of the leading edge of the cold air, so that the leading edge of the cold air seeks out a level within the moist mT air or above it where the potential density is the same as at the base of the cold air. Thus the leading edge of the mP air overrides the surface mT air. Occasionally there is sufficient instability and moisture in the overriding mP air for the development of high-

level showers or thunderstorms entirely within the polar air aloft.

Somewhere within the *mP* air mass, usually between fifty and three hundred miles from its leading edge aloft, the potential density is equal to the potential density of the *mT* air at the ground over the Great Plains region. The *mP* air at that point begins to displace the *mT* air at the ground. The line thus formed between surface *mP* and surface *mT* air has in general the characteristics of an occlusion of the warm-front type, though temperatures at the ground are usually the same immediately on either side of the line. There is usually a marked temperature gradient toward colder air on the *mP* side, and a sharp difference in moisture content between the moist *mT* and the dryer *mP* air. This "front" is not accompanied by precipitation but may mark the end of showers in the *mT* air mass. As the system moves eastward or northeastward across the western Plains and Mississippi Valley a more important surface cold front normally becomes evident in the system between the *mP* air and a colder *cP* air mass to the north.

The conditions thus produced as a result of the eastward flow of cold air off the Plateau become somewhat similar to those pictured in Figs. 1 and 2, and a well-marked instability line tends to develop. The height at which the cold air aloft advances farthest ahead is, however, considerably lower than is shown in Fig. 2. Above that level the boundary of cold air slopes upward to the left as with a conventional cold front.

The existence in the United States of an extensive area of low elevation open on the south to a plentiful supply of warm moist air and bounded on the west by a plateau, combine to make the United States east of the Rockies a favorable region for the production of well-marked instability lines. This is not only because of the tendency for cold air off the Plateau to override *mT* air, but apparently also because the Rockies provide a north-south barrier to lower-level air masses, favoring southward flow of cold air and northward flow of warm air, the zone of interaction between the pure *mT* air and the cold air being most often within the latitude of the United States.

Unique geographical and topographical conditions are, however, by no means necessary to the formation of the instability line. It may be found anywhere in middle or subtropical latitudes and perhaps in other regions. But tornadoes in the United States appear to develop most often with the type of instability line that results when cold air from the Plateau overrides *mT* air over the Great Plains.

Further Discussion of Physical Processes

The pseudo-cold front, suggested by Harrison and Orendorff [5] as important in formation of the pre-coldfrontal squall line, is observed in nearly all cases. Unquestionably its relation to other factors involved must be considered in any complete explanation of the mechanism. However, it seems equally certain that this factor must be secondary to other initial causative factors, because the existence of the shallow layer of rain-cooled air in the proper location appears to be the result of showers already occurring. The same authors suggested that convergence within a potentially unstable air mass might act together with the formation of the pseudo-cold front as a pair of factors to produce the squall line. This might be taken as a suggestion that convergence first produces the line of squalls and that the resulting pseudo-cold front then assists in maintaining the squalls and perhaps in keeping them along a line rather than allowing them to disperse in a disorganized manner. Low-level convergence is of course a necessity along the instability line and needs, itself, to be explained in terms of other factors. Are there other factors which first act to produce low-level horizontal convergence and upper-level divergence, or is vertical motion merely the result of vertical instability and thus the full cause of the convergence-divergence pattern?

The Olivers [9], in discussing forecasting of frontal weather from winds aloft, stated that if the wind component perpendicular to a cold front increases with height through the frontal surface, no weather is produced by the front, but that convergence in the frontal trough may produce weather in the warm sector which ceases abruptly on passage of the surface cold front. They extended this argument further as an explanation of the pre-coldfrontal squall line, that is, the air flowing downslope over the frontal surface may extend to near the ground for some distance ahead of the front and, being dry as a result of subsidence, would prevent shower conditions from forming along or immediately ahead of the cold front, while convergence in the trough would continue to produce showers in the *mT* air. This explanation, as far as it goes, seems to agree well with observation, and if some of the rain falls through the dry subsiding air the condition would be favorable for formation of a pseudo-cold front in the warm sector as envisioned by Harrison and Orendorff. A difficulty is that air at the ground back of the squall line is indistinguishable from the *mT* air at the ground ahead of the line, except for cooling which can be accounted for by rain. However, this difficulty is not serious if the dry air does not extend all the way to the ground, or if its moisture content is increased sufficiently by the rain falling through it.

It has also been suggested by various persons that, from vorticity considerations, the stronger flow of air above the frontal surface will tend to form a pressure trough as it flows into the warm sector, the argument being that subsidence will be greater at low than at high levels, resulting in vertical stretching and therefore increased cyclonic vorticity along and immediately ahead of the surface cold front. This argument is strengthened by its analogy to the explanation of the pressure trough which normally forms in the lee of extensive mountain ranges.

H. B. Wobus (U. S. Weather Bureau, Washington, D. C.) suggested to the author some years ago that the existence of a warm tongue in the mean temperature through a deep layer is favorable for the production of instability along a line near the axis of the tongue, if

it is assumed that the wind at all levels is instantaneously adjusted to the gradient value. Instability lines normally occur near the axis of a warm tongue as represented in the horizontal distribution of mean virtual temperature through a layer, say, 20,000 ft in depth. The curvature of the component of gradient wind flow perpendicular to the axis of the warm tongue must, because of this temperature distribution, become more anticyclonic (or less cyclonic) with height. The component of *gradient* flow across the axis must therefore tend to increase with elevation, and to carry the upper colder air over the lower-level position of the axis of warm air, thus decreasing the vertical stability. Also, in a manner pointed out by Rossby [10], a decrease of mean virtual temperature northward along the axis, as is generally observed, must cause an increase of geostrophic wind from the west and thus normally act in the same direction as the effect of curvature to produce greater eastward advection of colder air aloft as compared to lower levels. These considerations do not, of course, take account of important nongradient components of flow that must exist.

Conditions near the center of a warm tongue are also favorable for upward vertical motion if the energy of the solenoidal field can be realized, and these same conditions provide a means for eviction of air aloft because of nongeostrophic components of flow as pointed out by Durst and Sutcliffe [2]. Release of latent heat of condensation tends to maintain the solenoidal field. Some low-level convergence into the warm axis can be accounted for by frictional effect in the pressure trough. Further low-level convergence is a possibility in the layer above surface frictional effects in the presence of dynamic instability as pointed out by Solberg [12]. However, it is not known from observation that the horizontal component of dynamic instability (as first discussed in this country by Fulks [4]) is of general importance in the case of instability lines, the requirement being that there is appreciable anticyclonic vorticity in the horizontal wind field. Probably this factor is important in the immediate vicinity of the squalls but, in general, instability in the vertical appears to be predominant.

Some Remarks on Tornadoes

Tornadoes occur normally along instability lines and must therefore involve some of the same dynamic factors, though not all instability lines are accompanied by tornadoes. Studies or discussions of the mechanics of tornadoes by Bigelow, Humphreys, Jakl, Showalter, and others in this country, Wegener in Germany, some Russian investigators, and others have given some hint of the factors involved, but much more work remains to be done. Basically, there are two main factors to be explained:

1. *Source of energy.* This has usually been assumed to be vertical instability, such as is known to exist in the vicinity of instability lines. This explanation is, however, not sufficient unless it can be shown that the difference between instability lines which produce tornadoes and those which do not is a matter of degree of vertical instability, or that other factors are absent in one case and not the other. Showalter [11] suggested that precipitation carried aloft and falling ahead of the squalls, cooling the air at higher levels, might be a means of producing a high degree of instability more or less spontaneously. In any case some nearly spontaneous local release of energy seems necessary. Another possible means of creating a vertically unstable lapse rate is for the convectively unstable air mass ahead of the squall line to be lifted bodily. This would itself require some source of energy and seems less likely than the cooling by precipitation aloft for which convective instability is also important in producing a steep vertical lapse rate. An important possible source of energy other than vertical instability, which cannot be ruled out without further study, is that of the already existing wind circulation in the vicinity of tornado formation. Existing kinetic energy fed into the whirl would appear to be at least a contributing factor once the tornado is formed.

2. *Source of rotation.* It seems necessary that the rotation of the tornado results from low-level horizontal convergence within an already existing field of cyclonic (or occasionally anticyclonic) vorticity. The convergence appears to take place first in the upper portion of the warm moist air, roughly 2000–5000 ft above the ground (above surface frictional effects), and the whirl appears to extend rapidly upward and to bore downward to the ground. Tornadoes normally occur along or near a line of cyclonic shear, but this shear line is usually between the warm air moving rapidly northward and the wedge of rain-cooled air moving eastward. Since the required vorticity, to be effective, must be entirely within the warm air because any injection of cooler air at lower levels would tend to damp out convection, it is unlikely that the line of cyclonic shear is a direct source of the vorticity. A more likely source of vorticity is in the interaction between the warm and cool air, especially in warm tongues that must break off from the northward-moving warm air mass and flow into individual convective cells. This turning from northward to westward would produce cyclonic curvature in the warm air as well as some shear, cyclonic on the southern side of the tongue and anticyclonic on the northern side, the latter being a possible factor in producing the rarely observed anticyclonic rotation. In this connection it is also necessary to consider the irregular nature of outflowing tongues of cold air as described by Williams [15] and their possible effect on warm air flow. Another source of rotation that has long been considered is that of the roll cloud of the thunderstorm, the roll being presumed to extend down to the ground, resulting in a cyclonic whirl on one end, or an anticyclonic whirl if the other end should reach the ground; this theory has not been fully refuted or confirmed, but appears unlikely in the light of descriptive reports.

Some Recent Studies

Tepper [13] recently proposed that a "pressure-jump line" is an important part of the mechanism of squall

lines. A sudden rise of pressure is normally observed upon arrival of the wedge of rain-cooled air. However, Tepper further theorizes that the "pressure jump" is associated with a gravitational wave aloft (following a suggestion by Freeman [3]), and he suggests that the wave originates from an acceleration of the cold front. No convincing evidence has been presented in support of this theory. It seems extremely unlikely that there can be any close association between the leading edge of outflowing cold air at low levels and a gravitational wave aloft, as would be required.

The most substantial work to date is that based on observations of the recent Thunderstorm Project. Williams [15] made a microanalysis of selected squall lines passing through the Wilmington net. The official report of the Thunderstorm Project [14] further discusses some of the factors observed in connection with squall lines, and a still more recent paper by Newton [8] attempts to give a more detailed picture of their physical structure and mechanics, based on Thunderstorm Project data. The latter two studies are too recent to have been taken into account in the preceding discussion. Newton's examples probably do not embody the salient characteristics of all types of squall lines but shed light on the type included in his studies. He suggests that squall lines in some cases appear to form first over the cold-front surface and subsequently to move into the warm sector, and he confirms by observation the existence of a pseudo-cold front along the warm-sector squalls. He further suggests that squall-line activity can be accounted for partly as a result of this "front" and partly by the continuous generation of new thunderstorms as a result of convergence-divergence patterns produced by the vertical transfer of horizontal momentum in pre-existent thunderstorms and augmented by solenoidal circulations. He suggests that kinetic energy brought down from higher levels is an essential source of energy for maintaining squall-line activity.

Future Research

Future progress toward a better understanding of the physical structure and mechanics of the instability line will require both the collection of more adequate detailed observational data and the analysis of existing and future data. Undoubtedly there is much to be gained by further analysis of the Thunderstorm Project data, though this project was not aimed so much at the specific problems of the instability line as of the thunderstorm. More observational data for upper levels in the immediate vicinity of tornadoes is especially needed; this is difficult to obtain but will be very important because of the frequent association of tornado conditions with the instability line. A more adequate theory of the mechanics of the tornado would be an important step toward understanding the instability line.

From the practical standpoint of the synoptic meteorologist, whose primary problem is to forecast development of the instability line before it occurs, much remains to be done. His problem is somewhat different from that of basic research in that he must develop better forecasting methods based only on the use of data which are available synoptically. Synoptic data are at best very incomplete, so that he must interpolate to a great extent; better basic knowledge of the phenomenon would aid in this interpolation. There is a large and relatively untouched field for synoptic studies, particularly from a statistical standpoint, in which only synoptically available data are used. There are important geographical differences involved so that similar types of studies need to be made for different geographical areas. While forecasting development of the line is the main synoptic problem, it is also important to develop criteria for determining the individual characteristics of each instability line (turbulence, intensity of rainfall, surface winds, icing, likelihood of tornadoes, etc.) and means of forecasting the dissipation of the line.

REFERENCES

1. Bjerknes, J., "Exploration de quelques perturbations atmosphériques à l'aide de sondages rapprochés dans le temps." *Geofys. Publ.*, Vol. 9, No. 9 (1930).
2. Durst, C. S., and Sutcliffe, R. C., "The Importance of Vertical Motion in the Development of Tropical Revolving Storms." *Quart. J. R. meteor. Soc.*, 64:75–84 (1938).
3. Freeman, J. C., Jr., "An Analogy between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5:138–146 (1948).
4. Fulks, J. R., *Some Aspects of Non-gradient Wind Flow*, unpublished. Paper presented before Joint Meeting of Amer. Meteor. Soc. and Amer. Inst. of Aero. Sciences, New York, January, 1946.
5. Harrison, H. T., and Orendorff, W. K., "Pre-coldfrontal Squall Lines." *United Air Lines Meteor. Dept. Circ.* No. 16 (1941).
6. Lloyd, J. R., "The Development and Trajectories of Tornadoes." *Mon. Wea. Rev. Wash.*, 70:65–75 (1942).
7. Means, L. L., "The Nocturnal Maximum Occurrence of Thunderstorms in the Midwestern States." *Dept. Meteor. Univ. Chicago Misc. Rep.*, No. 16 (1944).
8. Newton, C. W., "Structure and Mechanism of the Prefrontal Squall Line." *J. Meteor.*, 7:210–222 (1950).
9. Oliver, V. J., and Oliver, M. B., "Forecasting the Weather with the Aid of Upper-Air Data" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed., pp. 813–857. New York, McGraw, 1945. (See pp. 815–817)
10. Rossby, C.-G., "Kinematic and Hydrostatic Properties of Certain Long Waves in the Westerlies." *Dept. Meteor. Univ. Chicago Misc. Rep.*, No. 5 (1942). (See pp. 32–37)
11. Showalter, A. K., and Fulks, J. R., *Preliminary Report on Tornadoes*. U. S. Weather Bureau, Washington, D. C., 1943. (See pp. 20–28)
12. Solberg, H., "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones." *P. V. Météor. Un. géod. géophys. int.*, Edimbourg, 1936. II, pp. 66–82 (1939).
13. Tepper, M., "A Proposed Mechanism of Squall Lines: The Pressure Jump Line." *J. Meteor.*, 7:21–29 (1950).
14. U. S. Weather Bureau, *The Thunderstorm*, 287 pp. Washington, D. C., U. S. Govt. Printing Office, 1949. (See pp. 122–130)
15. Williams, D. T., "A Surface Micro-study of Squall-Line Thunderstorms." *Mon. Wea. Rev. Wash.*, 76:239–246 (1948).

LOCAL CIRCULATIONS

Local Winds *by Friedrich Defant* 655

Tornadoes and Related Phenomena *by Edward M. Brooks* 673

Thunderstorms *by Horace R. Byers* 681

Cumulus Convection and Entrainment *by James M. Austin* 694

LOCAL WINDS*

By FRIEDRICH DEFANT

University of Innsbruck

INTRODUCTION AND DEFINITION

The terms of the complete equation of atmospheric motion are determined by a number of forces which, through their interplay, cause the movements of the atmosphere. These effective forces are gravity, hydrostatic pressure, friction, and the Coriolis force.

If the Coriolis and friction terms are negligibly small, we deal with *Eulerian wind* equations in which the acceleration can be measured by the pressure gradient. However, if the Coriolis term is large as compared to the acceleration and friction terms, so that the pressure gradient is balanced only by the deflective force of the earth's rotation, we speak of *geostrophic winds.* As a third possibility, the friction terms may be so large that they are of the same order of magnitude as the pressure gradient term, and the Coriolis and acceleration terms may be neglected. In that case the wind blows approximately in the direction of the pressure gradient and it is called an *antitriptic wind* [42].

The antitriptic wind introduces the field of winds restricted to relatively small areas. Their range is of the order of 100 km or less. In this category belong the land and sea breezes, the mountain and valley winds, the jet-effect (*Düseneffekt*) winds, and, at least in so far as their characteristics are determined by the orography of the ground, the foehn and bora winds. These winds of locally restricted influence are classed under the general name *local winds.* They are an elementary yet meteorologically very interesting part of the movement of the atmosphere and a phenomenon which always attracts the attention of the layman.

In the following sections the basic principles of these local winds will be discussed. These winds very closely fulfill almost all conditions of the antitriptic wind since they blow roughly at right angles to the isotherms and isobars and are limited in the vertical to a relatively low altitude.

LAND AND SEA BREEZES

Description of the Phenomenon. In coastal areas, especially on tropical coasts and on the shores of relatively large lakes, we can observe in the course of a day the reversal of onshore and offshore winds, called land and sea breezes. The phenomenon takes the following course:

A few hours after sunrise, depending on location and season, a sea breeze develops, particularly on calm summer days. This sea breeze usually has a noticeable cooling effect, particularly in the tropics; it continues throughout the daylight hours and dies down around sunset. After that the seaward-blowing land breeze appears. This phenomenon, which is restricted to the coastal area proper and which extends only in tropical regions over a 100-km range, must be attributed to the difference in heat response between land and water. On warm, clear, summer days, the horizontal temperature difference between land and water provides the energy that leads to the development of vertical and horizontal air currents of strictly circulatory character. To the ground observer, only the horizontal parts of the circulation are noticeable; the vertical branches can be observed only indirectly, for example, through the formation of clouds over the land. The daytime sea breeze considerably surpasses in intensity the nocturnal land breeze. This is understandable because of the greater daytime arc of the sun during summer and the increased instability and consequently increased vertical austausch in daytime. The nocturnal cooling, on the other hand, produces an immediate stabilization of the air layers near the ground.

It is noteworthy that the direction of the sea breeze does not remain constant in the course of the day, and that this breeze often sets in with considerable gustiness, sometimes in the form of a protrusion similar to a cold front.

It is logical that the gradient wind, as determined by the over-all weather situation, should be superimposed on this local wind system and should at times obscure or even conceal it completely.

As an example of a typical day with land and sea breezes, the wind, temperature, and humidity recordings of the Danzig airfield, 3.35 km inland from the Baltic Sea coast, from June 3 to June 5, 1932, are reproduced in Fig. 1. This diagram clearly shows the onset of the sea breeze from the northeast and north-northeast, respectively, at 1420 on June 3, and at 1430 on June 5, while on June 4 a gradient wind from west-southwest–west-northwest completely conceals the sea breeze. After a calm from 1930 to 2245 on June 3, and from 2100 to 2220 on June 5, the land breeze sets in from southwest and west, respectively, and continues as a mild breeze until 0600 the following day. During the night of June 4–5, a calm from 0030 to 0130 and a shift in wind direction to the southeast are the only indications of the onset of the land breeze, but here also a gradient wind that subsequently appears hinders its development until 0600 in the morning. The temperature and humidity curves of Fig. 1 show characteristic irregularities in their normal trend at the onset and cessation of the sea breeze.

Figure 2 gives an example of the above-mentioned change in velocity of land and sea breezes in the course of the day, as observed at Hock van Holland on July 31, 1938. We can see from this figure the increase in the speed of the sea breeze up to the daily maximum between 1300 and 1400 and a concurrent steady shift

* Translated from the original German.

to the right, which will be explained later. The dying down and continued shifting to the right takes place in similar fashion.

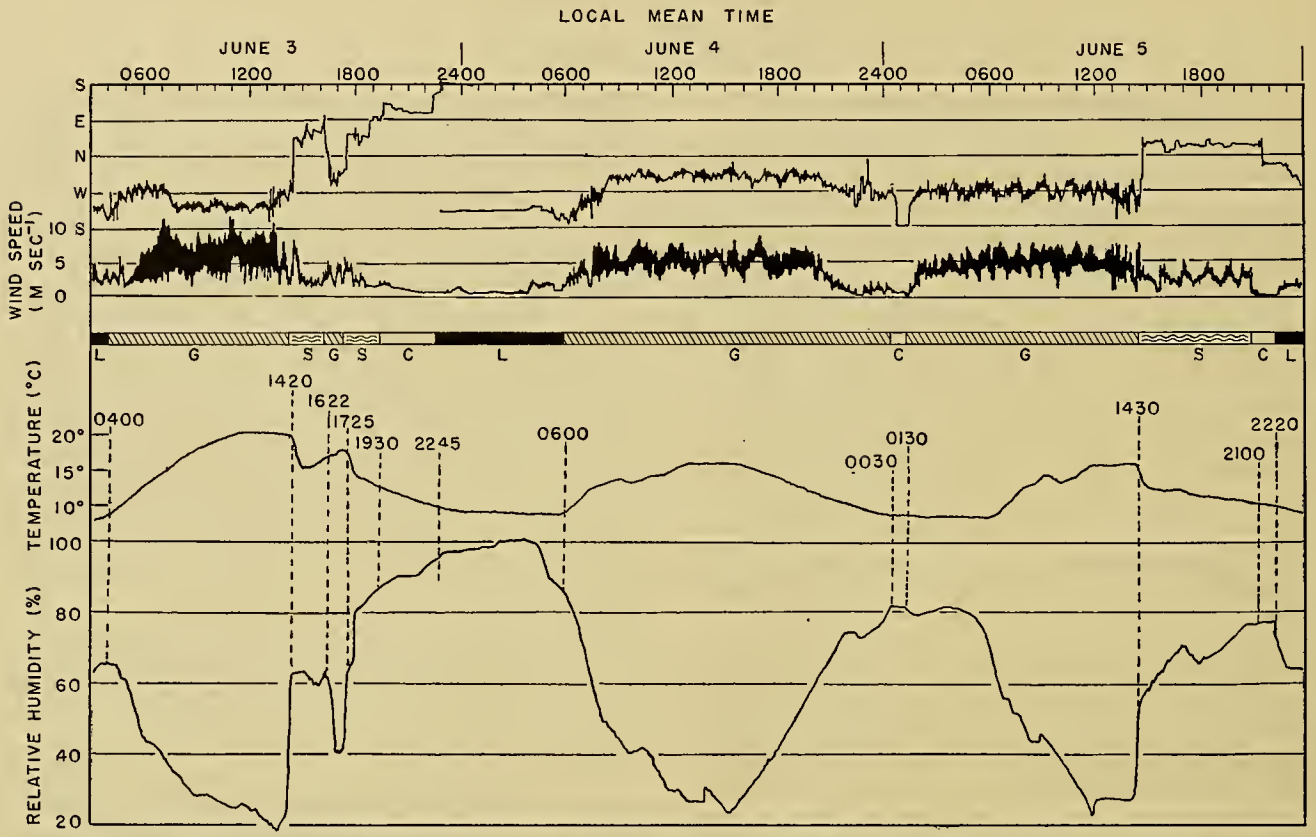


FIG. 1.—Registrations of wind speed and direction, temperature, and relative humidity during the characteristic land- and sea-breeze days from June 3 to June 5, 1932, at Danzig (L = land breeze, S = sea breeze, G = gradient wind, C = calm). (*After Koschmieder* [54].)

As an example of the tropical form of the land and sea breezes the diagram of velocity isopleths by van Bemmelen of the land and sea breezes at Batavia is reproduced in Fig. 3. It is also an example of the vertical velocity distribution during the course of a day.

Explanation of the Land and Sea Breezes. The cause of the land and sea breezes must undoubtedly be sought in the different behavior of land and water under the influence of an equal external heat supply. Water, as compared to soil, has a larger thermal capacity and its specific heat per unit volume reaches a value 40 per cent larger than that of soil. Although water should have a smaller temperature variation for the first reason, the amplitude of these periodic variations would be 1.414 that of the land for the second reason. This difference is equalized through the much smaller heat conductivity of water, and measurements reveal that the surfaces of water and sandy or rocky ground have temperature variations of comparable order. The depth to which radiation penetrates can also not be considered responsible for large temperature differences since the infrared radiation is immediately absorbed in the upper water layers.

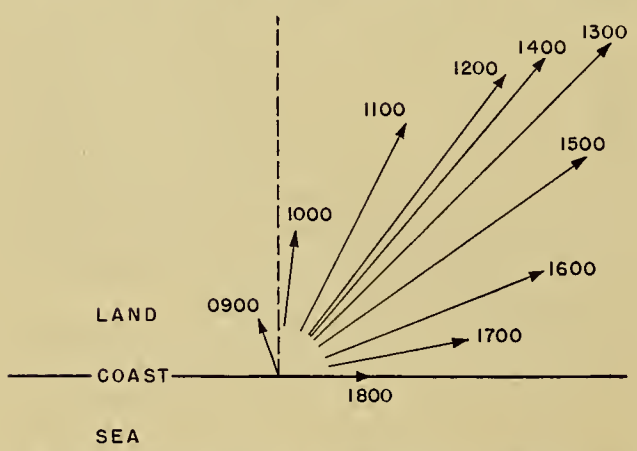


FIG. 2.—Hourly variation of relative wind velocity on a land- and sea-breeze day at Hoek van Holland as recorded on July 31, 1938. (*After Bleeker and Schmidt* [67].)

However, if we direct our attention to the turbulent mixing of the water by wind and waves, which effects a continuous downward transport of surface heat through large masses of water, we recognize that this mixing is the cause of the relatively small temperature variations. It is now clear that the complete absorption of radiant heat in the surface layers of the ground and the weaker influence of this form of energy on the deeper layers result in an entirely different thermal behavior of land and of water. Thus the temperature conditions of the ground are determined almost exclusively by its physical properties, whereas those of

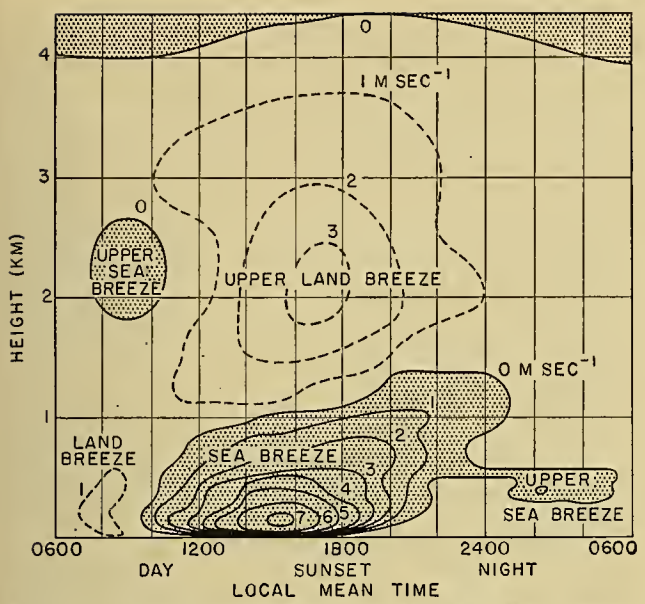


FIG. 3.—Velocity isopleths for the land and sea breeze in Batavia. (*After van Bemmelen* [70].)

the water are governed also by apparent conduction or turbulent mixing. In contrast with the strong heating of the air over the coastal region, the air over the strip of water offshore is only mildly warmed, and, as a result, a temperature difference between land and water develops. This difference diminishes toward sunset and reverses during the night.

Temperature Differences, Pressure Differences, and Pressure Distribution during Land and Sea Breezes. The maximum temperature differences (Δt) are given in the literature as follows: Kaiser [46] gives a range of Δt from 1.6C to 10.9C over a distance of 130 km between Wüstrow and Adlergrund Lightship (anchored in the middle of the Baltic Sea, about 100 km out of Swinemünde) as averages of a period of twenty summer days. Grenander [31] found differences in temperature between 3.6C and 7.6C at 1400, and between 1.4C and 3.1C at 0700 and 2100. These measurements, taken on the Swedish east coast, involved a distance between land and sea stations of 115 km. However, maximum values were as high as 10C or more; on the other hand, very small temperature differences occurred on some sea-breeze days. Measurements of Δt at lakes, as for instance at the Lake of Constance, likewise show a large range of temperature differences, namely values between 0.9C and 4C at 1400. We may conclude from these few measurements that Δt covers a wide range of values, and it is certain that a temperature gradient from sea to land exists in the morning hours. During forenoon a reversal takes place, and in the afternoon an increasing land-sea temperature gradient develops which is the driving force in the formation of the sea breeze. Over inland lakes, where opposite shores show this same behavior, the center of the lake must be a neutral zone. In general it may be assumed that the heating over land during daytime can reach a value five times that over water. Such temperature differences force the development of a pressure gradient and circulation system.

At the beginning of the day, the air pressure at higher altitudes over the land rises, while there is only a negligible increase over the water. As a consequence a drainage of the upper air from the land toward the sea takes place, and during forenoon the pressure close to the surface of the sea begins to rise, while it starts to fall over the land. This developing sea-land pressure gradient is accompanied by an air current in the same direction, that is, the sea breeze. A countercurrent, blowing toward land, is established at upper levels above the surface sea breeze. The circulation in daytime is completed by cumulus-forming convection over land and cloud-dissolving subsidence over water. In the evening the land and the overlying air cool faster than the sea and its overlying air, and a reverse nocturnal circulation develops. This circulation must be of smaller intensity and vertical extent because of the lack of instability and convection.

The land-sea pressure gradient near the surface at night and toward morning, as compared to a sea-land gradient during the day, is clearly discernible in Fig. 4.

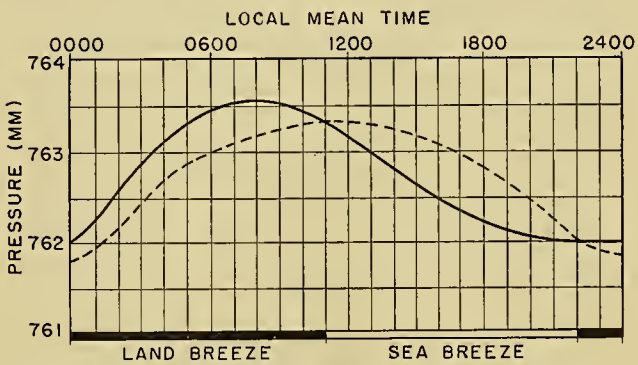


FIG. 4.—Average daily period of the air pressure on twenty sea-breeze days at the Baltic Sea. The solid line refers to Swinemünde; the dashed line, to Adlergrund Lightship. (*After Kaiser* [46].)

It is to be expected that the air current, according to the pressure distribution, is, at first, nearly at right angles to the coast line; only later, during the afternoon, a gradual change in the direction of the sea breeze appears because of the effect of the Coriolis force. However, the influence of the Coriolis force remains slight because of the short distances that these winds travel.

Conditions are somewhat complicated by the addition of a gradient wind associated with the over-all weather situation. A gradient wind blowing parallel to the coast line has no particular influence, since the pressure gradient causing it also acts normal to the coast line and thus only weakens or strengthens the pressure gradient associated with the land or sea breeze. However, a pressure gradient parallel to the coast line with, for example, an offshore gradient wind causes a mass transfer of air seaward. In that case a kink appears in the isobars where they protrude over the sea. Then the condition of gradient force plus Coriolis force equal to zero, which was valid in the previously discussed case, is no longer fulfilled, and unbalanced shoreward components of the gradient wind create a possibility for the development of a strengthened sea breeze. In simple schematic cases these conditions can also be

demonstrated numerically by superimposing the pressure fields of land and sea breezes and gradient wind [54].

Development of Land and Sea Breezes as a Function of Geographical Location, Season, and Time of Day. On tropical coasts, the land and sea breezes appear with great regularity [9, 11, 53, 70], because the clear sky there causes large variations in temperature over the land in the course of the day. Also the usually weak general air motion does not interfere with the development of local winds. It is only in India that the land breeze is completely obscured from May to September by the strong monsoon, which acts as a steady sea breeze during that season. Partial superimposition of mountain and valley winds on land and sea breezes may also cause peculiar wind conditions, as for instance on the coast of Samoa. In higher latitudes these local winds appear almost exclusively, or at least preferably, during the warmer seasons, since only then can sufficiently large temperature or pressure differences develop. In the cooler climates of higher latitudes, as for instance at the shores of the Baltic Sea [46], we can expect land and sea breezes, even in summer, on not more than about 20 per cent of the days. The role of solar radiation becomes apparent in a brief summary (Table I) of the probability of a sea-breeze day for different

Table I. Relationship between Cloudiness and Sea-Breeze Probability

Cloudiness (per cent)	0–50	60–80	90–100
Probability of a sea-breeze day (per cent)	90	39	27

amounts of cloudiness at the Black Sea. Little cloudiness and strong sunshine are decisive in promoting the occurrence of land and sea breezes. In polar regions the phenomenon disappears almost completely and occurs only once in a while on particularly clear summer days.

In the tropics (Batavia) [70] we find about 40–50 per cent of land-breeze occurrences and 70–80 per cent of sea-breeze occurrences to be fairly reliable values for the dry season. During the rainy season both frequencies increase; the land-breeze probability rises to between 60 and 80 per cent, that of the sea breeze to more than 80 per cent. Ramdas [63] gives frequencies for extratropical land and sea breezes in Karachi, India, (25°N) for every month of the year (Table II). We can

Table II. Annual Variation of Land- and Sea-Breeze Frequency at Karachi, India
(*After Ramdas* [63])

Month	J	F	M	A	M	J	J	A	S	O	N	D
Per cent	29	39	31	27	100	100	100	100	100	26	30	42

see from this table a 100 per cent occurrence during the summer months in contrast with the low percentage in the winter. Similarly, in etesian climates (as for instance the Mediterranean climate) spring has land and sea breezes on 31 per cent of the total number of days, the first half of June on 82 per cent, July on 91 per cent, and autumn on only 35 per cent of the days.

Finally, the phenomenon is in almost all regions a function of the time of day, since its periodic course is a consequence of the diurnal temperature variation. Usually, the sea breeze starts between 1000 and 1100, reaches its maximum velocity around 1300 to 1400, and subsides toward 1400 to 2000, whence it is replaced by the nocturnal land breeze. These approximate times naturally vary with the season and with climatic and local differences.

Intensity, Vertical Extent, and Range of Land and Sea Breezes. The height of the sea-breeze layer varies with climatic and local conditions. Its altitude ranges from 150 m at medium-sized lakes to 200–500 m at large lakes and the seacoast. Extending to 1000 m in moderately warm climates, the sea breeze reaches altitudes of 1300–1400 m in tropical coastal regions, as can be clearly seen in Fig. 3. In India, maximum altitudes of 2 km have been observed. In these areas, the nocturnal land-breeze layer is rather shallow by comparison. In Batavia, for instance, it reaches only to about 200–300 m. The difference in height between land and sea breezes is smaller in the temperate zones.

The intensities of the sea breeze cover the entire Beaufort scale. This breeze is of small force at lake and sea shores in the temperate zone (0 to 3 Beaufort); only at the seacoast do some peak values reach 4 to 5 Beaufort at noon. In the tropics, however, the wind may rise to storm intensity with the onset of the sea breeze. A particularly strong increase occurs on coasts with cold ocean currents offshore. While the horizontal speeds are of the order of meters per second, the vertical components are only of the order of centimeters per second.

The landward range of the sea breeze is estimated by many observers at 15–50 km in the temperate zones. Some values, for instance, are 16–32 km in New England, 20–30 km at the Baltic Sea, 30–40 km in Holland, up to 50 km in Jutland, 15 km on the Flemish coast, 40 km in Albania, more than 50 km on the northern coast of Java, and 40–50 km in Sweden. However, land and sea breezes are often augmented by mountain- and valley-wind effects which are difficult to separate from them. In tropical countries the sea breeze reaches 50–65 km, sometimes even 124–145 km into the interior. The seaward range of the much weaker land breeze appears to be everywhere much smaller. At the Baltic Sea, for instance, it extends only to 9 km.

Regarding the vertical temperature distribution in the temperate zone where condensation is relatively rare, nearly adiabatic or slightly superadiabatic gradients are to be expected. In the tropics, however, superadiabatic gradients are the rule, but probably reach only to the upper boundary of the sea breeze and decrease rapidly above it.

Conrad's Minor Sea Breeze, or Sea Breeze of the First Kind. The wind designated by Conrad [13] as "minor sea breeze" does not progress from the sea

toward the coast, but is formed on the boundary line between sea and land and progresses seaward. It is caused by the different heating processes on the flat, sandy beach. This minor sea breeze precedes the major sea breeze, the beginning of which is somewhat retarded by it until the air over the land has become considerably warmer than that over the water. The border line between warm land air and cool sea air, which has been displaced seaward by the minor sea breeze, is eventually overrun, and the cool sea air reaches the coast with an impact. This minor sea breeze, designated as a sea breeze of the first kind, will occur wherever strong pressure gradients connected with the large-scale weather situation do not hinder its development.

The Cold Front-Like Sea Breeze, or Sea Breeze of the Second Kind. The character of the sea breeze having a normal, front-free, and steady development is modified if a wind determined by the general weather situation hinders its regular course. This sea breeze, which develops in opposition to the gradient wind, is characterized by its retarded beginning (usually as late as 1500 to 1600), by its formation at sea and its slow progress toward the coast, and finally by a pronounced break-through of a cold front-like character (distinct gust with wind shift of 180 degrees). The development of this cold-air invasion can be considered to take place in the following way:

In the morning, the offshore gradient wind carries warmed air from the land out to sea and thus displaces seaward and weakens the pressure gradient between land and sea. An air-mass boundary is thus formed at sea against the cool sea air. The warmed land air, which accumulates in a nearly adiabatic layer, is highly turbulent and is able to carry along some of the stably stratified, cold sea air and is forced to rise with it. For a while, a stationary equilibrium between the two air masses may be maintained by an increasingly steepening frontal surface as is depicted in Fig. 5. However,

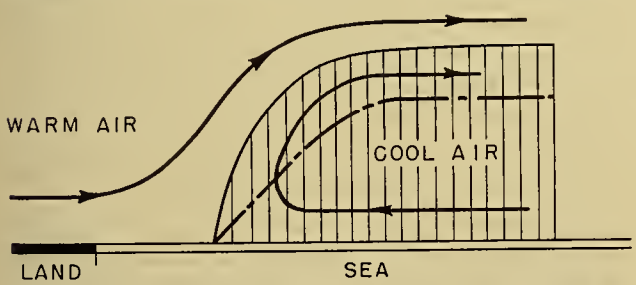


Fig. 5.—Stationary condition before outbreak of the sea breeze. The dash-dotted line is the boundary layer between currents of opposite direction in the cool air mass. (*After Koschmieder* [54].)

with further heating this equilibrium breaks down, the system becomes unstable, and the sea air now breaks through toward the land in the form of a front. This sufficiently explains the gusty onset and the cold-air character. If we consider that the largest vertical temperature gradients are not reached until 1200 to 1400, the retarded onset of the sea breeze is also understandable.

Theory of Land and Sea Breezes. A complete theory of the land and sea breezes must necessarily consider (in addition to the gradient force caused by land-water temperature difference) (1) the influence of the vertical turbulent heat exchange, (2) the turbulent friction of the air motion, and (3) the influence of the earth's rotation. In all existing theories these contributory factors receive only partially satisfactory consideration. Usually, the land and sea breezes are treated as simple, antitriptic currents caused by the unequal heating of land and water.

An older treatment of a stationary circulation is that by Jeffreys [42], who applied it primarily to the sea breeze and to monsoon winds. He considers friction, pressure, and Coriolis force and reaches a solution in which the daily wind change is in phase with the daily temperature oscillation. This result does not conform to reality. The application of this theory to the extended monsoon currents allows the calculation of the height of the monsoon reversal as well as of the amplitude of the surface pressure variations connected with the circulation. The agreement of these calculations with observation proves to be satisfactory.

V. Bjerknes and his collaborators [8] consider the land- and sea-breeze circulation a periodic current around isobaric-isosteric solenoids in which the wind is ninety degrees out of phase with the change in density. Accordingly, the sea breeze would start at the time of the greatest heating and would reach its greatest intensity at the time of the smallest temperature difference between land and sea in the evening. This is not in accordance with observation.

Kobayasi and Sasaki [52] as well as Arakawa and Utsugi [2] base their theories on Lord Rayleigh's convection theory [64]. They consider vertical and horizontal heat transfer in addition to turbulent friction of the air motion, whereby their basis for calculation becomes more complete than that of Jeffreys. Their solutions of the problem allow a comparison between theory and observation, although in order to reach complete agreement a heat conductivity one hundred times greater than that derived by Taylor from observations must be assumed. A further shortcoming of the theory is that the height to which the circulation (including the countercurrent) extends must be known, whereas it actually should result from the boundary conditions of the theory. This theory still neglects friction, which, as had already been pointed out by Godske [30], is responsible for the phase shift between density and wind-velocity variations.

A simple elementary theory of land and sea breezes is due to F. H. Schmidt [67]. In this theory he is less concerned with giving a complete explanation of the circulatory movement than with clarifying characteristic phenomena of the land and sea breezes, such as the phase shift between wind and temperature or the influence of the earth's rotation. A definite temperature distribution in accordance with observations is assumed, and the entire system of currents is calculated, taking the compressibility of the air into account. The Coriolis force is also introduced into the calculations, and thus

Schmidt succeeds in explaining quantitatively all facts of this atmospheric phenomenon in a satisfactory manner. Most certainly, his theory has greatly contributed to a better understanding of local winds.

Schmidt's assumption concerning temperature is characterized by the fact that the entire horizontal and vertical temperature distribution is given. He superimposes on the normal vertical temperature decrease the daily radiation temperature wave which is assumed to reach its maximum amplitude near the coast at a distance inland amounting to half the wave length of the temperature oscillation. This amplitude is supposed to decrease exponentially with altitude. However, for reasons of simplicity, he neglects the fact that the maximum amplitude is a function of time which, after all, should be of some importance to the theory. The variation in air density caused by the radiation temperature wave can be calculated; its amplitude also decreases exponentially with altitude. Part of the pressure variation is ascribed to the resulting density variation, and the remainder is caused by divergence and convergence of the compressible air. This influence of divergence is also assumed by Schmidt to decrease exponentially with altitude. If the pressure gradient is known, the horizontal velocity component can be calculated by application of the Guldberg-Mohn friction formula. Finally, the friction constant is assumed to be a quantity that decreases with altitude according to the expression e^{-rz}, where r is a constant representing the decrease of friction with height and z is the height. This is necessary in order to obtain sufficient variation with altitude of the sea breeze's starting time. We shall not discuss here how far such assumptions are justified; moreover, it would be preferable if the theory itself would yield the variations with altitude of all these quantities. Nevertheless, the theoretical determination of the deviation of the wind direction from the perpendicular to the coast and of the shift of the wind in the course of a day gives a satisfactory result.

The most recent work on the theory and observation of land and sea breezes has been published by Pierson [61]. In this work the theoretical considerations of F. H. Schmidt are considerably improved. Pierson makes assumptions regarding the temperature contrast between land and water that closely approximate reality and he takes into consideration not only the Coriolis force but also that type of friction which is used in the derivation of the Ekman spiral. A solution is given of the Navier-Stokes equations for laminar flow with consideration of the apparent eddy viscosity. Theoretical hodographs illustrate the variations of land and sea breezes under the influence of this type of friction, the Coriolis force, and the variable pressure-gradient force at all altitudes. The temperature contrast between land and water and its periodic variation during the course of a day as well as its variation with altitude are assumed, with due consideration for eddy diffusion (W. Schmidt, see [61, p. 9]), in a manner similar to that employed by F. H. Schmidt. From this the density and pressure distribution can be computed, and the wind field can be obtained from the equations of motion; the equation of continuity is unnecessary here. A comparison of the theoretical results with observations made at Boston (42°N), Madras (13.4°N), and Batavia (6°S) shows good agreement.

Recently, Haurwitz [37] made an interesting and important contribution to the theory of the land- and sea-breeze circulation. In contrast to previous investigators, he chooses Bjerknes' circulation theorem as his point of departure. With it he proves that the intensity increases, not only as long as the land-water temperature difference increases, but that it keeps growing until this difference disappears. Thus, the phase shift between temperature difference and wind maximum would be a quarter of the period, that is, six hours. Haurwitz shows that the introduction of frictional influences causes a considerable decrease of this phase shift which leads to a better agreement with observation. The daily shifting of the sea breeze, which was clearly observed in several locations, can be explained without difficulty as an effect of the Coriolis force. Haurwitz' work excels in its logical, all-inclusive consideration of all factors that are of importance for the development of land and sea breezes. However, as in the work by F. H. Schmidt, the theory of the land- and sea-breeze circulation is incomplete. In all these investigations the temperature contrast between land and water at the surface and at all levels above it is assumed. A complete theory should furnish the temperature distribution with altitude as well as the circulation from a given temperature contrast *at the surface.* The temperature distribution with altitude depends not only on the vertical turbulent heat exchange, but also on the circulation itself. For this reason, with a given boundary condition of temperature at the surface, the equation of heat conduction (as in the theory of the slope winds, see p. 666) must be incorporated in the theory.

If we approach the land and sea breeze as a single circulation cell in the sense of Lord Rayleigh's convection theory [64], we come considerably closer to the problem of these local winds. Furthermore, with this method we can take vertical as well as horizontal heat transfer and turbulent friction into account. The solution must yield not only the entire temporal development of the periodic current system, but also its dimensions as a function of heat supply or of the land-water temperature difference, respectively. I have recently attempted such a solution [18], which actually furnishes the required results in full conformity with observations. For the land-water temperature difference near the surface, which we have assumed as given, I used the simple harmonic function $\vartheta = Me^{i\Omega t} \sin lx$, where x is the normal to the coast, ϑ the potential temperature, M the amplitude of the temperature variation, $\Omega = 2\pi/(\text{sidereal day}) = 7.292 \times 10^{-5}$ sec^{-1}, and the length of the circulation cell is fixed as $L/2 = \pi/l$. This solution permits the utilization of any desired form of the land-water temperature difference, provided it is expressed by a Fourier series; the solution given above then holds for every one of its terms.

This solution, based on the assumption that ϑ is

proportional to sin lx, naturally yields a continuous chain of circulations. If a number of such circulations with variable l resulting from the Fourier series are superimposed on one another, we obtain a single circulation of definite horizontal extent. Thus, this extent becomes solely a function of the form of the land-water temperature difference. When the influence of friction (Guldberg-Mohn's friction equation, where σ is the coefficient of friction) and the Coriolis parameter ($f = 2\Omega \sin \phi$) are taken into consideration, the equations of motion become

$$\begin{aligned} \frac{\partial u}{\partial t} &= fv - \sigma u - \frac{1}{\rho}\frac{\partial \tilde{\omega}}{\partial x}, \\ \frac{\partial v}{\partial t} &= -fu - \sigma v, \\ \frac{\partial w}{\partial t} &= -\sigma w - \frac{1}{\rho}\frac{\partial \tilde{\omega}}{\partial z} + \gamma\vartheta. \end{aligned} \tag{1}$$

In addition to these equations we have the continuity equation in the form:

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0, \tag{2}$$

and the heat transfer equation which, with the omission of negligible factors, assumes the form:

$$\frac{\partial \vartheta}{\partial t} + \beta w = \kappa\left(\frac{\partial^2 \vartheta}{\partial x^2} + \frac{\partial^2 \vartheta}{\partial z^2}\right). \tag{3}$$

In these equations the y-axis is parallel to the coast, and the x-axis perpendicular to it, with the positive direction toward land. The letter ϑ stands for the potential temperature; $\gamma = g\alpha$, where α is the coefficient of heat expansion; β is the vertical gradient of the potential temperature; κ is the coefficient of turbulent heat conduction; and, as before, $\Omega = 7.292 \times 10^{-5}$ sec^{-1} expresses the day frequency. According to Rayleigh [64], $\tilde{\omega}$ is a quantity that fixes the perturbation pressure.

The solution may be assumed in the form:

$$\begin{aligned} u &= u(z)e^{i\Omega t}\cos lx, \qquad \tilde{\omega} = \tilde{\omega}(z)e^{i\Omega t}\sin lx, \\ v &= v(z)e^{i\Omega t}\cos lx, \qquad \vartheta = \vartheta(z)e^{i\Omega t}\sin lx. \\ w &= w(z)e^{i\Omega t}\sin lx, \end{aligned} \tag{4}$$

Then, by substitution in (1), (2), and (3), differential equations are obtained for u, v, w, $\tilde{\omega}$, and ϑ which are functions of z only and can be solved by exponential functions:

$$\begin{aligned} w &= Ae^{+az} + Be^{-bz}, \\ \vartheta &= Ce^{+az} + De^{-bz}, \end{aligned} \tag{5}$$

and similar expressions for u, v, and $\tilde{\omega}$.

The boundary conditions are

$$z = 0, \qquad w = 0, \qquad \text{and} \qquad \vartheta = M, \tag{6}$$

and for large values of z the circulation becomes negligibly small; then, the constants A, B, C, and D are fixed whereas a and b are determined by the day frequency and the other values given above.

The solutions for w and ϑ are:

$$\begin{aligned} u &= \frac{rM}{(a^2 - b^2)l}[ae^{+az} + be^{-bz}]e^{i\Omega t}\cos lx, \\ w &= -\frac{rM}{(b^2 - a^2)}[e^{+az} - e^{-bz}]e^{i\Omega t}\sin lx, \\ v &= -\frac{f}{i\Omega + \sigma}u, \\ \vartheta &= M\left[e^{-bz} + \frac{b^2 - s}{b^2 - a^2}\{e^{+az} - e^{-bz}\}\right]e^{i\Omega t}\sin lx. \end{aligned} \tag{7}$$

The factors r and s can also be expressed by the constants given above. It should be noted that all these quantities, including a and b, are complex numbers and that with $e^{i\Omega t}$ a separation of the real and imaginary terms would require extensive calculations.

When the solution (7) is worked out for latitude $\phi = 45°$ and various values of the friction coefficient σ, it yields circulation systems of the land and sea breezes that are in full agreement with observations. Table III lists the basic factors of the circulations for different values of σ with and without consideration of the Coriolis parameter. As is usually the case, the phases are referred to a maximum land-water temperature difference at 1200. It can be seen from Table III that the height of the land or sea breeze lies at roughly 400 m and rises with increasing friction from 320 m to 500 m. It is interesting to note that this altitude is somewhat reduced under the effect of the Coriolis force. Naturally, friction diminishes the velocity of the land and sea breezes to a considerable extent, namely from about 5.5 m sec^{-1} to 2 m sec^{-1} for every centigrade degree of temperature difference. For average friction conditions and a maximum land-water temperature difference of 5C over the distance $L/2$, we arrive at a maximum sea-breeze intensity of about 10 m sec^{-1}, which agrees quite well with observations. Under these conditions the vertical velocities reach maximum values of about 2 cm sec^{-1} per centigrade degree of temperature difference, which is also a reasonable value.

Of special interest is the phase of the land and sea breezes. If friction and the Coriolis force are neglected, the phase shift between the maximum temperature difference and the maximum intensity of the sea breeze is 4.7 hr. This shift decreases rapidly to 1.4 hr with increasing friction. In the case of $\sigma = 2.5 \times 10^{-4}$ sec^{-1} the maximum intensity of the sea breeze follows the maximum temperature difference closely, as is borne out by most observations. The influence of the Coriolis force on the component perpendicular to the coast is small, as is to be expected. Naturally, a cross velocity v appears instead, whose phase shift with the velocity perpendicular to the coast is 10.9 hr, or nearly 12 hr. This phase shift stays constant up to the height of the zero layer and then changes sign. Figure 6 is a vector diagram for the case $\sigma = 2.5 \times 10^{-4}$ sec^{-1} and $f = 1.031 \times 10^{-4}$ sec^{-1} ($\phi = 45°$). The figure shows clearly

the oblique position of the flow ellipse relative to the perpendicular to the coast, in good agreement with observations (see Fig. 2).

weak maximum that barely reaches a quarter of the intensity of the land and sea breezes near the surface. Theoretically, further circulations exist above this circu-

Table III. Circulation Characteristics Computed for Various Values of Friction Coefficient and Coriolis Parameter

σ $\times 10^{-4}$ (sec^{-1})	f $\times 10^{-4}$ (sec^{-1})	v Phase (hr)	v Amp. near surface (m sec^{-1})	u Phase (hr)	u Amp. near surface (m sec^{-1})	u Alt. of zero layer (m)	u Upper countercurrent Max. amp. (m sec^{-1})	u Upper countercurrent Alt. of max. wind (m)	w Phase (hr)	w Max. amp. aloft (cm sec^{-1})	w Alt. of max. wind (m)
0.0	0	12	M	16.7	5.43M	320	1.64M	650	16.6	4.21M	320
0.5	0	12	M	15.1	4.46M	340	1.11M	670	15.1	3.45M	340
1.0	0	12	M	14.1	3.68M	365	0.80M	700	14.1	2.86M	365
2.5	0	12	M	13.4	1.70M	500	0.26M	920	13.6	1.37M	500
2.5	1.031	12	M	13.1	1.84M	500*	0.25M	920*	13.1	1.76M	500*
				v							
2.5	1.031	12	M	2.2	0.72M	500*	0.11M	920*			

*Approximate.

Above the zero layer there is a countercurrent whose vertical extent exceeds that of the land- and sea-breeze layer by a factor of 4 to 5. This countercurrent has a

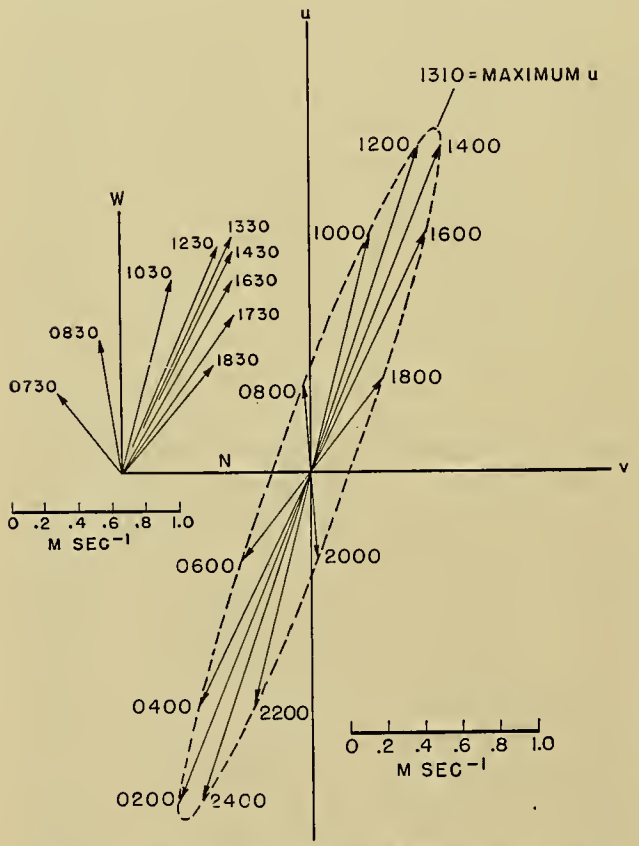


Fig. 6.—Theoretically calculated flow ellipse of the land and sea breeze under the influence of friction ($\sigma = 2.5 \times 10^{-4}$ sec^{-1}) and the Coriolis force ($f = 2\Omega \sin \phi = 1.03 \times 10^{-4}$ sec^{-1}; $\phi = 45°$) when the maximum temperature difference between land and sea is at 1200 LMT. The vector diagram at the left shows the mean winds during the sea-breeze period (0730–1830 EST) at Logan Airport, Boston, Mass. (based on 40 cases). (*After Defant* [18].)

lation cell of land and sea breezes and its countercurrent. However, their intensities are so low as to be negligible for all practical purposes. Thus, in contrast to others, this theory fixes a definite upper boundary to the land- and sea-breeze circulation. This upper boundary is, according to the theory, independent of the intensity of the land-water temperature contrast and a function only of the structure of the atmosphere, the Coriolis force, and the turbulent heat transfer. Friction tends to raise this boundary, while the Coriolis force tends to lower it.

MOUNTAIN AND VALLEY WINDS

The Mountain-Wind Circulation and Its Components. As in the case of coastal areas, local temperature and wind conditions occur in the vicinity of large mountain ranges that are often so strong in their effect that, locally, they modify or even obscure the general weather conditions. Here, thermal differences create a circulation system which, in daytime, consists of a lower current toward the mountains and an upper current in the opposite direction. In the region of the European Alps, Burger and Ekhart [12] have actually shown that this upper compensation current flows away radially from the mountains toward the neighboring plains in daytime. A corresponding flow system was found on the east slope of the Rocky Mountains by Wagner [71] and Ekhart [23]. This upper compensation flow is naturally less pronounced (speeds of the order of 15 cm sec^{-1}) than the lower current, since it is more strongly affected by the wind system determined by the general pressure distribution, the influence of which is difficult to separate from the local effect. Figure 7 shows these conditions schematically.

The Thermal Slope Wind. The difference in temperature between the air heated over the inclined mountain slopes and the air at the same altitude over the center of the valley causes the phenomenon of air rising in daytime along the slopes of moun-

tains, well known to every mountain climber. These winds start one fourth to three fourths of an hour after sunrise, and blow uphill in daytime. They

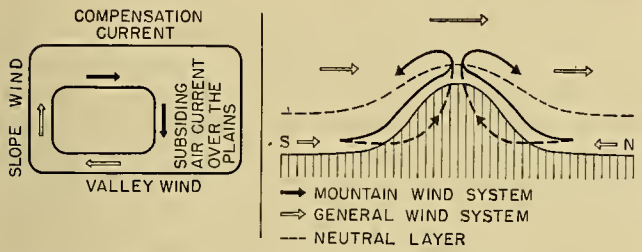


FIG. 7.—Schematic illustration of the air circulation during daytime in a cross section through the Alps. (*After Burger and Ekhart* [12].)

reach their greatest intensity at the time of maximum insolation and reverse their direction in the evening (about one fourth to three fourths of an hour after sunset). Because of the stronger insolation, they are especially well developed on the southern slopes and are weaker or almost nonexistent on the northern slopes. This wind prefers the ravines and gullies of the usually eroded slopes and is hardly noticeable on the projecting ridges. Numerous pilot-balloon observations in the up-and-down drafts on the slopes of the mountains north of the Inn valley near Innsbruck [43–45, 65] have clearly demonstrated the existence of such currents.

Here the thickness of the slope wind layer, as measured perpendicular to the slope, varies periodically with the wind intensity. Maximum values up to 260 m have been measured. However, as a rule, the thickness of the layer lies between 100 and 200 m. The thickness is less for the nocturnal downslope wind. The uphill wind continuously entrains, along its path, air from the space over the valley, so that the thickness of the affected layer is steadily increased in the direction of the uphill flow, and the layer becomes wedge-shaped. Naturally, the intensity of these slope winds varies greatly with the local differences in the slope and its exposure. Also, these winds can seldom be observed in their pure form and are often weakened or strengthened by extraneous wind conditions. On the average, the slope-wind speeds in the direction of the slope amount to about 2–4 m sec^{-1}, according to measurements. Projecting parts of the slope cause a detachment of this current from the slope and thereby an increased vertical movement which can be utilized by soaring pilots to gain altitude. In fact, the entire phenomenon of thermal slope winds is of greatest importance in soaring. The existence of thermal slope winds is often indicated by isolated cumulus clouds over summits or chainlike cumulus formations along ridges. Velocities of 13 m sec^{-1} have been measured in these updrafts. The nocturnal downslope wind shows a lesser vertical extent and lower velocities.

The highest velocity of these slope winds does not occur close to the slope surface, but at a definite distance from it. Higher up, it rapidly decreases again or is supplanted by other wind conditions. The current is extremely sensitive to changes in the insolation, and a temporary shading of the slopes will cause an immediate response by the wind. Also the difference in the starting time of the current is determined by the varying time at which insolation begins on slopes of different exposure. Thus, the phenomenon is often weakened at the time of the maximum temperature because of shadows cast on the slopes by extensive cloud formations.

Little is known about the thermal structure of the affected layer, because temperature measurements normal to the slope would be necessary for its actual determination. Such measurements have been made only up to about 20 m above the slope, since greater heights are difficult to reach. It is known that the layer of air adjacent to the slope shows strong superadiabatic gradients. During the cool downdraft at night, on the other hand, a strong increase of potential temperature (inversion) is often noticeable. The time span during the reversal of the wind is characterized by an isothermal air layer near the ground. At present, we have only a theoretical picture of the structure of the entire slope-wind layer, which, however, is probably very close to reality (see Figs. 11 and 12).

A variant of the thermal slope wind is the *glacier wind.* This wind, a shallow downdraft along the icy surface of the glacier, continues all day regardless of insolation. Its thermal cause is the continuously present temperature difference between the glacier surface and the free air at the same altitude. For this reason, the glacier wind has no diurnal period as does the slope wind. Since the wind is a function of this temperature difference, it reaches its maximum intensity and greatest vertical extent (between 50 m and 400 m) in the early afternoon, according to measurements by Tollner [69] and Ekhart [21]. The glacier wind always appears, even on cloudy days. In daytime, it fades out soon after leaving the glacier because its kinetic energy is dispersed by the ground friction. The glacier wind often collides with the upvalley wind and then slides under it. At night, after leaving the glacier, it blends into the mountain wind which has the same direction. A characteristic of the glacier wind is its strongly turbulent flow which, in a more moderate form, is also a feature of the nocturnal slope wind.

The Mountain and Valley Winds. The phenomenon of the daily wind change along the axis of large valleys is known in all mountainous countries. In daytime, from about 0900 to 1000 until sunset, an upvalley, or so-called valley wind blows. At night an opposite downvalley or so-called mountain wind appears which continues into the early morning hours after sunrise. Numerous investigations of this phenomenon, including soundings of its vertical structure, have been made in many mountainous countries [19]. Mountain and valley winds are best developed in the wide and deep valleys of the Alps. The shape of the valley's cross section and the inclination of the valley bottom are of little influence on these winds. As a matter of fact, in fairly level valleys, such as the valley of the Inn River in Tirol, these winds are particularly well developed. They occur most frequently during persistent high-pressure situations in summer and are thus a typical fair-weather

phenomenon of summer. Nevertheless, mountain and valley winds may also occur on cloudy days and in winter when they become manifest mostly in a modification of the general wind. The vector diagram of the winds in the Inn river valley near Innsbruck in Fig. 8 shows a good example of mountain and valley winds.

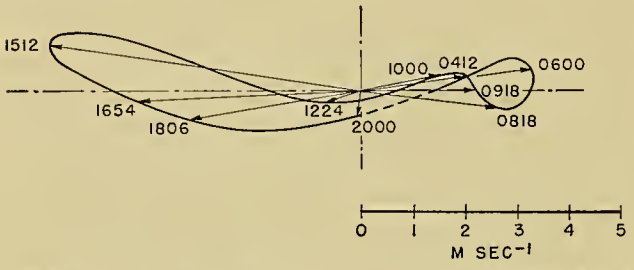


FIG. 8.—Vector diagram of the mountain and valley winds in the Inn valley (Innsbruck) on June 19, 1929, 100 m above the valley bottom. The dash-dotted lines indicate the axes of the Inn valley; upstream is to the left. Numbers indicate local mean time. (*After Ekhart* [19].)

This diagram of the wind conditions in the course of a clear day, June 19, 1929, shows that the maximum of the valley wind with speeds of 5–6 m sec^{-1} occurs shortly after 1500. The maximum of the mountain wind with speeds of 3–4 m sec^{-1} in the opposite direction occurs at around 0700. In general, the winds blow in the longitudinal direction of the valley's axis.

As an example of the height and intensity of the mountain and valley winds at Innsbruck, an isopleth diagram by Ekhart is reproduced in Fig. 9. The wind

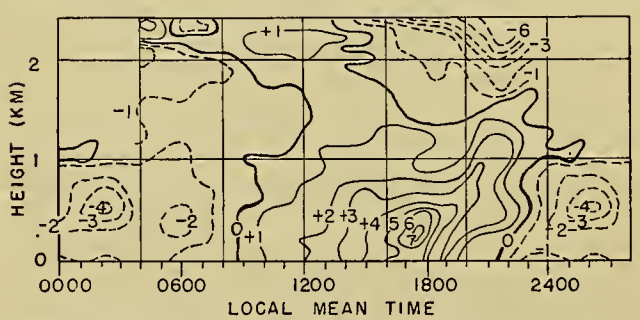


FIG. 9.—Mean velocity isopleths of the mountain and valley winds at Innsbruck (numbers are wind speed components (m sec^{-1}) in the direction of the valley; + upvalley, − downvalley). (*After Ekhart in Hann-Süring*, Lehrbuch der Meteorologie, *5. Aufl., p. 552.*)

velocity distribution shows the characteristic transition from mountain to valley wind. The values are averages of several fair summer days in 1929 and 1931. According to the diagram, the wind maximum is to be found at an altitude of about 200–400 m, while the velocities near the ground are reduced because of the influence of friction. This increase of the valley wind with altitude is characteristic of all valley winds, as is the fact that their average velocities are higher than those of the nocturnal mountain wind. Furthermore, the diagram shows that the valley wind starts almost simultaneously in all layers up to relatively high altitudes. The starting time of the valley wind is closely related to the width of the valley and thus to the size of the mass of air involved. This starting time changes with the season, that is, with the magnitude of the diurnal temperature variations.

The height of the valley wind usually reaches to, or somewhat above, the flanking mountain ridges. The more stably stratified mountain wind, on the other hand, is confined to lower levels. The upper boundary of the mountain and valley wind system usually arches several hundred meters over the mountain crests. There, the transition into the general wind system usually takes the form of an abrupt wind shift or a calm.

The Maloja Wind. The Maloja wind, named after the windshed between Engadine and Bergell, Switzerland, where it appears particularly well developed, is a variation of the mountain and valley wind [4, 6, 7, 10, 35, 36, 39, 50, 58, 59, 72, 74]. It is a mountain wind which blows downvalley both day and night. The phenomenon must be attributed to the fact that the valley wind of one valley reaches over a pass into another valley. In the mountain and valley wind system there, it acts as a disturbance, or, more specifically, as an abnormal development of the valley wind. The development of this anomaly is decisively determined by which one of the valleys involved has the larger diurnal temperature amplitude and thus effectively extends its circulation into the other valley across the pass. The question has not been satisfactorily answered whether this phenomenon can be entirely explained by the further increase of the diurnal temperature variation beyond the pass, or whether it is a purely inertial extension of that valley wind which is the more strongly developed. Sometimes, as was shown at the Arlberg Pass in Tirol [22], a strong upslope wind in one valley can, after crossing a pass, augment a valley wind in another valley.

Even in the absence of a pass, the interplay between thermal slope winds and typical mountain and valley winds sometimes produces considerable anomalies of the latter. Veering or backing of the valley wind on respective orographic slopes and cross circulation in narrow valleys, owing to strong insolation on one slope, have been observed. Cyclic wind shifts in the course of a day are caused by the fact that the upslope wind starts before the valley wind, and the downslope wind before the mountain wind.

The Theory of Mountain and Valley Winds. The development of the theory through many decades [15, 27, 33, 34, 50, 66, 74] finally culminated in the work by Wagner [71–73], who made an intensive study of the daily pressure and temperature variations in the free atmosphere within the valleys of the Alps. These investigations led to the result that at a certain altitude above the valley, usually at about the height of the surrounding ridges, the otherwise different diurnal pressure variations are completely equalized. This level of equalization was designated by Wagner as the "effective ridge altitude." A further important result of these investigations was establishment of the fact that the diurnal temperature variations in the valleys up to the

height of the ridge were more than twice as large as the variations within a similar layer over the plain. As a consequence, a pressure gradient from the plain to the valley must exist during the day, whereas the reverse gradient must appear at night. The equalization of pressure differences at the effective ridge altitude requires that the pressure differences be largest close to the valley bottom and that they decrease with height so as to disappear completely at the effective ridge altitude.

From these deductions an adequate theoretical explanation of the mountain and valley winds can be derived. In full accord with observations, the unequal rates of decrease in the amplitude of the temperature variations over the plain and the valley bottom create a pressure gradient of the observed magnitude whose direction is toward the valley during the day and away from it at night. The air current thus generated fills the entire valley and decreases with altitude in accord with observations. Thus, there is a thermal cause for the pressure differences, and the air current is theoretically established as a circulation with a reversal of direction at night. This theory also includes the observed circulation between the plains and the mountains and the necessary existence of the upper compensation current. Furthermore, the pressure gradient, according to Wagner, is the result of the combined effects of the inclinations of valley bottom and ridge line. A nonparallel course of these inclinations, such as a downslope of the valley bottom and simultaneous upslope of the ridge line, would cause an extension of the existing gradient. In other words, if the ridge line rises beyond the pass altitude toward the neighboring valley, the gradient existing in the first valley would run in the same direction also in that neighboring valley beyond the pass. In this case the wind would reach over the pass as a Maloja wind. Thus, this abnormal phenomenon of the mountain winds also fits into Wagner's theory.

This theory furthermore requires two wind systems to preserve the stationary condition in the valley. One of them consists of the horizontal inflow of the valley wind, produced by the static conditions; the other one is the thermal slope wind system which takes care of the outflow over the flanking ridges. The combined wind systems have an additional effective source of energy in the heat given off by the heated slopes. The slope winds have a further important function in that they continuously add heated slope air to the air over the valley through that branch of the slope-wind circulation which descends over the valley center. Thus, the temperature in the center of the valley cross section is continuously increased during the day and decreased at night, when the heating conditions are reversed.

The mechanism interlocking the upslope and downslope winds with the mountain and valley winds in the course of a day, described in great detail by Wagner [73], is shown in Fig. 10. The great importance of the thermal slope winds as integral members of the larger circulation between plains and mountains is obvious. On the basis of a theoretical treatment of the stationary slope wind by Prandtl [62], the present author recently extended these considerations to the nonstationary case [17].

Prandtl's theory has particular significance, because a deeper insight into the mechanism of the thermal slope-wind current was gained through the introduction of turbulent heat conduction and turbulent friction. A disturbance of the temperature field, as well as static instability, always appears in the air layer above a heated surface and thus tends to establish turbulence.

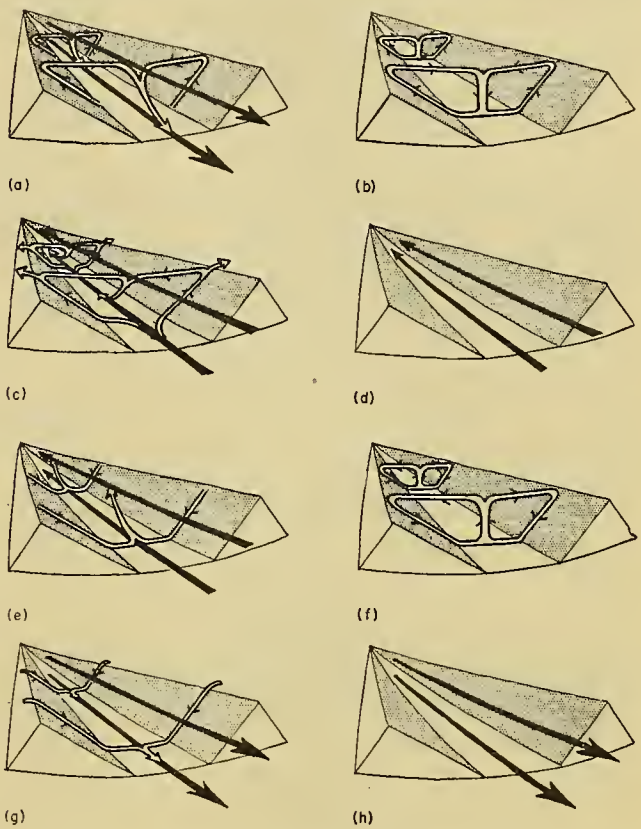


FIG. 10.—Schematic illustration of the normal diurnal variations of the air currents in a valley. (*After F. Defant* [17].)
(*a*) Sunrise; onset of upslope winds (white arrows), continuation of mountain wind (black arrows). Valley cold, plains warm.
(*b*) Forenoon (about 0900); strong slope winds, transition from mountain wind to valley wind. Valley temperature same as plains.
(*c*) Noon and early afternoon; diminishing slope winds, fully developed valley wind. Valley warmer than plains.
(*d*) Late afternoon; slope winds have ceased, valley wind continues. Valley continues warmer than plains.
(*e*) Evening; onset of downslope winds, diminishing valley wind. Valley only slightly warmer than plains.
(*f*) Early night; well-developed downslope winds, transition from valley wind to mountain wind. Valley and plains at same temperature.
(*g*) Middle of night; downslope winds continue, mountain wind fully developed. Valley colder than plains.
(*h*) Late night to morning; downslope winds have ceased, mountain wind fills valley. Valley colder than plains.

Let us assume the spatial distribution of the potential temperature ϑ, together with the temperature disturbance which stems from the heat transfer along the heated slope, in the form:

$$\vartheta = A + Bz + \vartheta'(n), \tag{8}$$

where A is a constant at $z = 0$, B is the lapse rate of potential temperature, z is the vertical, n is the normal to the slope, and ϑ' is the disturbance in potential temperature caused by heat conduction in the direction n. We can then expect the velocity w of the stationary current along the slope to be a function of n only.

The interplay of heat transfer and heat conduction furnishes the differential equation,

$$g\beta\vartheta' \sin\epsilon + \frac{\nu\kappa}{B \sin\epsilon} \frac{\partial^4\vartheta'}{\partial n^4} = 0, \qquad (9)$$

where β is the coefficient of expansion, $g\beta\vartheta' \sin\epsilon$ is the acceleration in the upslope direction, ϵ is the angle of the slope, ν is the coefficient of turbulent friction, and κ is the coefficient of turbulent heat conduction.

The usual solutions of equation (9) are

$$\vartheta' = Ce^{-n/l} \cos\frac{n}{l}$$

and

$$w = C\sqrt{\frac{g\beta\kappa}{B\nu}}\, e^{-n/l} \sin\frac{n}{l}, \qquad (10)$$

where

$$l = \sqrt[4]{\frac{4\kappa\nu}{g\beta B \sin\epsilon}},$$

and C is the temperature disturbance at the surface of the slope. The actual form of the solution is shown schematically in Fig. 11.

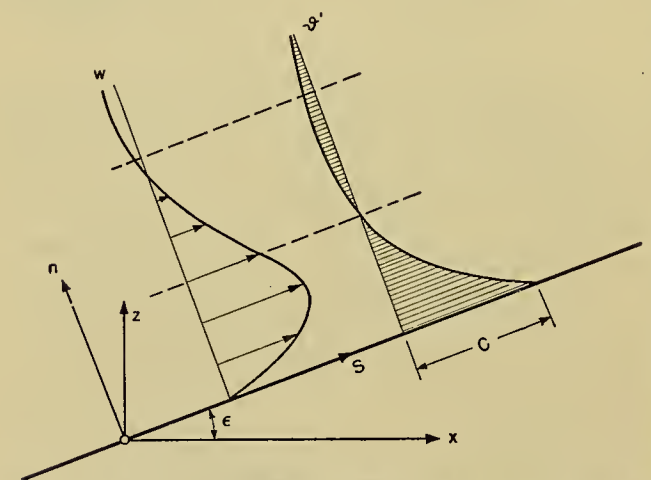


Fig. 11.—Schematic profiles (normal to the slope) of the wind speed w and temperature ϑ'. (*After Prandtl* [62].)

I have checked the correctness of these theoretical results by a detailed comparison with actually measured average values of ϑ' and w and have found a splendid agreement, except for the disturbing gradient influences in the upper layers. Furthermore, it is noteworthy that the resulting values of the austausch due to impulse transport (virtual friction) and of that due to transport of the heat content (virtual heat conduction) are of a plausible magnitude. Figure 12 presents the theoretical distribution of the potential temperature during upslope and downslope winds, respectively, for which no observations are available. The characteristics of the stratification become very apparent.

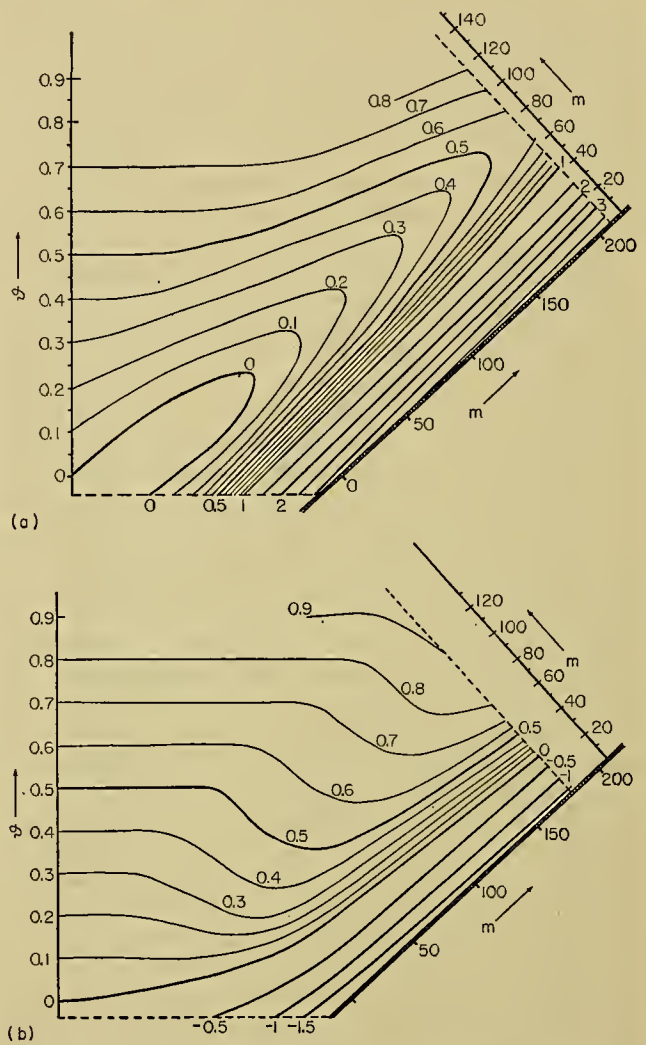


Fig. 12.—Theoretical distribution of the potential temperature over a mountain slope during (*a*) upslope wind and (*b*) downslope wind. (*After F. Defant* [17].)

The theory can be extended to include the oscillatory nature of the slope wind by simply multiplying, as a first approximation, the solutions (10) of the stationary case by the factor $\cos \Omega t$. Because of the fourth root in the expression for l, variations in the values of ν and κ [17, pp. 441–444] are of little influence on the solution.

The average air transport by the slope winds can be roughly computed from the calculated and the observed average profiles of the slope wind. We can then estimate how long it would take until the rising heated slope air has replaced the air masses over the valley bottom. The interest of this question becomes apparent if we consider that this process of warming the air in the valley center has an effect on the pressure gradient and thus on the generation of the mountain and valley winds. Calculations show that a slab of air 1 m thick, 1500 m long (the width of the valley), and 1750 m

high (the height of the valley) is completely replaced in a closed circulation by heated slope air in 4½ hr.

Thus, if we assume the heating to begin at 0830, the maximum pressure gradient between plains and valley would occur at about 1300. This time would be the beginning of the main phase of the valley-wind development. A comparison with the time of the valley-wind maximum at Innsbruck (between 1500 and 1600) proves satisfactory if we consider the distance traveled by the air from the valley entrance to Innsbruck. These results confirm the importance of the return of the upslope wind to the valley center for the theory of the mountain and valley winds, as postulated by Wagner.

DISTURBANCES OF THE WIND THROUGH OROGRAPHIC INFLUENCES

The Foehn. In almost all mountain areas, conspicuous local winds can be observed which blow from the ridges down into the leeward lowlands. At times these winds can assume gale intensity. Their chief characteristics, after reaching the plains, are abnormal temperature rise and dryness, which are of climatic significance over a more or less wide belt in the lee of the mountains. These local winds have become known under various names; however, today the general term *foehn* is used.

Although particularly well developed in the European Alps, especially in Switzerland and Tirol, such foehn winds blow also in Greenland from the inland ice over the mountain ranges down to the coast. In North America the foehn, known as the *chinook*, has a climatic influence on a very wide belt east of the Rocky Mountains. In Argentina, this wind is known as the *zonda* and blows down from the Andes. Local winds with foehn characteristics are also well known in Japan, New Zealand, and in eastern and central Asia. There is hardly a mountain range where the foehn is completely lacking. After all, the foehn will appear wherever prevailing winds must pass over a mountain barrier. Such barriers thereby become great divides of weather and climate.

The foehn is also noted for characteristic weather elements other than high temperature and low humidity. There is always extraordinarily good visibility; the mountains appear unnaturally close and clear and assume steel-blue to purplish hues. The clouds associated with the foehn are lens-shaped (*lenticularis*) at medium altitudes, and elongated banks often suggest wave formation. The peaks of the mountains and the upper part of the windward slope are shrouded in rather flat, cumuliform clouds. These clouds, the so-called "foehn wall," are in a continuous process of forming and dissolving, because they remain stationary in spite of strong winds.

The foehn is a gusty wind, and temperature and humidity curves therefore always show irregularities to a varying degree during a foehn. In the Alps the foehn is always strongest where north-south valleys open into the plains or into large east-west cross valleys. This latter form is characteristic, for instance, at Innsbruck. It was there that the investigation of this phenomenon was most intensively pursued (see, for example, the work by Trabert, v. Ficker, A. Defant, Ekhart, and Pernter [14, 20, 24, 26, 60]).

During the winter, the great increase in temperature causes a rapid melting of the snow. However, floods seldom occur because the melting is very localized and decreases rapidly with altitude. Also, during a foehn, evaporation is very rapid because of the low relative humidity, and precipitation is confined to the passes, the peaks, and the windward side of the mountains. This distribution of precipitation is the chief characteristic of the foehn and has decisive climatic consequences for the areas on both sides of the range.

The thermodynamic explanation of the foehn is essentially due to Hann [32]. Since the theory belongs to the basic principles of theoretical meteorology, the reader is referred to pertinent textbooks. In general, it can be said that observations are in good agreement with Hann's theory: On the windward side, cloud formation (particularly the stationary foehn wall) and precipitation must occur at a certain altitude as a result of the condensation. This removal of moisture together with the compression of the air during its descent on the lee must cause dissolution of clouds, warmth, and dryness there.

When air flows across a mountain range it is subjected to the above-mentioned foehn processes, as explained by Hann, and exhibits foehn characteristics on the lee side of the mountain. Such an air flow normal to the mountain ridge can be maintained only by a pressure gradient that is parallel to the ridge. Accordingly, such an air flow exists only when the general weather situation has a very definite pressure pattern, as shown schematically in Fig. 13. The sinusoidal deformation of the isobars, characteristically produced by the thermal pressure effect, forms a bulge of the isobars toward the high pressure on the lee side of the mountains and toward the low pressure on the windward side (often called the "foehn nose").

Depending on the orientation of the mountains, the air currents, after passing through the thermodynamic process, show the foehn phenomena to a greater or lesser extent. In the case of mountain ranges oriented from west to east (*e.g.*, the European Alps or the Pyrenees) a south wind will blow across the mountains if the low pressure is to the west and the high pressure to the east (see Fig. 13*a*). Because of its original warmth, the air from the south is well suited to the development of very marked foehn phenomena. For this reason, the *south foehn* is the most striking type of foehn from the viewpoint of the meteorologist as well as of the layman (see v. Ficker [28, pp. 25–37]). If low pressure lies to the east and high pressure to the west, air from the north is transported across a mountain range oriented west-east (see Fig. 13*b*). The cold air piles up to the top of the mountains and then descends on the lee side while undergoing the thermodynamic process. This air, however, exhibits foehn properties only when the foehn process changes its cold-air character to such an extent that, upon arriving on the lee side of the mountains, it

is markedly warmer and drier than the air that was there before. This type of foehn, called *north foehn* [25; 40; 29, pp. 48–53], is relatively rare.

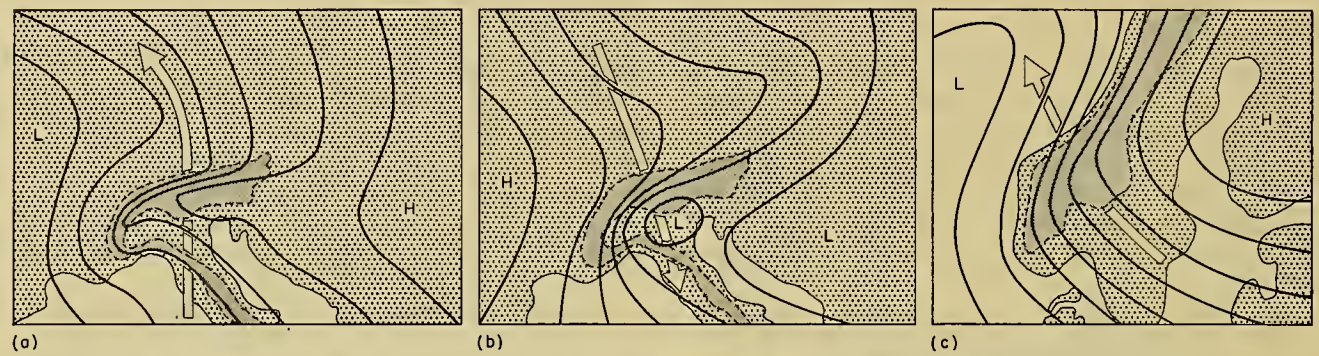


FIG. 13.—Windward and leeward effects of mountains on the isobaric pattern at sea level (schematic): (*a*) south foehn in the European Alps, (*b*) north foehn in the European Alps, and (*c*) southeast current over the Scandinavian mountains.

In the case of a north-south orientation of a mountain range, the development of a foehn requires low pressure to the north and high pressure to the south, or vice versa. Then, a west-east current passes across the mountains (*e.g.*, the Rocky Mountains, the Andes, or those of Greenland, New Zealand, and Scandinavia) (see Fig. 13*c*). On the lee side, considerable temperature differences in the meridional direction occur; for example, in western Canada the broad warm foehn current, after passing the Rocky Mountains, meets the extraordinarily cold Canadian polar air. The result is increased cyclogenesis or rapid deepening of existing cyclones, which must be considered a consequence of the foehn process. This phenomenon, which is to a much lesser extent associated with the south or north foehn, has perhaps an analogy in the formation of the Genoa cyclone south of the Alps and the Apennines.

The thermal pressure effects during foehn cause in all cases a great increase of the horizontal pressure gradient in the direction of the mountains (as much as 9 mb per 100 km at sea level or at the level of the valleys). The mountain barrier prevents the transformation of such gradients into air motion; or, in reverse, the establishment of such gradients is made possible only by the mountain barrier. Although this gradient is less strong at the level of the mountain ridges, it cannot be explained by the thermal contrast alone.

The factors determining the various types of foehn are the general synoptic situation, the orientation of the mountain ranges, and the type of air masses passing over the mountains. In turn, the thermodynamic foehn process causes changes in the pressure field and in the properties of the air masses involved, which affect the general weather situation.

The warming and drying of air, observed in anticyclones, bear great similarity to the foehn and are caused by the dynamic heating during the descent of air from aloft. As with foehn, clouds dissolve and fair weather without precipitation sets in. In these cases, warm air is always present aloft as is evident from observations at mountain stations or from aerological soundings. This warm air rests on the cold air that lies in the valleys and in most cases shows up as a sharp inversion and decrease in humidity. The descending motion is very slow and must be ascribed to a divergent flow at the surface, directed away from the mountains. This phenomenon is called *high foehn* or *free foehn* in the literature [28, pp. 53–58]. Since an anticyclonic situation usually precedes the south foehn, the free foehn over the mountains often forebodes the subsequent development of a foehn current.

The warm foehn current on the lee of the mountains descends into the lowlands rather than ascends as one would expect. The cause of this descent into the valleys is a meteorological problem. Following older concepts, v. Ficker [28, pp. 34–37] has answered this question by resolving the foehn development into several phases (preliminary phase, anticyclonic phase, stationary foehn phase). A period of foehn is generally preceded by an anticyclonic weather situation as the *preliminary phase*, with very stable vertical temperature distribution. Cold air lies in the valleys, with dry warm air above it and separated from it by an anticyclonic subsidence inversion. Before the onset of the foehn, the cold air moves out of the valleys and away from the mountains under the influence of the pressure distribution. The upper boundary of the cold air is thereby lowered, and warm air from aloft supplants it. When this warm air reaches a station, the *anticyclonic phase* sets in for this station. The temperature rise and humidity drop connected with very little air motion indicate the anticyclonic foehn. Thus, the warm air does not actually break through the cold air to the valley floor, but follows the cold air which must first flow out of the valley into the plains while its upper boundary subsides. Only when the foehn wall forms on the windward side of the mountains and the horizontal pressure gradient is increased in the direction of the mountains, does the *stationary foehn phase* set in. In this phase, the foehn proceeds as explained by Hann.

The foehn does not always reach the valley floor as a warm current. If a remnant of shallow cold air remains in the lowlands, the foehn current moves over it. Occasionally, some cold air remains in certain parts of the valley, and the warm air reaches the valley floor only at certain *foehn islands* which are of great climatic significance. Also at night, when the intensity of the

foehn current diminishes and the cold air remnants in the valleys are augmented, these remnants may again combine into a shallow cold-air layer covering the entire valley floor. The foehn current then lifts from the valley floor, a fact that shows up in meteorological recordings as marked temperature and relative humidity jumps called *foehn pauses*. Figure 14 shows a particularly good example of the temperature and humidity records during the foehn period from February 2–5, 1904, at four alpine (Inn valley) stations at different altitudes. In this figure all features of the individual foehn phases, foehn pauses, and the characteristic response of the stations to the south foehn are evident.

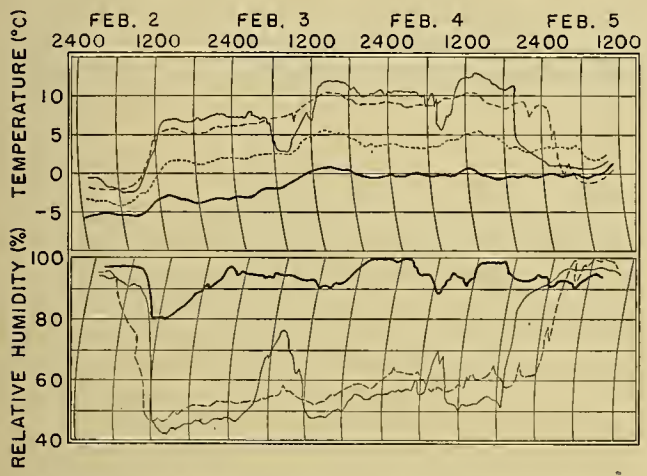


Fig. 14.—Typical south foehn registrations from the Innsbruck foehn area from February 2 to 5, 1904. Thin solid line—Innsbruck (573 m); dashed line—Igls (876 m); dotted line—Heiligwasser (1240 m); thick solid line—Patscherkofel (1970 m). (*After v. Ficker* [24].)

According to A. Defant, remnants of cold air in valleys form closed systems and may be excited into oscillations (similar to waves on lakes) by the passage of a foehn current over them. These oscillations show up as temperature fluctuations that are not to be confused with the short, periodic pressure fluctuations that occur during foehn.

Of special interest is the wave form of the air flow in the free atmosphere on the lee side of hill chains, mountain ridges, or other extended elevations of the ground. When the wave crests of this leeward current reach above the condensation level, the stationary wave becomes visible as a fixed "Moazagotl cloud" on the lee side of the mountains. The Moazagotl cloud consists of one or more cloud banks parallel to the mountain barrier; it forms continuously on the windward side of the wave and dissipates on the lee side. Sailplanes may reach great altitudes in the Moazagotl updraft. Prandtl, Küttner, and recently Lyra [55] have conducted theoretical investigations of these Moazagotl or foehn waves in the lee of mountains. In the development of the theory an increasing number of factors have been taken into consideration; Lyra's recent extension of the theory to include any polytropic stratification brought the theory of sinusoidal air currents on the lee side of barriers to a preliminary conclusion. Figure 15*a* shows the theoretically computed streamline pattern over a mountain barrier. Such foehn waves with their attendant cloud systems have been observed over numerous mountains. The waves caused by the Rocky Mountains in the northwestern United States have been investigated in detail by Hess and Wagner [38]. A particularly interesting case, in which the waves on the lee side reach an altitude of 40,000 ft, is shown in Fig. 15*b*.

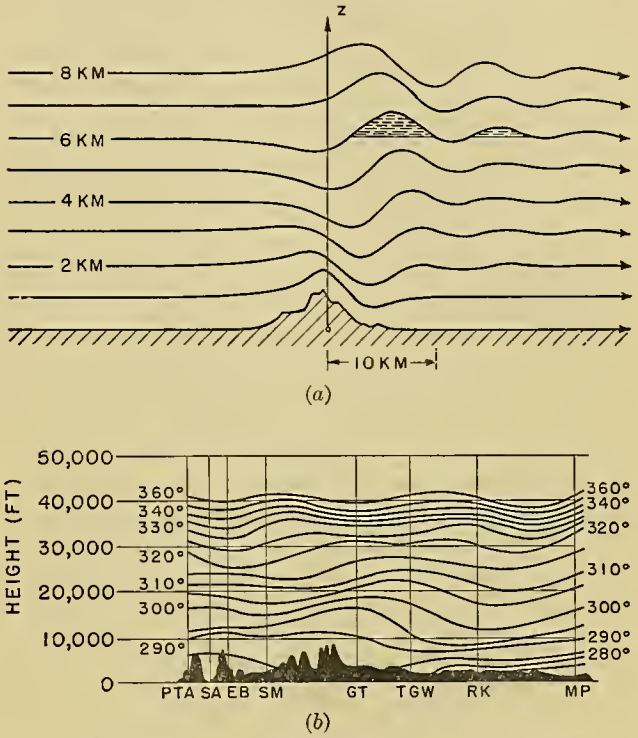


Fig. 15.—Stationary waves in the free atmosphere on the lee side of mountains. (*a*) Streamlines across a mountain range with Moazagotl clouds on the lee side. (*After Lyra* [55].) (*b*) Cross section from Tatoosh Island, Wash., (PTA) to Minneapolis, Minn., (MP) for 0300Z, January 14, 1945; isolines of potential temperature in degrees absolute. Vertical scale: units of pressure-altitude (ft) of U. S. standard atmosphere. (*After Hess and Wagner* [38].)

During well-developed foehn currents, marked physiological disturbances in men and animals occur on the lee side of mountains. These disturbances are ascribed to the short-period pressure fluctuations during foehn [28, pp. 65–105], but their causes have not been completely explained as yet.

The Bora. Hann's theory, which includes all fall-wind phenomena in the mountains, must also explain the cold fall-wind phenomenon. If a very cold air mass passes over a mountain barrier, or if it blows from the interior of a cold land mass over a plateau, the dynamic temperature rise during its descent on the lee side may not suffice to turn this fall wind into a warm foehn. Rather, this wind will be cold upon its arrival in the plain and is called *bora*, after the best-known of these local winds, which occurs on the coast of Dalmatia [48, 49, 57]. There, the cold, continental outbreaks of winter air from Russia find their way through Hungary and over the mountains of Dalmatia to drop over a

low, but relatively steep, slope to the otherwise warm coast of the Adriatic Sea.

The outstanding characteristic of the bora is its extraordinary violence, which often causes heavy damage. The cause of these stormlike fall winds is the liberation of potential energy when the cold air begins to drop to the sea. There are calms between violent gusts of 50–60 m sec^{-1} called *reffoli*, and barometric fluctuations of 4 mm *Hg* are not rare in these cases. The bora does not extend far over the sea. However, it induces a heavy sea and atomizes the wave crests to such an extent that a cloud of mist (*fumarea*) is formed over the ocean. The bora of the Adriatic Sea has a pronounced diurnal period of intensity and frequency of occurrence. The maximum intensity occurs between 0700 and 0800, the minimum around 2400. It occurs most frequently about 0600 to 0700, most rarely about 1400.

We may distinguish between cyclonic and anticyclonic bora, according to the origin. The cyclonic bora is dependent on the existence of a low-pressure area over southern Adria, with an especially warm sirocco blowing from the south over the front portion of this low and aloft over the bora. Consequently, the sky is usually cloudy, and there is precipitation. The bora then blows more steadily and covers the entire Adriatic Sea. For an anticyclonic bora to develop, a strong high-pressure area must exist over central Europe with an extension of high pressure over Dalmatia which need not be opposed by a cyclone with closed isobars in the south. In that case the bora exhibits a violent character, but does not extend far out to sea (about 10 nautical miles). The sky remains clear, with the exception of a foehn wall over the mountains.

Under favorable conditions, well-developed local winds of bora character occur in many areas in the world. These conditions are (1) a short topographic drop, and (2) a sharp climatic division between a cold plateau and a warm plain so that, under suitable pressure conditions, cascading of the deeply chilled continental air can take place. Especially well known are the bora winds of Novorossiisk on the northern Caucasian shore of the Black Sea [1, 3] and those of Novaya Zemlya [75].

A. Defant [16] has dealt theoretically with the question of the drainage of cold air along a slope. His method of attack, in which he bases a somewhat schematic model on the equations of motion and continuity in a doubly stratified atmosphere, allows a very good estimate of the influences of gravity and pressure gradient on this drainage of cold air masses. It is interesting to note that these two influences are of almost the same order; in other words, not only gravity, but also the pressure gradient, has an influence on the drainage of the cold air. If the steady current is disturbed, it may remain stable up to a critical slope angle, in which case small initial disturbances propagate as waves with a definite period and eventually fade out under the influence of friction. If the critical angle is exceeded, small initial disturbances grow exponentially with time, and the current assumes an unstable and turbulent character. A ground slope of 1:100 seems to be the critical value, fully confirmed by observation. An example of a stable current is the orderly, nocturnal outflow of the mountain wind through slightly inclined valleys; one of unstable flow is the extremely turbulent bora in Dalmatia.

The Mistral and Jet-Effect Winds. The bora-like local wind of Provence and the French coast of the Mediterranean up to Perpignan [5, 29, 51] is called the *mistral*. It is a combination of a bora and a wind that is increased through the so-called jet effect. The mistral is due to the drainage of cold air initiated by a high-pressure ridge which frequently extends from the Azores to France, and by a permanent low-pressure area over the warm Gulf of the Lion. Its intensity is increased and its duration prolonged through the constriction of the geographic gate between the Pyrenees and the western Alps. If fresh polar or arctic air enters the Mediterranean basin in the rear of a cyclone that moves off to the east, the bora and the jet effect build up the mistral to a destructive intensity. It should be noted that it is not the shallow cold surface air of the plateau of central France, but the deep break-through of polar air into the western Mediterranean basin that favors the development of the mistral. Similar fall winds and jet-effect winds are caused on the Pacific coast of North America by the break-through of polar continental air; they are known as *northers* [41].

The jet effect, that is, a purely local increase in wind intensity because of certain orographic configurations, is observed in many localities. The convergence of streamlines in a constricted path necessitates a substantial increase in wind speed. The pressure gradients responsible for the jet-effect wind extend only over very short horizontal distances and can be detected only by special investigations. A very detailed study of such conditions in the Vienna basin was made by Margules [56]. In that area, the air masses that stream through the gate between the Kahlenberg and the Bisamberg during a west wind display all the phenomena of the jet effect, and the corresponding pressure disturbances are clearly revealed by barograms.

SURVEY OF DESIRABLE STUDIES OF LOCAL WINDS

Land and Sea Breezes. Further intensive investigation of the vertical structure of land and sea breezes on especially suitable coasts by means of continuous aerological measurements over land and water at various distances from the shore would be most desirable. As regards the theoretical aspects, the problem appears adequately solved.

Mountain and Valley Winds. Accurate aerological cross sections along favorably situated slopes up to the ridge are still needed. Likewise, systematic upper-air soundings should be carried out in a particularly favorable valley location from its deepest recesses out into the plains. Special attention should be given here to the interrelation between the slope wind and the valley

wind. Such investigations should cover the entire cross section of the valley up to the ridges of the flanking mountain chains.

The cell circulation for an inclined slope and the connection of the slope-wind circulation with the system of the mountain and valley winds still constitute a theoretical problem.

Foehn Winds. As an extension of the work of W. Schmidt [68], a comprehensive monograph on the occurrence of the foehn in all areas of the world, with special emphasis on its climatic importance, is needed.

As far as the foehn theory is concerned, its thermodynamic aspects appear to be completely explained. The dynamic side of the problem, which is a purely hydrodynamic problem, can be advanced only by special aerological investigations in the different mountain ranges of the world. Special attention should be paid to the problem of wave formation in the lee of the range.

REFERENCES

I. General references on the subject of local winds.

Conrad, V., "Die klimatologischen Elemente und ihre Abhängigkeit von terrestrischen Einflüssen," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. 1, Teil B. Berlin, Gebr. Borntträger, 1936. (See "Land- und Seewind," pp. B226–252; "Die Beeinflussung der Luftströmungen durch die Gebirge," pp. B306-332.)

Ficker, H. v., und de Rudder, B., *Föhn und Föhnwirkungen*. Leipzig, Becker & Erler, 1943.

Hann, J. v., und Süring, R., Hsgbr., *Lehrbuch der Meteorologie*, 5. Aufl. Leipzig, Keller, 1943. (See pp. 545, 555-556)

Koschmieder, H., *Kleinräumige Luftbewegungen*, Bd. 3, *Physik der Atmosphäre*. Leipzig, Becker & Erler, 1944. (See [54])

II. References cited in text.

1. Alisoff, B. P., "Le phénomène de bora dans le Kazakstan de l'est." *J. Geophys. Leningrad*, 5:115 (1935).
2. Arakawa, H., and Utsugi, M., "Theoretical Investigation on Land and Sea Breezes." *Geophys. Mag., Tokyo*, 11:97–104 (1937).
3. Arndt, A., Über die Bora von Noworossisk." *Meteor. Z.*, 30:295–302 (1913).
4. Barschall, H., "Über die Gebirgswinde in den Mittleren Vogesen." *Meteor. Z.*, 36:137–142 (1919).
5. Berjoan, L., "Sur un cas remarquable de mistral (4 au 8 mars 1926) à Marseille." *La Météor.*, 3:215–217 (1927).
6. Billwiller, R., "Der Talwind des Ober-Engadin." *Z. öst. Ges. Meteor.*, 15:297–302 (1880).
7. —— "Der Talwind des Oberengadin." *Meteor. Z.*, 13:129–138 (1896).
8. Bjerknes, V., and others, *Physikalische Hydrodynamik*. Berlin, J. Springer, 1933.
9. Braak, C., "Beobachtungen über den Seewind." *Ann. Hydrogr., Berl.*, 56:190–192 (1928).
10. —— "Der Malojawind." *Meteor. Z.*, 50:231–232 (1933).
11. —— "Het klimaat van Nederlandsch-Indië." *Verh. magn. meteor. Obs. Batavia*, Nr. 8, Deel I, 272 pp. (1923). (See p. 91)
12. Burger, A., und Ekhart, E., "Über die tägliche Zirkulation der Atmosphäre im Bereiche der Alpen." *Beitr. Geophys.*, 49:341–367 (1937).
13. Conrad, V., "Beobachtungen über den Seewind an einem flachen Sandstrand." *Ann. Hydrogr., Berl.*, 56:1–3 (1928).
14. Defant, A., "Periodische Temperaturschwankungen bei Föhn und ihr Zusammenhang mit stehenden Luftwellen." Innsbrucker Föhnstudien II. *Denkschr. Akad. Wiss. Wien.*, 80:107–130 (1907).
15. —— "Zur Theorie der Berg- und Talwinde." *Meteor. Z.*, 27:161–168 (1910).
16. —— "Der Abfluss schwerer Luftmassen auf geneigtem Boden, nebst einigen Bemerkungen zu der Theorie stationärer Luftströme." *S. B. preuss. Akad. Wiss.*, 18:624–635 (1933).
17. Defant, F., "Zur Theorie der Hangwinde, nebst Bemerkungen zur Theorie der Berg- und Talwinde." *Arch. Meteor. Geophys. Biokl.*, 1(A):421–450 (1949).
18. —— "Theorie der Land- und Seewinde." *Arch. Meteor. Geophys. Biokl.*, 2(A):404–425 (1950).
19. Ekhart, E., "Zur Aerologie des Berg- und Talwindes." *Beitr. Phys. frei. Atmos.*, 18:1–26 (1931).
20. —— "Einiges zur Statistik des Innsbrucker Föhns." *Meteor. Z.*, 49:452–459 (1932).
21. —— "Neuere Untersuchungen zur Aerologie der Talwinde, die periodischen Tageswinde in einem Quertale der Alpen." *Beitr. Phys. frei. Atmos.*, 21:245–268 (1934).
22. —— "Die Windverhältnisse des Arlbergs." I und II. *Beitr. Geophys.*, 48:313–324 (1936); 49:7–25 (1937).
23. —— "Zum Klima der freien Atmosphäre über USA." I, II, und III. *Beitr. Phys. frei. Atmos.*, 26:50–66, 77–106, 210–242 (1940).
24. Ficker, H. v., "Beiträge zur Dynamik des Föhns." *Denkschr. Akad. Wiss. Wien*, 78:83–163 (1906); 85:113–173 (1910).
25. —— "Der Transport kalter Luftmassen über die Zentralalpen." *Denkschr. Akad. Wiss. Wien*, 80:131–200 (1907).
26. —— "Föhnuntersuchungen im Ballon." *S. B. Akad. Wiss. Wien*, Abt. IIa, 121:829–873 (1912).
27. —— "Die Flächen gleichen Druckes bei Berg- und Talwind." *Ber. preuss. meteor. Inst.*, Nr. 320, SS. 56–67 (1924).
28. —— und de Rudder, B., *Föhn und Föhnwirkungen*. Leipzig, Becker & Erler, 1943.
29. Galzi, L., "Le mistral à Nîmes." *La Météor.*, 3:213–214 (1927).
30. Godske, C. L., "Über Bildung und Vernichtung der Zirkulationsbewegungen einer Flüssigkeit." *Astrophys. norveg.*, 1:11–86 (1934).
31. Grenander, S., *Über das Erscheinen der Seebrise an der schwedischen Ostseeküste*." Diss., Uppsala, 1912.
32. Hann, J. v., "Zur Frage über den Ursprung des Föhn." *Z. öst. Ges. Meteor.*, 1:257–263 (1866).
33. —— "Zur Meteorologie der Alpengipfel." *S. B. Akad. Wiss. Wien*, Abt. IIa, 78:829–866 (1878).
34. —— "Zur Theorie der Berg- und Thalwinde." *Z. öst. Ges. Meteor.*, 14:441–448 (1879).
35. —— "Zur Theorie der aufsteigenden Talwinde." *Meteor. Z.*, 27:492–499 (1910).
36. —— "Über die Theorie der Berg- und Talwinde." *Meteor. Z.*, 36:287–289 (1919).
37. Haurwitz, B., "Comments on the Sea-Breeze Circulation." *J. Meteor.*, 4:1–8 (1947).
38. Hess, S. L., and Wagner, H., "Atmospheric Waves in the Northwestern United States." *J. Meteor.*, 5:1–19 (1948).

39. Heuer, W., "Über die Ursachen des Malojawindes." *Meteor. Z.*, 27:481–488 (1910).
40. Hoinkes, H., "Über Nordföhnerscheinungen auf der Nordseite der Alpen." *Arch. Meteor. Geophys. Biokl.* (in press).
41. Hurd, W. E., "Northers of the Gulf of Tehuantepec." *Mon. Wea. Rev. Wash.*, 57:192–194 (1929).
42. Jeffreys, H., "On the Dynamics of Wind." *Quart. J. R. meteor. Soc.*, 48:29–46 (1922).
43. Jelinek, A., "Beiträge zur Mechanik der periodischen Hangwinde." *Beitr. Phys. frei. Atmos.*, 24:60–84 (1937).
44. —— "Über den thermischen Aufbau der periodischen Hangwinde." *Beitr. Phys. frei. Atmos.*, 24:85–97 (1937).
45. —— und Riedel, A., "Über die Schichtdicke der periodischen Lokalwinde im Inntal." *Beitr. Phys. frei. Atmos.*, 24:205–215 (1937).
46. Kaiser, M., "Land- und Seewinde an der deutschen Ostseeküste." *Ann. Hydrogr., Berl.*, 35:113–122, 149–163 (1907).
47. Kanitscheider, R., "Beiträge zur Mechanik des Föhns, Ergebnisse von Doppelvisierungen in Innsbruck und Umgebung." I, II, und III. *Beitr. Phys. frei. Atmos.*, 18:27–47 (1932); 24:26–44 (1938); 25:49–58 (1938).
48. Kesslitz, W. v., "Zum Borasturm in der Nordadria vom 31. März 1910." *Meteor. Z.*, 27:233–235 (1910).
49. —— "Über die Windverhältnisse an der Adria." *Meteor. Z.*, 31:248–251 (1914).
50. Kleinschmidt, E., "Zur Theorie der Talwinde." *Meteor. Z.*, 38:43–46 (1921).
51. Knoch, K., "Der Mistral Südfrankreichs." *Petermanns Mitt.*, 56(1):297–298 (1910).
52. Kobayasi, T., und Sasaki, T., "Über Land- und Seewinde." *Beitr. Phys. frei. Atmos.*, 19:17–21 (1932).
53. Köppen, W., und Geiger, R., Hsbgr., *Handbuch der Klimatologie*. Berlin, Gebr. Bornträger, 1936.
54. Koschmieder, H., "Danziger Seewindstudien." *ForschArb. meteor. Inst. Danzig*, 8:1–44 (1936); 10:1–39 (1941).
55. Lyra, G., "Theorie der stationären Leewellenströmung in freier Atmosphäre." *Z. angew. Math. Mech.*, Bd. 23, Nr. 1 (1943).
56. Margules, M., "Material zum Studium der Druckverteilung und des Windes in Niederösterreich." *Jb. ZentAnst. Meteor. Wien*, 35:D1–16 (1898); 37(V):1–15 (1900).
57. Mazelle, E., "Kälteeinbruch und Bora in Triest, Januar, 1907." *Meteor. Z.*, 24:171–172 (1907).
58. Mörikofer, W., "Beobachtungen zur Theorie des Malojawindes." *Jber. naturf. Ges. Graubünden*, 63:69 (1924).
59. —— "Weitere Beobachtungen zur Theorie des Malojawindes." *Verh. schweiz. naturf. Ges. Luzern*, II. Teil, 112/113 (1924).
60. Pernter, J. M., "Über die Häufigkeit, die Dauer und die meteorologischen Eigenschaften des Föhns in Innsbruck." *S. B. Akad. Wiss. Wien*, Abt. IIa, 104: 427–461 (1895).
61. Pierson, W. J., Jr., "The Effects of Eddy Viscosity, Coriolis Deflection, and Temperature Fluctuation on the Sea Breeze as a Function of Time and Height." *Meteor. Pap. N. Y. Univ.*, Vol. 1, No. 2, 29 pp. (1950).
62. Prandtl, L., *Führer durch die Strömungslehre*. Braunschweig, F. Vieweg & Sohn, 1942. (See pp. 373–375)
63. Ramdas, L. A., "The Sea-Breeze at Karachi." *India Meteor. Dept. Sci. Note* No. 41, Vol. 4, pp. 115–124 (1931).
64. Rayleigh, Lord, "On Convection Currents in a Horizontal Layer of Fluid When the Higher Temperature Is on the Under Side." *Phil. Mag.*, 32:529–546 (1916).
65. Riedel, A., *Zur Aerologie der periodischen Gebirgswinde an einem Südhange, Ergeb. v. Doppelvisierungen im Innsbrucker Nordkettengebiet*. Diss., Innsbruck, 1936.
66. Saigey, J. F., *Petite physique du globe*. Paris, L. Hachette, 1842.
67. Schmidt, F. H., "An Elementary Theory of the Land- and Sea-Breeze Circulation." *J. Meteor.*, 4:9–15 (1947). (See p. 14)
68. Schmidt, W., "Föhnerscheinungen und Föhngebiete." *Wiss. Veröff. dtsch.-öst. Alpenvereins*, Innsbruck, Nr. 8 (1930).
69. Tollner, H., "Gletscherwinde in den Ostalpen." *Meteor. Z.*, 48:414–421 (1931).
70. van Bemmelen, W., "Land- und Seebrise in Batavia." *Beitr. Phys. frei. Atmos.*, 10:169–177 (1922).
71. Wagner, A., "Klimatologie der freien Atmosphäre," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. I, Teil F, 1931.
72. —— "Hangwind—Ausgleichsströmung—Berg- und Talwind." *Meteor. Z.*, 49:209–217 (1932); "Neuere Theorie des Berg- und Talwindes." *Ibid.*, 49:329–341 (1932).
73. —— "Theorie und Beobachtungen der periodischen Gebirgswinde." *Beitr. Geophys.*, 52:408–449 (1938).
74. Wenger, R., "Zur Theorie der Berg- und Talwinde." *Meteor. Z.*, 40:193–204 (1923).
75. Wiese, W., *Bora auf Nowaja Semlja*. Leningrad, Hydrometeor. Publ. House, 1925. (Reviewed by M. Kofler, *Meteor. Z.*, 44:158 (1927).)

TORNADOES AND RELATED PHENOMENA

By EDWARD M. BROOKS

Saint Louis University

Introduction

Definition. The word "tornado" comes from the Spanish *tronada* (thunderstorm). In western and north central Africa, "tornado" refers to a thunderstorm. But elsewhere it means an intense spiral motion around a vertical or inclined axis, averaging about 250 yards across, with its lower part [17] often characterized by a narrow pendant cloud [12] extending from a cumulonimbus cloud base to or nearly to the ground.

Frequency. A tabulation of tornadoes which have occurred over a long period of years is needed to determine frequencies. Unfortunately, many severe windstorms are difficult to classify correctly. The number of tornadoes is erroneously decreased by unreported tornadoes and by tornadoes classified as other storms, and it is erroneously increased by nontornadic storms reported as tornadoes [20].

Tornadoes occur on all continents, but are rare except in Australia and the United States. The numbers of tornadoes reported in these two countries [1] average about 140 and 145 per year, respectively, and the frequencies per unit area are similar. Australian tornadoes, however, do not become as severe. In the United States the highest annual frequencies per ten thousand square miles are found in northeastern Kansas (3.2), central Arkansas (3.0), and throughout Iowa (2.5 to 2.3). However, some of these differences between states may be due to their different methods of classifying windstorms. Because the area swept by a tornado averages only about two-thirds of a square mile, the probability of a tornado's hitting a given square mile of the Midwest is generally less than 1 per cent per century.

The number of tornadoes in any one year may vary considerably from the annual average. The seasons, listed in order of tornado frequency for the United States, are spring, summer, fall, and winter. May has the greatest number; December, the least. For an individual state, the month of maximum tornado frequency is usually one month before the maximum hailstorm frequency and two months before the maximum thunderstorm frequency. Tornadoes appear most frequently in the following months: in the southern states in March; in the central states in April or May; and in the northern states in June or July [12]. Multiple tornadoes within a general cyclone are most likely in April or May [12].

Over 80 per cent of reported tornadoes occur between noon and 9 P.M. This preponderance of afternoon storms seems to be least pronounced in winter.

Weather Conditions Associated with Tornadoes

Synoptic Situations. The usual United States weather map, favorable for tornadoes, shows a deep extratropical cyclone with central sea-level pressure below 1002 mb located in the central or northern part of the country, usually with a curved trough extending generally southeastward from it and having a large wind shift across it. The most likely place for tornadoes is within the portion of the trough generally 100–600 mi southeast of the center of the parent low. Tornadoes rarely occur within 50 mi of the primary, or parent, low center, but when they do they are usually very violent.

The following extratropical frontal conditions are associated with many tornadoes. Along the trough there is usually a cold front aloft, or a surface cold front where a maritime tropical air mass is being displaced by a modified continental polar or maritime polar air mass in 90 per cent of the cold front cases and by a fresh continental polar air mass in 10 per cent. Sometimes the trough contains an occluded front or a warm front instead.

Other tornadoes occurring with a deep parent cyclone are not located on a front. Tornadoes occasionally occur in the maritime tropical air warm sector southeast of the parent low center. These tornadoes might be located on a squall line, or at the points of intersection of two or more squall lines [21], or on several closely associated squall lines separated by relatively quiet areas. Sometimes tornadoes occur on the north or west side of the parent low center in the polar air cold sector which may have maritime tropical air over it. Occasionally the parent cyclone is a tropical hurricane, in which tornadoes occur in maritime tropical air usually north rather than southeast of the center. About four-fifths of the tornadoes in the United States are "cyclonic" and the rest "convective." By definition, "cyclonic" tornadoes occur with a well-developed parent low, whereas "convective" tornadoes are either frontal or nonfrontal and occur in a weak parent low, in a weak pressure trough, or even in a high-pressure area. Tornadoes are known to occur along a tropical or intertropical front [9].

Typical aerological soundings [19] of the maritime air in which tornadoes develop show that it has high relative humidity usually only up to about 1–3 km. A thin stable layer, which may be an inversion, separates the maritime air from the dry superior air aloft, which is characterized by a steep lapse rate. The thin layer is convectively unstable because of the rapid decrease of humidity with height.

Nearby Weather Conditions and Preliminary Signs. Tornadoes are not to be expected unless there are some preliminary signs of cumulonimbus activity, since the pendant cloud, sometimes not visible, extends downward from a cumulonimbus cloud base. The under side of the cumulonimbus frequently exhibits a mammatus appearance before the tornado [9]. Along the path of a

single tornado, repeated reports of two cumulonimbus clouds or thunderstorms coming together just before the tornado appears indicate that the junctions are a result rather than a cause of the tornado circulation [11]. In the absence of much light coming through the thick clouds from above, reflections from the ground presumably cause the green- and yellow-tinted clouds near tornadoes. The cumulonimbus clouds are often very dark because of their great vertical extent.

Often a tornado is preceded by a violent thunderstorm (northeast of the tornado) [11]. More often a tornado is followed by a thunderstorm (southwest of the tornado), often a severe one. Or a thunderstorm center may follow a path parallel to the path of the tornado (usually northwest of the tornado). Green-colored lightning and even ball lightning have been reported. On the other hand, sometimes a tornado occurs even though no thunderstorm is reported.

Rain almost always accompanies a tornado [17], but heavy rain is not likely within the tornado itself. Heavy showers are common, especially following a tornado, and frequently yield the greatest precipitation just northwest of the tornado path. Hail is also common [17], sometimes occurring as much as two hours before and/or after a tornado, often northwest and sometimes southeast of the tornado track. Its duration is shorter than that of rain, but it may cause more damage than a minor tornado. Very large hailstones [6], ranging up to discs 7 in. in diameter (weighing 3 lb), and hail accumulating to a depth of several inches have been reported near or even along tornado paths.

The air preceding a tornado is usually warm and humid, whereas the air following it is usually cool and dry. Research at Parks College of Aeronautical Technology showed that 111 out of 142 tornadoes occurred on the line of highest surface dew point. The drop in temperature with the arrival of the first thundersquall usually exceeds the difference between the temperatures of the preceding and following air masses.

Within a few miles outside the tornado path it is not unusual for pressure drops of over 3 mb to occur, and in extreme cases the change amounts to about 10 mb. This indicates that the tornado is surrounded by a low-pressure area which, with its attendant winds, has a radius of about 5–10 mi [2], which is intermediate in size between the parent low and the tornado. In this respect, the tornado is similar to an ordinary thunderstorm, which has an inflow current coming from as far as about 5 mi away from the edge of the active cloud [22].

A tornado is usually preceded by a general southerly wind and followed by a general westerly wind. Research at Parks College showed that 92 out of 132 tornadoes occurred on the line of highest surface wind speed. Winds in the vicinity of the tornado are likely to be strong and variable, sometimes reaching full hurricane force outside the tornado proper. The reasons for these strong winds are that (1) the parent cyclone causing the prevailing winds is more intense than is usual for a cyclone, (2) the secondary low surrounding the tornado has steep pressure gradients, (3) there may be heavy thunderstorm squalls, and (4) small local whirlwinds are often present. Any one of these effects alone or in combination with the others may account for a zone a few miles wide with scattered damage to buildings and trees.

Properties of Tornadoes

Central Pressure and Wind Speed. The pressures and winds within a tornado must be determined largely by indirect methods since there are very few instrumental measurements. These methods include a detailed analysis of the tornado damage and a theoretical study of the relationship between wind and atmospheric pressures. Two fundamental causes of damage are (1) the terrific winds, exerting pressures and suctions, and carrying missiles, and (2) the sudden changes of the prevailing atmospheric pressure. Since both are related to pressure, the combined pressure effects observed on structures cannot be attributed either to atmospheric pressures or to wind pressures alone without the introduction of errors. Cracks in the ground [12] and completely smashed objects cannot be used in estimating wind pressures when the true causes of the damage are unknown.

Although Hazen [11] thought that tornadoes were local high-pressure areas, barographs in tornadoes show pressure drops usually up to 25 mb. In one case an aneroid barometer fell 200 mb (Minnesota, Aug. 20, 1904). The actual atmospheric pressure drops are probably greater because the pressure changes inside buildings are sluggish and because the buildings may not be in the exact center of the tornado path. Espy's experiments [4] showed that low pressure was not the cause of the forced removal of feathers from chickens. Neither this denudation nor the breaking and peeling of bark from tree trunks can be used as an argument for such pressure. Explosions of windows, doors, walls, and roofs not under the direct influence of tornadic winds (when the vortex is just off the earth's surface) are clear indications of the sudden arrival of a very low atmospheric pressure and a very large horizontal pressure gradient. For a frictionless vortex with inflow, the gradient would be inversely proportional to the cube of the distance from the center (except very near the center).

Determinations of maximum wind speeds in tornadoes are rough approximations. Calculations of wind based on structural failures and impacts of flying objects yield values ranging from less than 100 mph to one case of more than 300 mph, enough to demolish the strongest buildings.

Ferrel [6] gives the following theoretical relationships between wind and atmospheric pressure in a frictionless vortex surrounded by a windless environment: The total kinetic energy of the wind is equal to the work that would be done by the pressure gradient in bringing the air into the vortex from the environment. The wind speed is the same as the free-fall velocity of an object dropped in a vacuum from the height having the same pressure in the environment as that at the ground in the vortex. For example, a pressure of 900 mb, or a decrease of 100 mb, would give a wind of about 300

mph. This depression is also equal to the theoretical wind pressure of the corresponding wind. Hence, the wind pressure and atmospheric pressure effects on closed structures should be of the same order of magnitude.

Flow Pattern. Although the rotation of air in a tornado is generally accepted as a fact today, there was no such agreement one hundred years ago [4, 17, 18]. With a few doubtful exceptions [12] the rotation in tornadoes is believed to be always cyclonic. The evidence for this is found in a wide belt of damage usually on the southeast side of a tornado path in the Northern Hemisphere. In this belt the tornado wind must be increased because of the prevailing wind (in approximately the same direction) which in some cases may be responsible for even more damage than the rotation.

Theoretically, the tangential velocity should be inversely proportional to the distance from the center for frictionless inflow, according to the principle of the conservation of angular momentum. Actually, the presence of a frictional torque acting against the wind makes the velocity profile somewhat less steep. Consequently, the vortex is not irrotational as it would be for zero friction, but has a cyclonic vorticity. The direction in which a house is twisted on its foundation is not a reliable indication of the sense of the tornado's rotation nor of its vorticity, as the twisting is dependent on variations (along the windward wall) in the structural resistance and in wind speed, which may be influenced by trees or by a strictly local whirl.

The radial motion in a tornado is inward near the ground. The winds in different tornadoes or even within one tornado probably blow inward at different angles to the radius, depending on the ratio of inflow and tangential speeds. Along many tornado tracks, evidence for the inflow is much clearer than evidence for the rotation, which suggests that, in these cases, inflow speeds may be greater than tangential speeds except near the center. The inflow winds are often strong enough to cause damage by suction over lee roofs or around lee walls, that is, on the side facing the tornado center [11]. This suction, added to the atmospheric pressure reduction, builds up an intolerable pressure deficit outside. On the other hand, the windward roofs and walls are damaged by excess pressure due to the wind (counteracted somewhat by the atmospheric pressure deficit) and by battering from missiles.

Continuity demands a strong vertical motion upward within the vortex. A remarkably strong upflow was once witnessed (Abilene, Texas, June 1938) when a tornado got ahead of its cumulonimbus cloud and immediately created a new cumulonimbus cloud which shot up to an elevation of 35,000 ft in one minute. The rapid vertical acceleration of the air requires a vertical pressure-gradient force upward much stronger than gravity. The central depression of the tornado must be much more pronounced just above the ground than at the ground. Buildings in the path of a tornado receive greatest damage in their upper floors.

Redfield visualized a double spiral, the air from outside the vortex spiralling first downward and inward toward the center, then within the vortex upward and outward. The latter motion was once observed [17]. Whirling air coming down at 1800 feet per minute has actually been encountered by a glider pilot in a desert dust whirl; and hot puffs of air about 20F warmer (because of adiabatic compression) than the prevailing air have been felt at the ground near some tornadoes.

The turbulence in tornadoes is extremely large because of strong surface and internal frictional effects. Mechanical eddies may contain destructive gusts of wind. In the central region of rapid convergence and upflow, speeds of secondary whirls increase as their radii decrease. The presence of such miniature tornadoes may be indicated by short pendant clouds overhead close to the main pendant cloud (for a photograph of this, see [7]).

Appearance and Visibility of a Tornado. A tornado may appear as two columns: (1) a pendant cloud extending downward from the general cumulonimbus cloud base, and (2) an ill-defined mass of dust and debris extending a short distance upward from the ground. The air within the pendant cloud has expanded adiabatically to pressures and temperatures lower than the condensation pressure and temperature of the air mass, which is assumed to have nearly uniform specific humidity before condensation. The freezing level [6] as well as the condensation level is lowered by adiabatic cooling. The core of a tornado felt ice-cold to one observer. At a central pressure of 900 mb the temperature could be as much as 10C or 18F cooler (rather than warmer [4]) than the surroundings.

The shape of the pendant cloud is determined by the slope of the condensation-pressure isobaric surface. The slope is usually steepest near the axis of rotation, but sometimes excessive friction greatly reduces the innermost slope, prevents the cloud from reaching the ground, and changes it from a funnel shape to a basket or balloon shape. If the air is dry enough or the central depression is weak enough for the central pressure to be just slightly below the condensation pressure, the pendant cloud is narrow, resembling an elephant trunk, a rope, or a serpent. After the waterdrops condense in the pendant cloud they tend to whirl outward from the center owing to their centrifugal force. This leaves a very small "eye" at the center [13], characterized by cloudless air and relatively light wind, as has actually been observed.

The dust column is usually wider than the pendant cloud and surrounds its base, because the wind normally reaches a speed great enough to raise dust before the pressure drops as low as the condensation pressure. Exceptions occur where the soil is too wet for dust or where the air is so moist that the pendant cloud is very broad, extending outward beyond the edge of the high winds. The black dust column may be mistaken for smoke rising from a fire. As it rises, most of the debris is thrown outward in all directions by centrifugal force, sometimes giving the appearance of a dark fountain. The condition for the occurrence of damage is not whether the pendant cloud strikes, but whether the usually dusty column of high winds strikes.

As many as nine pendant clouds have been seen from one place during a tornado situation [1, 12]. In many

cases, vortices have occurred about a mile apart [12]; and in extreme cases, only 150 or 100 ft apart. On the other hand, in some of the most destructive tornadoes, no funnel clouds were recognized. Either the pendants were much wider than their vertical extent or they were obscured by darkness at night or even during the day because of heavy precipitation or dust, a very low general cloud base, hills, or buildings. A higher percentage of funnel clouds are seen in the Great Plains than in the eastern part of the United States [12].

The axis of a tornado initially may be nearly vertical, but since the top and bottom usually progress at different velocities, the axis becomes tilted away from the vertical in the direction of the vertical shear until it may be nearly horizontal. The base usually drags behind, because the translatory wind is reduced by friction at the surface. Finally, it may become separated entirely from the cumulonimbus cloud originally over it.

Audibility and Scent. A tornado reaching the ground produces a roaring or buzzing sound which has been heard as long as one hour before it arrived [11] and as far as 25 mi away; this distance is comparable to that at which thunder can be heard, but is much greater than the ½ mi at which the sound of heavy hail can be heard. As this noise still occurs when a whirl is aloft (though to a lesser extent), it is not due entirely to the destruction being caused by the wind, but is due also to vibrations created by frictional effects in the strong wind shear of the whirl. Such sounds are augmented by long rolls of thunder, which may overlap to make a nearly continuous background of rumble.

During one tornado, the air seemed to have a suffocating and burning tendency. Odors during a tornado are rarely noticed. They have been reported as resembling the smell of ozone or burning sulphur. These odors are more likely to be due to the effects of lightning discharges [17] than to the effects of the tornado.

Paths. The diameter of a tornado, or width of a tornado path, which averages close to 250 yards, varies from zero up to about 1 or 2 mi. The length of a tornado path averages about 4½ mi, and ranges from only about 100 ft up to 300 mi. "Cyclonic" tornadoes usually have much longer tracks than "convective" tornadoes, because they move faster and last longer. The most common path length is only 1½ mi, probably because the vortex frequently rises and skips over a stretch of land before starting a new path. Repeated lifting of the funnel may even occur while the vortex is stationary (Miami, April 5, 1925).

About 90 per cent of all Northern Hemisphere tornadoes move in a direction from between south and west because they are embedded in and consequently move with the warm moist air [16], which usually blows from this quadrant. In hurricanes, tornadoes usually move from the east because of their location north of the center. Tornadoes sometimes move from the northwest, especially in Texas, where they may be steered by an upper wind blowing around the high aloft, characteristic of the southwestern United States in summer. A translation from unusual directions [12] may be found where the parent cyclone, if present, is not northwest of the tornado. Sometimes tornadoes within a single system move in widely different directions [12, 16].

The translation speed of a given tornado is not uniform [17]. Also the average translation speeds of different tornadoes vary, ranging from nearly zero to nearly 150 mph, with a mean close to 35 mph. The average tornado durations are computed to be ¼ min at a single point and 8 min on a path along the ground. An extreme duration of over 7 hr along the ground has been observed (Illinois, May 26, 1917).

The tornado path is often nearly straight when it extends for a considerable distance over flat country. For example, it may not deviate more than a mile or two in over 150 mi. Yet such a deviation would make the most common path (1½ mi long) appear to be strongly curved. There is no preferred direction of curvature as the vortex zigzags along. However, experimental research indicates that a vortex rotating in a cyclonic sense is forced to the left of a horizontal current (looking downwind). Possibly the path curves to the left if the tornado is moving faster than the accompanying secondary low, or to the right if moving slower.

Hills and even buildings make tornado paths more crooked. Like a dust devil [14], a tornado on the slope of a hill or ridge tends to be deflected upslope. When ridges and valleys are oriented at right angles to the direction of translation, the tornado tends to skip, but skipping can also occur over water or flat land [17]. It appears that leeward slopes of ridges are more often severely hit than windward slopes, and that valley bottoms [18] are no safer than hilltops. However, some meteorologists do not agree with these effects of topography on tornadoes [8]. Rough terrain destroys many tornadoes, but it sometimes creates whirlwinds by orographic convergence.

Many cases are reported of two or more whirls combining into a single whirl [11], or of one whirl breaking up into several whirls. Under these conditions there is usually only one major whirl.

Theories on the Generation and Maintenance of Tornadoes

Energy for Vertical Motion. A major problem in explaining the formation of a tornado is to find the source of the potential energy and the manner in which it is converted into kinetic energy.

In the nineteenth century, Hare [10] theorized that the chief source was electrical because of lightning in and around the funnel, the fiery appearance [4] of the funnel, the smoky appearance [11] of the dust column, and the scorched appearance [4] of vegetation (owing to the effects of the sun on broken plants) along the path. Hare [10] supposed that the surface air was electrically charged and, being attracted upward by an oppositely charged thundercloud, started a rapid upward motion. Such a flow would tend to wipe out the vertical gradient of electric potential, as lightning does in a thunderstorm. Electrical phenomena are usually considered a result [17] of severe convective activity rather than a

cause of it. Jones [15] finds that lightning in tornadic thunderstorms gives more intense radio static (sferics) than lightning in ordinary thunderstorms.

Espy [4] theorized that heat and latent heat of vaporization supplied the energy for the strong convection in a tornado. To supply thermal energy fast enough, a violent updraft of warm moist air would be required. Since the process cannot get started by itself, the heat of vaporization is a maintaining factor rather than the initial cause, which must be an external source of energy. In other words, the pendant cloud formation aids the tornado, but does not cause it. A California oil fire (1923) heated the surface air sufficiently to generate hundreds of whirlwinds, one of which was violent enough to kill two people and wreck a house.

Hail has been regarded by some meteorologists as being the result of vortex motion [17], and a tornado as being an unusually severe hailstorm cloud. On the other hand, Showalter [19] recently suggested that hail might be the initial mechanism that causes a tornado. Hail falling from the overhanging top of a cumulonimbus cloud would cool the layer of dry air below it by conduction and evaporation until the original dry inversion (or stable zone) would reach unstable equilibrium. As this instability would not extend above the cooled layer, the available convective energy would be concentrated in the lower atmosphere.

Horizontal Motion as a Source of Energy. Showalter [19] points out the necessity of lifting or horizontal convergence for the establishment of free convection. In either case, horizontal inflow is required. Concerning frontal lifting, Humphreys [13] states that vortex motion cannot usually be generated within the warm air mass being lifted, regardless of how much windshift there is at the front. For vortex motion to develop in the warm air it is necessary for the warm air itself to have angular momentum about the center of convection. This requirement may be fulfilled in the trough of a squall line. Horizontal convergence at a front is effective in producing vortex motion if both air masses rise, for in this case the frontal windshift supplies the necessary angular momentum. Humphreys [13] says this condition is best fulfilled at an upper cold front, which Willett [24] qualifies as having a small horizontal temperature contrast.

Taylor [20] regards a tornado as the extreme development of a small intense secondary low formed on a front. There is much evidence in favor of the existence of a secondary low ("tornado cyclone") around a tornado [2]. Before the tornado forms, there is ample angular momentum about the center of the tornado cyclone. To generate and maintain a tornado, a certain minimum rate of inflow is required to create an acceleration of the wind sufficient to outweigh by far the frictional losses.

Garbell [9] mentioned that a tornado may follow a combination of (1) strong vertical motion indicated by a rapidly growing cumulonimbus cloud, and (2) cyclonic circulation. Incipient vortices are present in most thunderstorms; it is only when the gyratory motion becomes very intense that they reach from the cloud base to the ground.

Wegener [23] introduced the theory that a tornado is the same as the rotating thunderstorm squall cloud, one end of which reaches down to the earth. Aside from insufficient observations to substantiate this theory, one objection is that the numbers of tornadoes rotating clockwise and counterclockwise are far from being equal, as would be required if there were no preference as to which end of the squall cloud would dip to the earth.

Cause of the Low Pressure in a Tornado. The biggest difficulty in explaining the maintenance of a tornado is to account for the failure of the low pressure at its center to disappear in the face of strong surface inflow.

A common explanation for the low pressure is that the pressure reduction is caused by the outward centrifugal force of the whirl. If the low pressure is caused by the whirl, the latter must exist first, probably developing between two strong countercurrents. For the low to develop there must be a slight outflow which, by the approximate conservation of angular momentum, would lead to a reduction in wind speeds to values less than those of the original countercurrents. The fatal objection to this explanation is that wind speeds in a tornado far exceed the speeds of the winds prevailing on either side of it.

If inflow rather than a slight outflow is assumed, the centrifugal force of the inflowing air is greatly increased (approximately inversely as the cube of the radial distance). Some have argued that such a terrific increase in centripetal acceleration requires a great strengthening of the pressure gradient and, of course, a very low pressure at the center. Actually the mass inflow would have exactly the opposite effect on the central pressure, namely, the filling of the low. After a slight inflow, the increase in winds would make the centrifugal force larger than the pressure gradient force, making the inflow zero or even reversing it. The fallacy is in the original assumption of unlimited inflow.

The foregoing discussion shows that a tornado cannot be generated by either inflow or outflow at the surface.

Hare [10] proposed that the volume of the uprushing air exceeds that of the inflowing air, thereby leaving relatively low pressure. Redfield disagreed, saying there is no such thing as "suction." A surface low-pressure area can be produced by excessive removal of air upward only if the pressure gradient force acting upward exceeds gravity before as well as during the tornado. This requires the formation of a low aloft, the generation of which still needs to be explained.

Divergence aloft is necessary [3] not only to generate and maintain the low aloft, but also to prevent inflow from filling the low at the surface. Since it is known that there is inflow at the surface and vertical ascent in the core, continuity demands outflow aloft. Such outflow may have been witnessed once before a tornado formed [17]. Espy thought there was enough eviction of air aloft to build up a ring of higher pressure in the surroundings. For deepening the surface low, the outflow aloft must exceed the inflow at the surface. This is possible due to the dragging away (entrainment) of

some of the environment by an upward current [3, 22]. It can also occur if the inflow is reduced, which would be the case with the development of a strong centrifugal force, as described above. Because of the smaller horizontal pressure gradient aloft, the centrifugal force of the whirl overbalances it, producing the outflow needed to maintain the tornado. This is an argument in favor of accepting rotation in addition to some inflow as a necessary feature of tornadoes. An analogy can be found in the case of water running down a drain, high horizontal water speeds being obtained only when there is a whirl. Redfield and Ferrel [6] in the nineteenth century realized that gyratory motion as well as an unstable condition was necessary for the formation of a tornado.

Protection of Life and Property

Forecasting and Tracking. Unlike the ordinary daily weather forecast, reliable specific tornado forecasts cannot be made at the present time [8]. At best, conditions may be found which are favorable for the development of tornadoes over a wide area. Ever since Finley [7] first attempted to limit this area to a quarter of a state, others have been trying to limit the area or time of occurrence by taking into account variations in surface and upper-air conditions.[1] For more than fifty years the U. S. Weather Bureau has been issuing warnings of severe local storms within the next 24 hr without mentioning tornadoes specifically.

Although the exact place and time a tornado will strike are not known, Lloyd [16] and others realized that, once a tornado had formed, its future course could be predicted with reasonable accuracy from a knowledge of winds aloft in the warm air mass. In some midwestern cities, the U. S. Weather Bureau has considered plans for the detection and tracking of tornadoes by telephone, by short-wave radio sets with independent power supplies, and on radarscopes. Identification and tracking should be improved by the use of sferics, a method now being tested experimentally [15].

Injuries, Loss of Life, and Public Safety Measures. The average annual death toll from tornadoes in the United States is about 245 [1]. Since the population of the United States exceeds 100,000,000, the average chance in the United States of being killed by a tornado in any given year is less than 1 in 400,000. Much of the loss of life and many injuries are caused by objects striking people's heads, and by fires starting after the tornado. The greatest loss of life in a single tornado was 689 (March 18, 1925) and the greatest on a single day was about 1200 (Feb. 19, 1884).

People can protect themselves better by learning to recognize local signs of a tornado and to watch the sky when public forecasts call for severe local storms. If a tornado cloud appears, it is advisable (if time permits) to shut off immediately the electric power and gas supplies and to extinguish all fires (in fireplace, furnace, etc.) so that a conflagration will not occur and burn to death someone trapped by heavy debris. The next step is to seek *shelter* quickly in a tornado cellar, or in the southwest corner of the basement of a frame house, or beside an inside partition on a lower floor of a reinforced concrete or modern steel building, but not in a house with brick walls. In a city, it is generally dangerous to try to get in a car and drive away from an approaching tornado because excessively high winds, often with flying debris and hail, could wreck the car and even kill the occupants. If a person is caught in the open without available shelter, to avoid injury or death he should lie flat in a ditch or culvert, hold on to a fixed object to keep from being blown away, and cover himself, especially his head, to protect himself from missiles.

Property Damage, Building Safety, and Insurance. The average annual United States property loss from tornadoes is over $11,000,000 [1]. The yearly damage is extremely variable, depending on the size of population centers hit. Also, the total amount of property and the value of the dollar undergo considerable changes over a long period of years. The greatest damage on a single day, with 57 tornadoes, was about $35,000,000 (Feb. 19, 1884); however, that much damage has been caused by only two tornadoes in St. Louis (1896, 1927). Indirect property losses are from fires and looting after the storm.

The Western Society of Engineers in 1925 recommended that a building be designed for wind pressures of 65 lb ft^{-2} with a factor of safety of 4, which would probably save it in a minor tornado or on the outer edge of a major tornado. A building, especially a public meeting place, should be better protected by being bonded together, being anchored to a basement, having sufficient openings or automatic vents to relieve pressure, and having a grove of large trees (preferably oak) on its southwest side to diminish the wind speed.

Finley, on the other hand, thought buildings should be constructed as if there were no tornadoes, and then be covered by insurance. Most tornado insurance is lumped together with all other windstorm insurance to avoid arguments about whether a windstorm was or was not a tornado. The highest risks in the United States are along the South Atlantic and Gulf Coasts because hurricanes there cause about seven and one-half times as much damage as do tornadoes in the Middle West.

Factors usually taken into account in writing windstorm insurance policies are (1) the location of the building, (2) the construction of the building, and (3) the susceptibility of the contents of the building to damage.

Other Whirlwinds

Since whirlwinds in general have received less attention than tornadoes, they are not well understood. The information given below is based on only a few good sources.

Both dust devils and waterspouts are, on the average, less violent and have higher central pressures

1. For an account of some recent work of this nature see p. 792 in "A Procedure of Short-Range Weather Forecasting" by R. C. Bundgaard in this Compendium.

than an average tornado. Yet, in special cases, they have been known to change into tornadoes.

Dust Devils. The dust devil [14, 25], ranging from less than 10 ft to more than 100 ft in diameter, is about $\frac{1}{100}$ to $\frac{1}{10}$ as large as a tornado. It has no pendant cloud, but a whirling column of dust or sand, in which the size of the particles increases with the distance from the center. Dust devils occur most frequently on very hot days over dry terrain, and are caused by strong surface convection in a slightly irregular wind field. Since the direction of rotation is accidental, it may be clockwise as often as counterclockwise. The horizontal rotation and upflow within a dust devil usually exceed 20 mph. The dust devil is reported to have an average height of 600 ft and to last approximately the same number of hours as it is thousands of feet high. If the prevailing wind is less than 3 mph, a dust devil tends to seek higher terrain rather than to travel with the wind.

Waterspouts. Waterspouts [13] fall into two classes: tornadoes over water, and fair-weather waterspouts.

Most waterspouts are tornadoes, with cyclonic rotation and the characteristic pendant from a cumulonimbus cloud. The spout may appear more transparent through its center than through its edges, because of the presence of a small eye, as in a tornado. Sometimes a waterspout appears to be double, the pendant of condensed water vapor being inside a spreading column, which appears to be a fountain of spray from the surface. Waterspouts frequently appear in groups, as many as thirty having been visible from a ship in one day. Owing to the small amount of friction over a water surface, the motion in a waterspout is more nearly tangential, with less inflow and upflow than in a tornado. The reduced friction and larger moisture supply over a water surface are not the chief factors determining intensity, for, if they were, waterspouts would be stronger than tornadoes.

The fair-weather waterspout, like the dust devil, is a low-level whirl, rotating in either sense. It is caused by irregular air flow and pronounced surface instability, more the result of high humidity than of high temperature.

The water surface under a waterspout is either raised or lowered depending on whether it is affected more by the atmospheric pressure reduction [17] or by the wind force. The waterspout does not lift any significant amount of water from the surface, as shown by the precipitation of mostly fresh water on a ship passing through a waterspout on the ocean. It often dissipates on reaching a shore line because of the absence of sufficient inward acceleration to compensate for frictional losses.

Concluding Outlook

The subject of tornadoes is very old—the ancient savants Pliny the Elder, Seneca, and Lucretius wrote about them. Since then, progress in the study of tornadoes has been slow. Everdingen [5] noted that the literature on tornadoes in the United States in the last half century has, for the most part, been confined to compilations of such statistics as the distribution and frequency of occurrence and description of resulting damage. What is lacking is the application of satisfactory hydrodynamical theories to the frictional vortices of tornadoes, dust devils, and waterspouts. For example, the horizontal velocity profile could be found by an approximate solution of the differential equation expressing the rate of change of angular momentum of a fluid particle in terms of the net torque from surface and internal friction, as has been done for gaseous spray nozzles.

Suggested items for study are (1) the outflow aloft (method of removal and destination of removed air), (2) the formation by action of hail, and (3) the sense of rotation, as examples of explaining facts, checking proposed theories, and verifying accepted ideas, respectively. For this work, many more surface and upper-air observations are needed.

Since the tornado is a local circulation, the meteorological data ought to be gathered over micronetworks which could detect the development of a small secondary cyclone and would accurately locate squalls, thunderstorms, and hail showers with respect to the tornado.

After a tornado occurs, careful surveys should be made of the damage to determine winds and atmospheric pressure drops. A standardized questionnaire [11] should be used in personal interviews and should be published in local newspapers with a request for replies. Copies of local photographs should be obtained for analysis, the best photographs being those which include the top of the cumulonimbus cloud formation above the pendant cloud rather than the pendant alone.

A more complete knowledge of the small-scale circumstances attending tornadoes is needed for the understanding of the nature and causes of tornadoes. Such knowledge will put tornado forecasting and tracking on a firmer basis.

REFERENCES

A valuable bibliography, which was published after this list of references was prepared, is that by Miss H. P. Kramer, "Selective Annotated Bibliography on Tornadoes." *Meteor. Abstr. & Bibliogr.* (publ. by Amer. Meteor. Soc.) 1:307–332 (1950).

1. Baldwin, J. L., "Preliminary Report on Tornadoes in the United States during 1943 and Totals and Averages, 1916–42, by States." *Mon. Wea. Rev. Wash.*, 71:195–197 (1943).
2. Brooks, E. M., "The Tornado Cyclone." *Weatherwise*, 2:32–33 (1949).
3. Brunt, D., *Physical and Dynamical Meteorology*, 2nd ed. Cambridge, University Press, 1939. (See pp. 303–306)
4. Espy, J. P., *The Philosophy of Storms*. Boston, Little Brown, 1841. (See pp. 304–373)
5. Everdingen, E. van, "The Cyclone-Like Whirlwinds of August 10, 1925." *Verh. Akad. Wet., Amst.*, 28:871–889 (1925).
6. Ferrel, W., *A Popular Treatise on the Winds*, 2nd ed. New York, Wiley, 1893. (See pp. 347–449)
7. Finley, J. P., *Tornadoes*. New York, C. C. Hine, 1887.
8. Flora, S. D., "Tornadoes in Kansas." *Mon. Wea. Rev., Wash.*, 57:97–98 (1929).

9. Garbell, M. A., *Tropical and Equatorial Meteorology*. New York, Pitman, 1947.
10. Hare, R., "Notices of Tornadoes." *Amer. J. Sci. Arts*, 38:73–86 (1840).
11. Hazen, H. A., "Tornado Losses and Insurance." *Science*, 16:15–19, 22–23 (1890). (See pp. 15–16)
12. Henry, A. J., *Tornadoes, 1889–1896*. Report of the Chief of the Weather Bureau for 1895–1896, U. S. Dept. of Agric., pp. xxiii–xl (1896).
13. Humphreys, W. J., *Physics of the Air*, 2nd ed. New York, McGraw, 1929. (See pp. 208–214)
14. Ives, R. L., "Behavior of Dust Devils." *Bull. Amer. meteor. Soc.*, 28:168–174 (1947).
15. Jones, H. L., *The Identification and Tracking of Tornadoes*. Commission Four of the International Scientific Radio Union Meeting, Washington, D. C., April 1950 (unpublished).
16. Lloyd, J. R., "The Development and Trajectories of Tornadoes." *Mon. Wea. Rev. Wash.*, 70:65–75 (1942).
17. Oersted, H. C., "On Water-Spouts." *Amer. J. Sci. Arts*, 37:250–267 (1839).
18. Redfield, W. C., "Remarks Relating to the Tornado Which Visited New Brunswick in the State of New Jersey, June 19, 1835, with a Plan and Schedule of the Prostrations Observed on a Section of Its Track." *J. Franklin Inst.*, 32:40–49 (1841).
19. Showalter, A. K., "The Tornado—An Analysis of Antecedent Meteorological Conditions" in *Preliminary Report on Tornadoes*, 139 pp. U. S. Weather Bureau, Washington, D. C., 1943.
20. Taylor, G. F., *Aeronautical Meteorology*. New York, Pitman, 1938. (See p. 249)
21. Tepper, M., "On the Origin of Tornadoes." *Bull. Amer. meteor. Soc.*, 31:311–314 (1950).
22. U. S. Weather Bureau, *The Thunderstorm*, 287 pp. Washington, D. C., U. S. Govt. Printing Office, 1949. (See Chap. I)
23. Wegener, A., "Beiträge zur Mechanik der Tromben und Tornados." *Meteor. Z.*, 45:201–214 (1928).
24. Willett, H. C., *Descriptive Meteorology*. New York, Academic Press, 1944. (See pp. 279–281)
25. Williams, N. R., "Development of Dust Whirls and Similar Small-Scale Vortices." *Bull. Amer. meteor. Soc.*, 29:106–117 (1948).

THUNDERSTORMS

By HORACE R. BYERS

University of Chicago

Introduction

Thunderstorms are shower clouds or aggregations of shower clouds in which electrical discharges can be seen as lightning and heard as thunder by a person on the ground. Discharges observed on airplanes flying in clouds do not constitute conclusive evidence of thunderstorms unless lightning is observed illuminating the cloud from another source, since most discharges on airplanes are believed to be autogenous in origin, that is, to result from an intensification of the electric field around the airplane brought about by the innumerable collisions with precipitation particles.[1] Since thunderstorms are showers that have reached a stage of producing lightning, it has been recognized for centuries that they represent an intense form of shower or an advanced stage in the development of convection in moist air. Modern observations indicate that thunderstorms will develop when large accumulations of condensed water, presumably liquid and solid, are being carried upward at heights where the ambient temperature is less than about −20C. It is only under conditions of moderate to high temperature and moisture that large accumulations of water can occur in the atmosphere below the −20C isotherm. Therefore thunderstorms are seldom observed without the occurrence of above-freezing temperatures through a considerable portion of the lower atmosphere. They are most likely in tropical air masses warmed from below in areas favorable for ascending motion, such as over mountains or in large-scale synoptic disturbances.

In summer, thunderstorms in the belt of the westerlies form two different distribution patterns easily recognized on long-range radarscopes. One consists of an irregular spacing of individual storms or masses of storms, and the other appears as a line of thunderstorms, usually running more or less parallel to the low-level wind flow. In low latitudes and in the tropics, the first type seems to predominate. In middle latitudes lines of thunderstorms are favored, as shown by the existence in Ohio of thunderstorm lines on thirty-two out of fifty-six days on which thunderstorms were observed by long-range radar. On only six of these days were the lines directly associated with fronts. Figure 1 is a photograph of the radarscope at Jamestown, Ohio, on July 12, 1947, at 3:25 P.M., showing both types of distribution. Thunderstorm echoes appear in irregularly arranged masses to the north and in a distinct line to the south and southeast.

The tendency for thunderstorms to form in lines, even in the absence of fronts, is not well understood.

Thunderstorms are sometimes observed imbedded within a general nimbostratus cloud cover. These are found in overrunning unstable tropical air in connection with warm fronts of extratropical cyclones. They are most common in spring. Sometimes thunder and lightning will be observed with the passage of a cold front or the center of a depression on a dark, stormy day in winter, even when it is snowing, but most commonly only with rain. In maritime climates winter thunderstorms are not unusual, occurring with fronts and depressions or with the characteristic showers of unstable polar-maritime air.

FIG. 1.—Thunderstorm radar echoes at 1436 EST July 16, 1947 of PPI scope at Jamestown, Ohio. Range circles are for each 50 miles; white area in center is local ground return.

The most thorough study of the nature of individual thunderstorms was that accomplished by the large U. S. Weather Bureau-Air Force-Navy-N.A.C.A. Thunderstorm Project from 1946 to 1949, reported in the government publication *The Thunderstorm* [43]. Most of the factual material given in the following pages is taken from that report, but some speculations, leading questions, theory, and new results are added.

Thunderstorm Structure and Circulation

Numerous investigators have noted that the thunderstorm consists of several cells, each having a distinctive convective circulation. Originally the cells have been separated but later they come together. After they have formed a single thundercloud mass the cells can be distinguished by the patterns of vertical motion measured on airplanes flying through them. The cell boundaries are identified as narrow zones of inactive or nonturbulent cloudy air.

1. Consult "Precipitation Electricity" by R. Gunn, pp. 128–135 in this Compendium.

The life cycle of the thunderstorm cell [7] naturally divides itself into three stages determined by the magnitude and direction of the predominating vertical motions. These stages are:

1. The cumulus stage—characterized by an updraft throughout the cell.

2. The mature stage—characterized by the existence of both updrafts and downdrafts, at least in the lower half of the cell.

3. The dissipating stage—characterized by weak downdrafts throughout.

Cumulus Stage. Throughout the cell there is an updraft which is strongest at the higher altitudes and increases in magnitude toward the end of this stage (Fig. 2). Converging air feeds the updraft not only from the surface but also from the unsaturated environment at all levels penetrated by the cloud. Thus, air is entrained into the cloud and is accommodated by the evaporation of some of the liquid water carried in the updraft. This entraining continues throughout all of the stages.

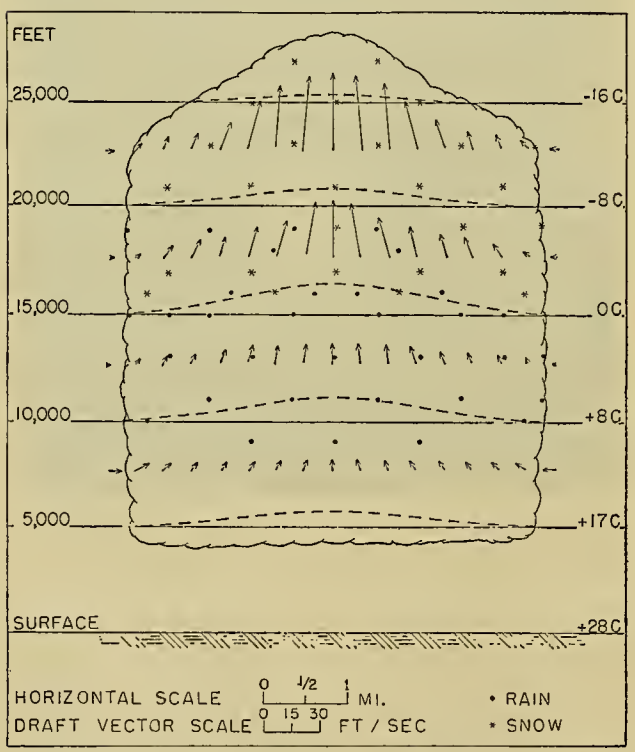


FIG. 2.—Vertical cross section in cumulus stage showing vertical motions, inflow, hydrometeors, and temperature distribution.

In-cloud temperatures in a strongly developing cell are higher than those of the environment at corresponding altitudes. The greatest differences between cloud and environment temperatures are in the regions of strongest updrafts, which are in the upper parts of the cloud and which are especially well developed at the end of the stage.

Although pilots flying through the clouds in this stage report rain or snow, particularly near the end of the stage, these condensation products appear to be suspended by the updraft, since no precipitation is observed at the ground. As a matter of practice, it is found that the end of the cumulus stage and beginning of the mature stage can be signalized by the occurrence of precipitation at the ground. In the cumulus stage there appears to be a considerable concentration of hydrometeors at or slightly above the freezing level in the form of liquid or solid or a mixture of the two. The stronger the updraft the greater is the vertical thickness of the transition zone between water and ice.

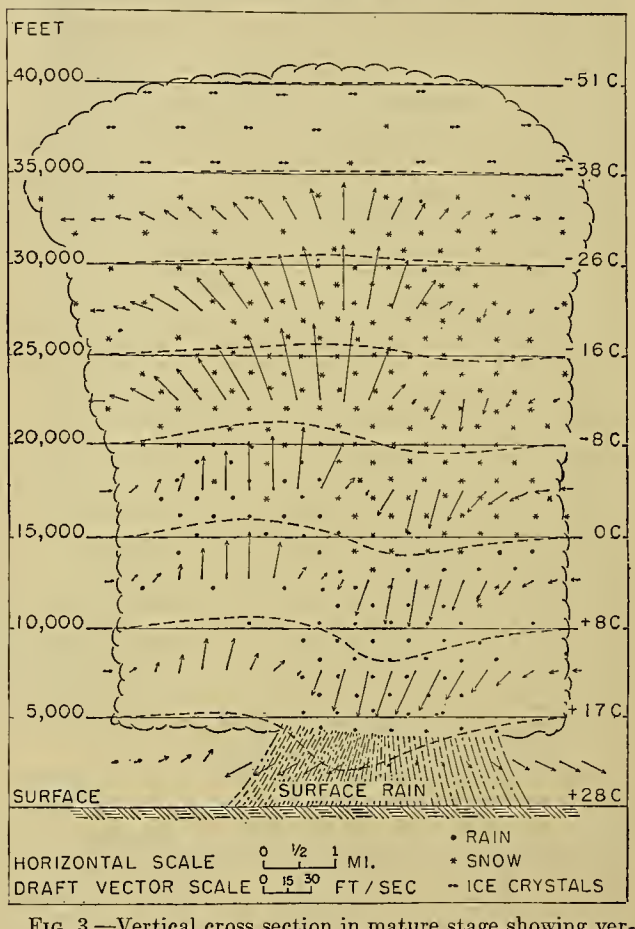


FIG. 3.—Vertical cross section in mature stage showing vertical motions, inflow, hydrometeors, and temperature distribution.

Mature Stage. The mature stage begins when rain first falls distinctly out of the bottom of the cloud. Except under arid conditions, the rain reaches the ground. The size and concentration of the drops or ice particles have now become so great that they cannot be supported by the updraft and they begin to fall relative to the earth. The frictional drag exerted by the precipitation helps to change the updraft into a downdraft which, once started, can continue without this frictional drive, as will be demonstrated later in the discussion of the thermodynamic process. The beginning of the rain at the surface and the initial appearance of the downdraft are nearly simultaneous. The downdraft appears to start in the vicinity of the freezing level, later growing in vertical as well as in horizontal extent (Fig. 3).

The updraft also continues and often reaches its greatest strength in the early mature stage in the upper parts of the cloud. Updraft speeds may locally exceed 80 ft sec^{-1}. The downdraft is usually not as strong as the updraft and is most pronounced in the lower part of the cloud, although weakening near the ground. The descending air is forced to spread laterally at the surface of the earth. Thus areas of rain, downdraft, and horizontal divergence are found together at the ground.

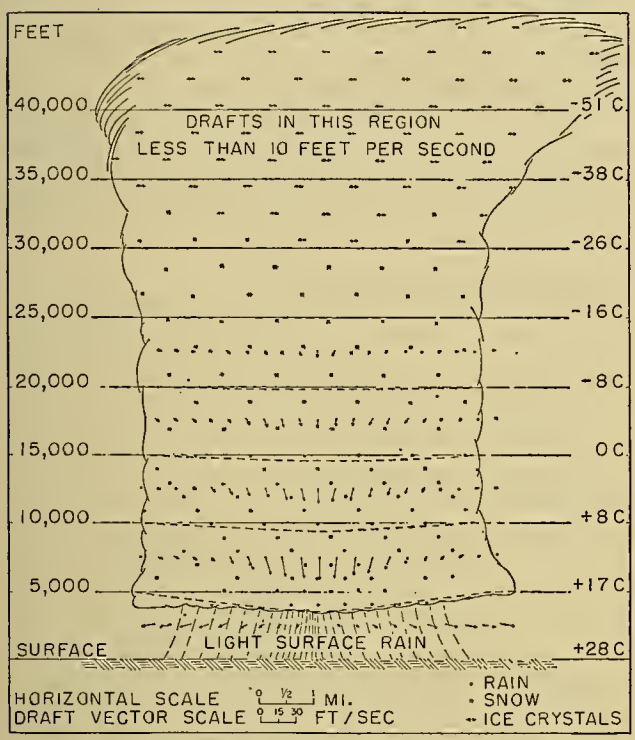


FIG. 4.—Vertical cross section in dissipating stage showing vertical motions, inflow, hydrometeors, and temperature distribution.

Temperatures are low in the downdraft, compared with the environment, and contrast especially with the updraft temperatures. The greatest negative temperature anomalies are found in the lower levels. As might be expected, there is a close association between strong updraft and high temperature and between strong downdraft and low temperature.

In the updraft, mixing of entrained air causes evaporation of some of the liquid water thus removing some of the heat gained from condensation. Instead of the simple wet-adiabatic rate, the updraft air cools at a somewhat greater "entraining-wet-adiabatic" rate. This does not permit the updraft air to become very much warmer than the environmental air. The downdraft seems to be characterized by reversible wet-adiabatic temperature increases in which evaporation counteracts to some extent the compressional effects. Since the downdraft starts out at a temperature very near that of the environment, its wet-adiabatic descent assures that it will be colder than the environment which has a lapse rate greater than the wet adiabatic. The cold downdraft spreads out at the surface as a cold air mass to form the well-known pseudo-cold front advancing against the warmer surrounding surface air.

The mature stage represents the most intense period of the thunderstorm in all its aspects. At the ground heavy rain and strong winds are observed while in the clouds the airplanes encounter at this stage the most severe turbulence, including in addition to the drafts the short, intense accelerations known in aeronautics as "gusts." Hail, if present, is most often found in this stage. From a study of radar echoes it has been concluded that parts of the cloud containing appreciable quantities of water in crystalline form extend to heights of 60,000 ft on some occasions. With very strong updrafts it is possible for liquid water to be carried well above the freezing level. On the Thunderstorm Project in Ohio one case of heavy rain at 26,000 ft, nearly 10,000 ft above the environmental freezing line, was reported.

Dissipating Stage. When the updraft disappears and the downdraft has spread over the entire area of the cell, the dissipating stage begins (Fig. 4). Dissipation results from the fact that there is now no longer the updraft source of condensing water. As the updraft is cut off, the mass of water available to accelerate the descending air diminishes, so the downdraft also diminishes. The entire cell is colder than the environment as long as the downdraft and the rain persist. As the downdrafts give out, the temperature within the cell rises to a value approximately equal to that of the surroundings. Then complete dissipation occurs or only stratified clouds remain. All surface signs of the thunderstorm and its downdraft disappear.

Thermodynamics of Entraining and the Downdraft

Computed and observed inflow rates in American thunderstorms show that the cumulus cloud which develops into a thunderstorm entrains environmental air at a rate of approximately 100 per cent per 500 mb of ascent, that is, it doubles its mass as it rises through a pressure decrease of 500 mb. This is a relatively low rate of entrainment.[2] With very much higher rates of entrainment, the updraft would be theoretically colder than the environment, but this is not found to be the case except in isolated or transitory conditions. Entraining rates greater than 100 per cent per 250 mb would prevent the growth of cumulus clouds to heights great enough to produce thunderstorms in American maritime tropical air. In less favorable air masses the critical entraining rate would have to be lower.

At the present stage of theory and observation of entraining, the updraft seems to follow the required thermodynamic pattern. The downdraft is a special case, however. In Fig. 5 the updraft entraining-wet-adiabatic rate in a typical, well-developed growing cumulus cloud is represented by line $A'B'$ and a typical environmental lapse rate for American tropical air is given by line AB. A saturated parcel displaced downward from C would follow the wet-adiabatic CD, if no

2. Consult "Cumulus Convection and Entrainment" by J. M. Austin, pp. 694–701 in this Compendium.

environmental air is entrained into it. With entraining it would warm at some other, less rapid rate, such as *CE*. If the parcel is dragged downward beyond *D* or *E*, it will become colder than the environment and sink. The frictional drag of the mass of liquid water provides the means whereby a parcel in a thunderstorm can thus be forced below point *D* or *E*, whence it continues as the thunderstorm downdraft. With a large quantity of liquid water available for evaporation, saturation can be maintained in spite of the increasing temperatures during descent and the parcels will reach the ground, arriving there with a temperature several degrees lower than the surface environmental wet-bulb temperature.

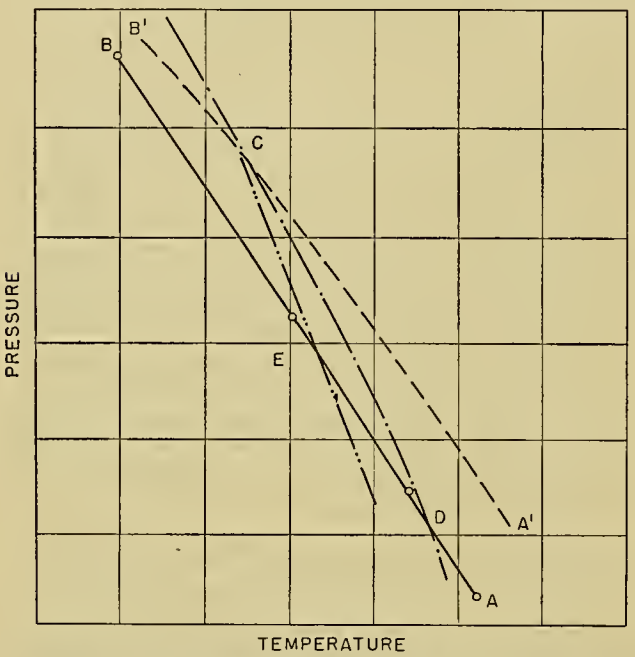


FIG. 5.—Diagram of temperature vs. logarithm of pressure, illustrating thermodynamic process in downdraft.

Thunderstorm Weather near the Surface

Rainfall. The rainfall pattern follows closely the arrangement of the cells and reflects to a considerable extent their stages of development. In some regions, such as the New Mexico area studied by Workman and co-workers [48], single, isolated thunderstorm cells are common. In more humid areas such as the eastern and southern United States, a single-celled thunderstorm is comparatively rare, and when it does occur, it is generally weak and not representative of the average thunderstorm. Usually a group of three or more cells join together to form a thunderstorm and each cell manifests itself in the rainfall pattern. Along with the downdraft and area of horizontal divergence, the rain from a newly developed cell first covers a very small area and then gradually spreads. However, the cold air of the downdraft is able to spread laterally from the cell while the rain falls directly to the ground, so that an expanding outer area of cold air without rain develops. In the dissipating stage this cold-air area continues to expand while the rain area contracts.

If the rainfall is considered with respect to the moving cell, it is found that the duration of moderate or heavy rain from a single cell may vary from a few minutes in the case of a weak, short-lived cell to almost an hour in a large, active one. At a fixed point on the ground the duration of the rain depends upon such factors as the number, size, and longevity of the cells passing over the point, the position of the point with respect to each passing cell, and the rate of translation of the cell. In the eastern and southern United States the average duration of thunderstorm rain at a given station is about twenty-five minutes, although it is highly variable from case to case.

The most intense rain occurs under the core of the cell within two or three minutes after the first measurable rain from that cell reaches the ground and the rain usually remains heavy for a period of five to fifteen minutes. The rainfall rate then decreases, but much more slowly than it first increased. Around the edges of the cell, lesser rainfall rates occur.

Wind Field. Early in the cumulus stage there is a gentle inward turning of the surface wind, forming an area of weak lateral convergence under the updraft. As the cell grows and a downdraft develops, the surface winds become strong and gusty as they flow outward from the downdraft region. The outward-flowing cold air underruns the warmer air which it displaces and a discontinuity in the wind and temperature fields is established. The discontinuity moves outward, pushed by the downdraft, resulting in strong horizontal divergence. Divergence values as high as 8×10^{-3} sec^{-1} or about 1000 times the values observed in intense cyclones, have been evaluated over a small area of the surface micronetworks of the Thunderstorm Project.

The outflow is radial in the slowest-moving storms but in most cases the wind field is asymmetrical with considerably more movement on the downwind side. The prevailing air movement of the lower layers nullifies the radial flow on the upwind side. With respect to the moving cell the outflow may still be radial in character, although not with respect to the ground. Thus the wind discontinuity in most cases is easily detected only in the forward portions of the storm, where it appears as a micro-cold front.

The cold dome of outflowing downdraft air has a form illustrated in Fig. 6. In this sketch the thunderstorm cell is considered to be in the mature stage and is moving from left to right. The cold air is represented as having spread out considerably farther on the downwind side of the cell than on the upwind side, as would be expected in a moving system.

In 15 to 20 min after the outflowing starts, the discontinuity zone has traveled about five or six miles from the cell center. The surface winds near the discontinuity are still strong and gusty but, well within the cold-air dome, the surface wind speeds have decreased so that the strongest winds are no longer underneath the cell itself. A continued settling of the outflow air, transporting momentum downward, causes the wind speed to increase as one approaches the discontinuity zone from within the cold air.

The most interesting of all surface weather features associated with the thunderstorm is the discontinuity zone, which everywhere marks the limit of the cold downdraft air. At a station passed by the cell core in the early mature stage, there is a sharp increase in the wind speed, occasionally to destructive force, and a marked reduction in temperature, occurring with the passage of the discontinuity zone. The sharp increase in wind has been termed the "first gust," since it often appears as the first major gust of a period of strong, gusty winds. After the cold air has spread outward from the cell core in the middle and late mature stage, the wind speeds near the boundary of this air decrease. It is often found that the discontinuity zone is displaced faster than the normal component of the winds immediately behind it. The rapid downflow of cold air toward the ground along the boundary zone thrusts the discontinuity surface forward with a momentum transported from above.

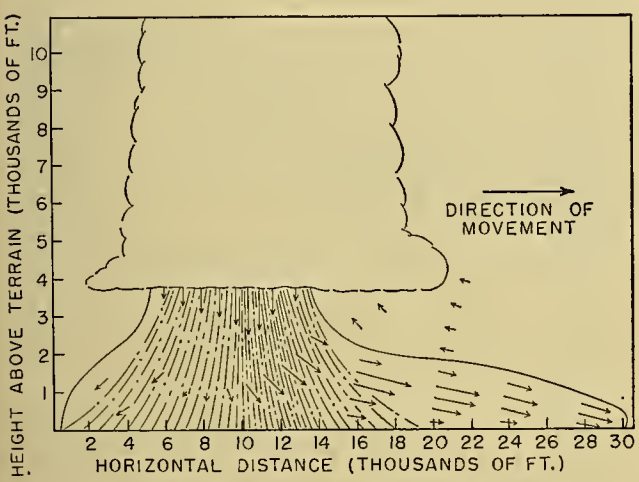


FIG. 6.—Schematic cross section showing formation of cold dome from downdraft. Stippling represents falling rain. Cold-air boundary outlined below the cloud.

With the discontinuity, the wind shows clockwise shifts in most cases. This is especially true in American tropical air currents in middle latitudes where the winds usually are from southwest or south and shift to west or northwest at the discontinuity.

Temperature. The "first gust" and the "temperature break," *i.e.*, the point on the thermogram where the temperature suddenly starts its drop, are two of the most pronounced features observed at the surface, and they occur essentially together. At the time of formation of summer afternoon thunderstorms in the United States the temperature is usually above 85F, often in the 90's. As a result of the rain and the downdraft, the temperature may reach a value as low as 65F without change in air mass. As the thunderstorm activity dies down after sunset, the temperature usually has recovered to an intermediate value representing a mixture of strongly cooled and less cooled or uncooled portions of the air mass.

The area affected by the cooling is many times greater than that over which rain falls, but the temperature change is, of course, most marked in the rain core (downdraft center). Cooling may be detected as much as fifteen to twenty miles downstream. Near the center of a mature cell the temperature reaches a minimum ten to fifteen minutes after the temperature break; farther from the cell the temperature drop is much slower. The amount of the temperature decrease observed in any given storm varies inversely with the distance of the observation point from a cell core. The temperature discontinuity is sharp and well defined under the core but is less pronounced farther away. Since the downdraft is only a few miles in diameter, the area over which the first and most rapid temperature fall occurs is relatively small. As a result, strong horizontal temperature gradients are created after the downdraft first reaches the ground. Gradients exceeding 20F per mile have been observed. As the storm ages, the cold air spreads out and the magnitude of the horizontal temperature gradient decreases. Regardless of the spread of the cold air, the area of minimum temperature remains in the general location where the cold downdraft made its first appearance at the surface, except in cases where a new cell with its own downdraft and rain core develops over another part of the cold dome.

Pressure. Early in the cumulus stage a fall in surface pressure almost invariably occurs. This fall is observed before the radar echo forms, and it is recorded over an area several times the maximum horizontal extent of the echo. When the radar echo appears, the pressure trace levels off in the region directly underneath it, but continues to fall, and frequently at a more rapid rate, in the surrounding areas. The pressure drops in the cumulus stage are usually small in magnitude—less than 0.02 in. (0.67 mb) below the diurnal trend—and take place over a period of 5 to 15 min. Following the fall, the pressure trace remains steady for as long as 30 min.

The pressure falls appear to be caused by the combined effects of vertically accelerated air motions, the expansion of the air due to the release of the latent heat of condensation, and the failure of the convergence near the surface to compensate fully the expansion or divergence aloft. Wind patterns in the vicinity of thunderstorms showed velocity convergence in the developing stages with divergence above 20,000 ft, suggesting a mass balance.

In the mature stage, two features of the pressure trace—the "dome" and the "nose"—are recognized. The dome is registered at all stations to which the cold outflow air penetrates. The pressure nose, the abrupt, sensational rise that some meteorologists regard as typical of the thunderstorm, really occurs only at stations that happen to be passed by the main rain and downdraft just after they have first reached the earth in the beginning of the mature stage. It is superimposed upon or may mark the start of the pressure dome.

The displacement of the warmer air by the cold outflowing air from the downdraft results in the pressure rise, initiating the pressure dome. A study of 206 thunderstorm pressure records from the surface micro-

network showed that in 182 of the traces there was a pressure rise associated with the arrival of the cold outflowing air. Since the rate and total amount of pressure rise depend on the slope of the cold-air mass, the temperature difference between the cold air and the displaced warm air, the depth of the cold air itself, and the speed with which the system travels, the most marked pressure changes are found near the cell-core and decrease with distance from it. The areal extent of the pressure dome is similar to that covered by the cold air. Therefore the pressure remains high for a period of from one-half hour to several hours, depending on the amount of cold air involved.

The distinction between the pressure nose and the pressure dome can be made only with difficulty on the conventional week-long barograph traces, but on the twelve-hour recording drums used on the Thunderstorm Project the distinction is clear.

At any given time the pressure nose was usually detected by but one station per cell in Ohio, indicating that the maximum diameter of the area in which it occurred was less than five miles It is thus apparent that it is caused by a transitory process which exists for only a brief period, namely the time of commencement of the mature stage when rain and downdraft reach the ground. Two effects which are thought to be important in creating the pressure nose are the weight of the suspended and falling water, and a lack of balance between convergence and divergence. Both of these effects are at a maximum at this time. An accumulation of an average of one gram of liquid water per cubic meter from the cloud base to 35,000 ft, a conservative value, would increase the surface pressure by about 1 mb, other factors being equal. When the downdraft becomes established, it reverses at least part of the vertical circulation pattern and reduces the divergence above 20,000 ft, or perhaps even reverses it to convergence. Should this happen before the divergence (outflow) in the surface layers becomes well established, a net increase in mass convergence could result in a substantial surface pressure rise. For example, an uncompensated convergence of 1 hr^{-1} (a conservative value for thunderstorm convergence) within a layer only 1000 ft thick at 20,000 ft would increase the surface pressure at a rate exceeding 1 mb in 3 min. Relationships of pressure changes to vertical motions in thunderstorms, which have been a favorite subject for theoretical treatment by meteorologists, could not be confirmed by data of the Thunderstorm Project. One of the difficulties is that a downdraft at the lower levels is under an updraft at a higher level, so that the vertical accelerations are opposed and tend toward compensating each other.

After the brief pressure nose, the pressure remains at the value of the pressure dome which prevails for the particular thunderstorm. The dome persists through the dissipating stage of the cell, after which the pressure returns to the trend prevailing before the passage of the storm. In the case of a thunderstorm associated with a cold front or a fast-moving squall line, the pressure remains high or even continues to rise as a result of cold-air advection or the passage of a wave in the pressure and wind fields.

Humidity. An extraordinary phenomenon is noted in the hygrograph traces from the surface micronetworks in the form of a sharp decrease in the relative humidity in the midst of the heaviest downpour of rain. With the onset of rain, the relative humidity rapidly approaches but usually does not quite reach 100 per cent, then, in many cases, drops suddenly to values as low as 60 or 70 per cent, returning to near saturation again in a few minutes. Such a fluctuation, termed the humidity dip, and illustrated in Fig. 7, is associated with the rain area. In practically all cases the humidity dip occurred in a region of divergence in the surface wind, therefore in the outflowing downdraft. The temperature was usually decreasing, but if it had already reached its lowest point, as was sometimes the case, a rise in temperature of 2 or 3F would sometimes accompany the humidity dip.

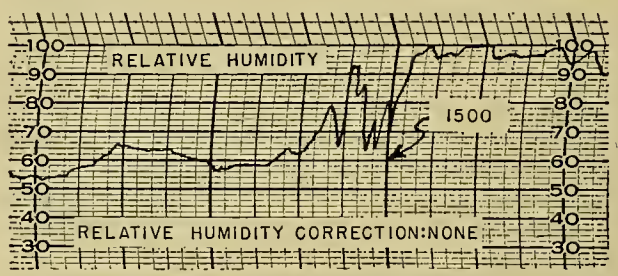


FIG. 7.—Example of humidity "dip" as shown on hygrogram during heavy rain under thunderstorm of August 14, 1947 near Wilmington, Ohio. Time divisions are for 5-min intervals.

The failure of surface air to become saturated during the heavy rain, together with the frequently observed relative-humidity dip, suggests that the downdraft air fails to maintain saturation as it descends, even in the presence of large concentrations of liquid water. Two processes are suggested: first, desiccation of the air by condensation on cold precipitation particles; second, a time lag in the evaporation of the waterdrops so that there is insufficient accommodation to the increase in the saturation mixing ratio as the air descends to lower levels.

Measurements of raindrop temperatures at three feet above the ground show that in afternoon thunderstorms the rain temperature in the first few minutes of the rain period is several degrees lower than that of the ambient air. This would be sufficient to keep the relative humidity from reaching 100 per cent. If the second process mentioned above is taking place, one would expect that the air of the thunderstorm downdraft would be heated at a rate between the moist- and dry-adiabatic and would reach the ground in an unsaturated state.

Development of New Cells. From a study of numerous cases of new cell development, the action of the cold outflowing air appears unmistakably as causing or contributing to the new growth. When two cells in the mature stage are within a few miles of each other, the cold outflows collide. The greatest frequency of new

cell development is in the area between two existing cells whose edges are three or less miles apart. A three-mile band downwind is next in importance, then the lateral edges, and least frequently the upwind or rear side.

In some cases the time interval between the beginning of the outflow and the appearance of the new cell on the radarscope is too short to permit explanation of the new development as a result of the underrunning cold air or similar time-consuming process. There are cases, as indicated by the radar echoes, in which one or a cluster of new cells comes into existence almost simultaneously with the initial or parent cell, which suggests that a preferred region of convergence and ascent favors the development.

Cloud Structure in Relation to Thunderstorm Electricity

Atmospheric electricity and its phenomenal aberrations in thunderstorms are treated in other articles in this Compendium.[1,3] A few observations might be made at this point relating the meteorological thunderstorm structure and dynamics to the electrical structure and electrodynamics.

In recent years abundant information has been presented indicating that thunderstorm electrification is somehow associated with the ice phase in the cloud or, more probably, with a heterogeneous mixture of liquid and solid condensates. For example, recent studies by Workman and Reynolds [48] and by the U. S. Thunderstorm Project [43] show that lightning does not occur until the visible cloud top has ascended to where the temperature is less than about −28C or the top of the echo seen on a 3-cm height-finding radar at a range of twenty miles has surpassed the isotherm of −20C. A number of studies indicate that the upper positive-charge center is located where the temperature is near −20C and the lower negative center is at a temperature of 0 to −10C (roughly). Between these two centers is the lightning hearth region, that is, the region where the first lightning develops. In these temperature ranges mixtures of liquid and solid precipitation particles are commonly observed in growing and established thunderstorms.

Workman and Reynolds [49] discovered an effect, earlier indicated indirectly by Dinger and Gunn [11], involving the freezing of water, which offers a clue as to the processes which might produce electrification in the freezing parts of thunderstorms. They found that water containing contaminants which are likely to be found in the atmosphere could produce, as it freezes, potentials of more than 150 volts across the ice-water interface. Extremely dilute solutions of the order of 10^{-6} to 10^{-3} normal, constituting very pure water by industrial standards, produce the desired effect. Calcium hydrocarbonate, $CaH(CO_3)_2$, formed from calcium carbonate in the presence of carbon dioxide, both found in the atmosphere, is observed to transfer negative ions across the interface to the ice, leaving the water positive and the ice negatively charged. Ammonium compounds and acids or substances with high *p*H produce the opposite polarity, that is, with the water negative and the ice positive. Workman suggests that in the cloud some of the water freezing around an ice pellet is torn off and carried upward as small droplets in the cloud while the frozen core, being larger and heavier, falls. If the cloud particles contain a contaminant, such as calcium hydrocarbonate, which produces positive water and negative ice in the freezing process, the carrying upward of the positively charged portion and the descent of the negative ice would produce the observed polarity of thunderstorms at these levels.

Observations by Weickmann [44] and Kuettner [19] on the Zugspitze in the Bavarian Alps, where the observatory was at or near the freezing level in most thunderstorms, indicate that a predominant form of solid precipitation other than ordinary snow is a graupel made up of an agglomeration of tiny supercooled droplets to form rime clusters. An ice core with a film of liquid around it, as might be expected in a hailstone, would be required for Workman's hypothesis. Further observations of the exact nature of precipitation particles in the freezing regions of thunderstorms are needed in order to clarify the discussion. The precise recognition of these precipitation forms is next to impossible from an airplane.

Kuettner also found that the region of greatest lightning activity coincided with the downdraft and heavy rain core. Furthermore in this center of activity there was frequently a small center of positive charge embedded within the main lower negative charge region and centered on the 0C isotherm. This corresponds to some extent with the picture given by Simpson and Robinson [32] of an isolated lower positive-charge center. Observations of what appeared to have been a lower positive charge in an area of heavy rain and downdraft were obtained in a series of traverses through the lower part of a thunderstorm on one of the Thunderstorm Project flights. This is another phenomenon that needs further investigation.

In seeking to explain the electrification of thunderstorms, most scientists have looked to aspects of the updraft. However, the Thunderstorm Project data indicate that the greatest lightning activity is in the region of downdraft and heaviest rain, as noted by Kuettner. It is not entirely logical to conclude from this, however, that the charge-generating mechanism occurs in the downdraft, because the boundary separating updraft from downdraft in the mature stage is not vertical and a region of updraft may be found over a downdraft.

On the Thunderstorm Project a study was made using recording point collectors located on the ground and a height-finding radar scanning the thunderclouds from a point about twenty miles away. The following conclusions concerning lightning were indicated by the data:

1. The cloud tops (echo tops) reach a height where the environmental temperature is around −20C before the first lightning occurs.

3. Consult "Universal Aspects of Atmospheric Electricity" by O. H. Gish, pp. 101–119.

2. The maximum lightning frequency occurs at the same time as the cell reaches its maximum height.

3. As the height of the cell top decreases, the frequency of lightning also decreases.

4. It appears that a greater cell height (or lower temperature of the cloud top) is necessary to initiate lightning than is required to maintain it once it has started.

5. In the life cycle of a thunderstorm the maximum frequency of lightning precedes the time of its maximum 5-min rainfall.

This last conclusion may have some meaning in connection with the possibility of the suspension of the raindrops by the electric field.

The Thunderstorm as Disclosed by Radar

A number of studies of thunderstorms have been made by radar. A large amount of useful data was obtained from the radars of the Thunderstorm Project. From the photographic records of the range-height-indicating (RHI) radarscope it was possible to obtain data on the altitude of first formation of 66 radar echoes from convective clouds. A graph was made of the frequency distribution of the differences between altitudes of the tops of the initial echoes and the concurrent heights of the freezing level. The distribution showed a pronounced mode at 1000 ft above the freezing level but the mean was at +2200 ft. In convective clouds the echoes appear abruptly after the cloud has been visible to field observers for some time, suggesting a sudden release of great quantities of large waterdrops after penetration above the freezing level, in accordance with the Bergeron theory.

It was found that the rate of vertical growth of the top of a radar cloud echo agrees closely with the updraft velocity measured in that portion of the cloud if a correction factor of about 3 ft sec^{-1} is added to the radar growth rate to account for the relative free fall of the attenuating snowflakes or other particles. The use of radar in this manner for measuring updrafts appeared to have a practical application in detecting hail possibilities, since all observed cases of hail accompanied strong updrafts.

It was noted that the thunderstorm tops ascended in a series of steps, appearing as the growth of new protuberances or "turrets," during their growth and, as a matter of fact, in 32 cases studied, the maximum height reached was correlated with the number of turrets thus formed by a coefficient of $+ 0.67 \pm 0.10$. Each successive turret was higher than the preceding one, and the mean lapse of time between successive turret peaks was 17.8 min. It is believed that each turret makes it easier for the following ones by increasing the moisture content in the cloud-top environmental air which must be entrained in further growth. Less heat of condensation is robbed from the new turret by entraining than would be the case if it were standing alone in a dry environment. Each turret underwent a growth period followed by an interval of subsidence. The growth period averaged about 16 min. The average vertical growth rate was about 18 ft sec^{-1} and the subsidence rate about 12 ft sec^{-1}.

The mean of the maximum heights reached by 199 Ohio storms as indicated from the radar was 37,500 ft. There was no significant difference in this respect between different types of thunderstorms, such as air-mass, squall-line, or frontal.

The boundaries of the radar cloud echoes were found to agree closely with the visible cloud limits except in the levels near the cloud base where non-echo-producing "outrider" clouds were common. Also the anvils and other layer-type lateral extensions usually did not return a radar echo. A positive correlation was found between the horizontal and vertical extents of thunderstorms; that is, the taller the cloud, the broader it was. The greatest areal coverage was at an altitude around 10,000 ft. The cross-sectional area was slightly less at the lower level and tapered at the top, forming a total cloud echo shaped like a rosebud.

From radar scans covering an area of over 55,000 square miles, it was found that on average thunderstorm days in Ohio 10 per cent of the area would be covered by cloud echo, and on a day of maximum thunderstorm activity, 40 per cent of the area would be covered. It is found from indirect comparison that at 20,000 ft the in-cloud areas would average 5 per cent and have a maximum of 22 per cent, indicating the better chance of avoiding thunderstorms by flying high. (Other measurements showed that these high levels are the worst choices for flights *within* thunderstorms.)

Effect of Environmental Wind Field

The causes and effects of vertical shear, including effects of entrainment and momentum transfer in the drafts, have been investigated by Byers and Battan [6] and Malkus [20].

With the aid of long-range radar and abundant upper-air wind data, the movements of thunderstorms in Florida and Ohio were studied in relation to the wind fields in which they were embedded. A method was devised whereby the translational component of the motion could be separated from the growth or dissipative component.

It was found that the motion of the storm corresponded most closely to the mean vector wind between the gradient level and 20,000 ft. The correlations were better in Ohio than in Florida owing to a stronger, more consistent wind flow in the former region. In direction, the correlations were 0.95 or better in both regions. It was found that the speed of the cloud was less than that of the vector mean wind from gradient to 20,000 ft. When the two speeds are plotted, a nearly straight line is formed, but the relationship is better represented as

$$U_w = 1.9 + 0.65U_c + 0.020U_c^2,$$

where U_w is the vector mean wind speed and U_c is the cloud speed.

Preferred Areas of Thunderstorm Development

The movement of thunderstorm cells is closely related to the question of preferred areas for their new develop-

ment, for movement is in part comprised of propagation which in turn results from new growths or extensions. In the preceding section, measurements were described in which the propagation and dissipation components were removed from the thunderstorm motions. When this is not done erratic motions are indicated. At the end of the section on thunderstorm weather near the surface it was shown how certain areas immediately surrounding existing cells are favorable for the formation of new ones. These frequently become attached to form a larger thunderstorm mass. In mountainous areas the propagation components in the motion usually outweigh the wind components. Thus thunderstorms are seen to propagate along a mountain range even though the wind is blowing across the range at all levels, and sometimes there will be few or no cells drifting into the valley so that the wind component in the motion is relatively ineffective.

Certain features of the earth's surface provide an ideal location for new thunderstorm development. Mountains and, in fact, rugged relief of any kind as contrasted with smooth terrain will be favorable. Islands, peninsulas, and other pronounced heat sources are favorable for the formation of afternoon storms. In Florida it is found that the convergence of the sea air over the peninsula from both sides in the afternoon is favorable for thunderstorm formation [10]. Locally on the peninsula the dry land areas as contrasted with the numerous swamps and lakes are most favorable for new formation.

In Ohio, a study of the initial appearance of 584 radar clouds on 21 days showed a larger number of echoes from day to day over the rougher parts of the terrain. However, plenty of new echoes formed over the flattest, smoothest areas. Neither in Ohio nor in Florida was it possible to find a well-marked region from which the first thunderstorm of a series would develop. The thunderstorm hearth (*Gewitterherde*) described by von Ficker [13] was not apparent.

In many parts of the world popular notions have developed concerning an apparent tendency for thunderstorms to follow rivers. The idea is often expressed that the moisture provided by the river favors propagation along the watercourse. Scientific studies indicate that, where thunderstorms propagate along rivers, conditions of the terrain rather than the presence of the water are the factors which are favorable for new growth. Many rivers have bluffs along their flanks and the immediately surrounding terrain is roughly dissected by tributary streams. In Ohio the maximum number of new formations occurred over the sharply dissected plains. The area studied is predominantly a peneplain at about 1000 ft through which streams cut down to the Ohio River (about 500 ft at Cincinnati).

Squall Lines

During the 127-day period from May 17 to September 21, 1947, there were in Ohio 56 thunderstorm days on which extensive radar data were obtained. On 32 of these days lines of thunderstorms were observed. The lines on 6 days occurred along surface fronts, on 19 days were ahead of surface cold fronts, and on the remaining 7 days apparently had no connection with the surface fronts. As is usual throughout the eastern United States, the pre-cold-front squall line was the predominating scene of thunderstorm activity.

It was found that the pre-cold-front squall line is not quite parallel to the cold front but has an orientation averaging 13 degrees clockwise from that of the front. Its orientation is somewhat counterclockwise from the orientation of the overlying 700-mb contour lines. The movement of the squall line, as distinct from the movement of the individual storms composing the line, is not very closely related to the speed of the following cold front; in many cases it moves faster than the front. The individual elements of the line, or storms, move with the prevailing upper winds as described in a preceding section.

The pre-cold-front squall line on the synoptic chart is frequently observed to last for periods as long as twenty-four hours, during which it may travel several hundred miles. When viewed on the radar, it was found that there is usually a zone of convective activity comprising several lines and, although the zone may persist for a considerable period of time, the lines within it are constantly undergoing change—new lines are forming and older ones are dissipating.

The elements in a line range from a single isolated echo one to two miles in diameter to a large aggregate of storms appearing on the scope as a single, more or less homogeneous echo having a maximum dimension of over thirty miles. The number of separate elements in a line varied from 4 or 5 small echoes in the early stages to as many as 40 or 50 at the time of maximum echo intensity. For the most part, elements were found to form and dissipate on a given squall line, suggesting the existence of a preferred line formation.

The squall line or, more properly, the squall-line zone was found to be at an average distance of 170 miles ahead of the cold front, with extreme values of 80 and 325 miles. Generally there was an area of relative inactivity between the cold front and the trailing edge of the squall zone; but on one or two occasions the convective activity finally extended back to the front itself.

Generally, the thunderstorms associated with the squall lines were little different from any others except for a tendency to be more severe, especially in producing effects at the surface, such as strong, gusty winds. Since the storms come in a line, there is a more or less connected discontinuity line from the cold outflowing downdraft which is often mistaken for a true front. The return to tropical air-mass conditions afterward shows the local character of the discontinuity. The pressure "dome" of the combined storms often results in the creation of a long pressure ridge along the squall line. Sometimes a squall line of this character may after a time produce a true frontal discontinuity involving more cold air than can be accounted for by the downdraft.

Of the seven squall lines that were not associated with fronts, three were related to some type of large-scale cyclonic disturbance, two were on the west side

of a warm anticyclone and the other two were in rather flat, uniform pressure fields. These lines were usually short compared with the prefrontal types and seemed to be less well defined.

Remarks on Thunderstorm Forecasting

It is not practical in this article to discuss all of the important points related to the forecasting of thunderstorms. The problem is intimately connected with the general subject of forecasting treated elsewhere in this Compendium.

Historically, thunderstorm forecasting has been of unique interest because it seemed to be the one atmospheric disturbance that could be treated quantitatively from a thermodynamic consideration of the buoyancy forces in an atmosphere in labile equilibrium. A series of rival thermodynamic diagrams was developed by different meteorologists, largely stimulated by the possibilities of forecasting convection through thermodynamic analysis of upper-air soundings. Refinements of the parcel method and finally the introduction of the slice method and variations of it were aimed at this problem.[2,4]

Some meteorologists drew attention away from the analysis of individual soundings to studies of upper-air charts which they believed would hold the key to thunderstorm forecasting. Namias [21, 22] introduced the isentropic chart as a means of forecasting thunderstorms, showing that the occurrence of summer thunderstorms depended on the presence of a moist tongue on isentropic surfaces in the vicinity of a potential temperature of 315K, usually found at an altitude around 3 to 4 km in the eastern United States in summer. The moist tongues, with dry "tongues" or areas between them, occur over tropical air masses that are more or less homogeneous horizontally in the low levels. Thus the upper-air humidities appear to be critical. Namias' work implied a mechanism similar to what is now known as entrainment in order to account for the arrested growth of convection in a dry upper-air environment.

The picture seems to be somewhat as follows. While on the lower isentropic surfaces there seem to be only slight gradients of water-vapor content within the tropical air masses, there is great variability of moisture on the surfaces around 315K. These variations can be tracked from one isentropic chart to the next in the form of moving moist and dry tongues. Analyses of separate soundings in terms of stability may lead to error because a dry tongue may replace a moist tongue or vice versa in the forecast period at a station.

If a strong dynamic cause for ascent exists in the tropical air mass, water vapor will be carried aloft to isentropic surfaces where it existed only in small quantities before. Thus, in addition to a chart of the moist tongues existing aloft some study of the possibilities of strong low-level convergence with accompanying ascent should be included in the forecast preparations. There is evidence to indicate that the two main summer moist tongues of the United States are caused by convergence and ascent over (1) the Mexican Plateau [45] and (2) the Florida Peninsula [10]. Some meteorologists conclude that thermal instability may be a necessary condition for thunderstorms but is not a sufficient one. A dynamic process, inducing ascent, is also considered to be required. If remote convergence and ascent which supplies the extensive moist tongues is included in this dynamic process, the conclusion probably is correct. Direct studies of the wind field disclose local areas of low-level convergence if their size is in the proper relationship to the density of the observation network.

The Thunderstorm and the Airplane

Besides the usual difficulties of flight in clouds, the thunderstorm presents the additional, unique hazards of lightning, hail, and heavy turbulence. Flight records show that turbulence is the most predominant danger in thunderstorms and may be the principal cause of thunderstorm accidents. The difficulties of maintaining proper flight attitudes or air speeds within highly turbulent clouds may lead to loss of control or structural damage. It is believed that hail damage is second in importance as a hazard, and that lightning is third. Since hail and lightning are covered in other articles in this Compendium, thunderstorm turbulence effects will be emphasized here.

It is convenient to recognize two classes of turbulence —gusts and drafts, corresponding to two fairly distinct types of response experienced on an airplane and measured by two different techniques. In a draft, the airplane is displaced in altitude in one direction over several seconds of time because of the mean upward or downward motions of the air. Gusts subject the airplane to a series of sharp accelerations without a systematic change in altitude. These accelerations are caused by abrupt changes in velocity of the drafts and by small vortices or whirling masses of air. The larger gusts are invariably associated with strong drafts. If the airplane is flown at constant power setting, attitude, and air speed, the draft velocities can be measured by rates of vertical displacement shown on a recording altimeter. The gusts can be obtained from an accelerometer trace. The techniques of both measurements have been developed to a high degree of reliability by the Gust Loads Section of the National Advisory Committee for Aeronautics and are described in N.A.C.A. publications [12].

In order to provide a gust measure that is applicable to studies of the gust loads imposed on airplanes, aeronautical engineers determine the "effective gust velocity." In simplest terms, this may be defined as the vertical component of the actual velocity of a "sharp-edged gust" that would produce the acceleration in question on a particular airplane flown in level flight at a known air speed and a given air density. The effective gust formula is of the form

$$U_e = K \frac{a}{\rho_0 V (\rho/\rho_0)^{\frac{1}{2}}} \text{ ft sec}^{-1},$$

4. Consult "Thermodynamics of Clouds" by F. Möller, pp. 199–206 in this Compendium.

where K combines a number of factors relating to the airplane used, a is the acceleration increment in g units, ρ and ρ_0 are air density at flight level and sea level, respectively (slugs ft^{-3}), and V is the true air speed (ft sec^{-1}).

The relationship shows that a given effective gust velocity will produce a greater acceleration the higher the speed of the flight. Through substitution of the values of airplane characteristics contained in K, the accelerations produced on one airplane can be translated into the accelerations on another of different characteristics.

From 1362 traverses through thunderstorms on the Thunderstorm Project it was determined that if on any one of these traverses the mean maximum U_e per 3000 ft of flight exceeded 8 ft sec^{-1} and four gusts per 3000 ft, the aircrews reported the turbulence as heavy. The flights were made by crews highly experienced in bad-weather flying. This information proved useful in convincing pilots and meteorologists that data collected by N.A.C.A. in which gust velocities rarely exceeded 25 or 30 ft sec^{-1}, really represented violent thunderstorms. The highest gust velocity recorded on any of these flights was 43 ft sec^{-1} which, incidentally, exceeds the maximum gust load that many airplanes are built to withstand at cruising speeds, free of maneuver loads. By contrast, sustained drafts of greater than 30 ft sec^{-1} were not uncommon and on a few occasions the values were in the vicinity of 90 ft sec^{-1}.

A significant difference among the different levels flown was noted in the magnitudes of gusts and drafts and in the frequency of high values. The lowest values of both gusts and drafts were, in the mean, found at the lowest standard flight levels of the project—6000 ft above ground in Florida and 4000 ft above ground in Ohio. A statistical treatment of the data shows a tendency toward increasing turbulence with height up to the highest levels flown (25,000 or 26,000 ft), but the differences between the various levels from 10,000 ft upward in this respect is slight. The draft velocities showed a distinct maximum at the highest levels. In general, the values of both types of turbulence were greater in Ohio than in Florida, possibly due to a more successful effort to get the airplanes into the thunderstorm before the very intense early-mature stage had passed. The mean of the maximum effective gust velocity above 4 ft sec^{-1} in 3000 ft of traverse in Ohio was 9.4 ft sec^{-1} and in Florida 8.9 ft sec^{-1}.

By the use of a range-height-indicating radar, measurements have been made on the rate of growth of thundercloud tops, some of which have been observed to extend above 55,000 ft. These measurements show that the mean rate of growth, up to an altitude about 10,000 ft below the top of the individual storm, increases with height. The rate of growth of these cloud tops is also a measure of updraft velocities. Therefore one might assume that the updrafts and, with them, the gust velocities, increase with height, at least during the growing stages, up to a level about 10,000 ft less than the maximum height reached by the storm cell.

Finally, it has been shown that areas of highest water concentration in a thunderstorm are the areas of heaviest turbulence. This supports the idea that radar can be useful for avoiding areas of excessive turbulence. In flights below the cloud base, the heaviest turbulence will be found where the darkest rain columns are seen.

Unsolved Problems and Future Research Needs

Although the research in the three years from 1946 to 1949 added more details to our knowledge of thunderstorms than had been accumulated in many decades previously, it also focused attention on a number of unsolved problems.

Some of the questions can be settled by further detailed observations or measurements. One question of detail having fundamental importance for the whole circulation and energy problem is brought up by a finding of the Thunderstorm Project that the strongest downdrafts as well as updrafts are frequently found in the upper parts of the cells. No rational picture of a thunderstorm cell with the downdraft decreasing downward has been devised. Additional data should be gathered, since the number of measurements of downdrafts is small at the high altitudes. Because of difficulty in maintaining the attitude of the airplane, more than 30 per cent of the drafts could not be evaluated. Another detail that could be studied concerns entrainment. It is not known to what extent entrainment involves lateral mixing. This could be solved by obtaining the horizontal gradient of water vapor from some distance outside the cloud up to its very edge by means of an airplane carrying a sensitive dew-point hygrometer. Further measurements are also needed concerning the tendency toward desiccation of the downdraft air, as indicated by the "humidity dip" at the ground in the rain core.

Additional observations should be made on thunderstorms in arid regions, especially in those cases where the condensation level is very high and the rain may sometimes evaporate before reaching the ground. The water and energy budgets of such cells must be radically different from those of humid regions. The problem of hail has not been solved in relation to thunderstorms. Hail, unless in the form of stones more than a centimeter in diameter, is hard to detect in a fast-flying airplane, since heavy rain itself makes a great deal of noise. The Thunderstorm Project did not obtain data from the region of maximum hail occurrence of the Great Plains and Rocky Mountain states. The problem of hail in the generation of thunderstorm electricity is a critical one. In spite of the nearly 200 years that have elapsed since Benjamin Franklin discovered that lightning was a form of electricity, we still are not sure what causes it.

The squall line and tornadoes in relation to thunderstorms are a wide-open field for investigation. Tepper [37] and Newton [23] have attacked these problems from novel viewpoints. The dynamics of the pre-cold-front squall line—whether it is a hydraulic jump phenomenon as emphasized by Tepper or has other special characteristics—need further investigation. In this connection it is interesting to note that, from radar photographs, it appears that squall lines crossing the isobars

at a slight angle and therefore having a spiral appearance, also occur in hurricanes and typhoons. Is there some predisposition for extreme convection in the atmosphere to occur in lines? If so, why?

At this writing, computations of the water budget and the energy budget reveal problems of a fundamental nature. One finding is that under average conditions only 10 per cent of the total water involved in a thunderstorm reaches the ground as rain. This is not only of significance to hydrologic studies but also places limitations on computations of the amount of precipitation producible by artificial nucleation. Computations of the energy budget of the thunderstorm show that ideas about thermal instability and its prognostic value not only need to be revised but may have to be abandoned altogether.

Two factors having no very direct connection with the thermal stability of the atmosphere appear to be important for the formation of thunderstorms: (1) the occurrence of local regions of convergence, often too small in area to be detected by the existing network of upper-wind stations, and (2) the absence of low moisture in the vicinity of the 315K potential-temperature surface. Why other potential-temperature surfaces are not very critical in American summer conditions is not understood. Perhaps the question is related to the fact that 315K is about the potential temperature of the dry-adiabatic midafternoon lower troposphere over the western plateaus. The importance of the Mexico-Arizona-New Mexico moist tongue is well known. It should again be pointed out, also, that convergence and high-level moisture are related.

The studies point to the desirability of undertaking further investigations of isentropic or similar charts and of using a net of upper-wind stations with a 50-mile or, at most, 100-mile mesh to obtain usable divergence-convergence charts.

Finally, one is not sure why thunderstorms should ever occur. The atmosphere seems to be able to take care of the vertical heat exchange without such violent manifestations. Since thermal instability seems to be a necessary but not a sufficient condition for thunderstorms, ordinary cumulus convection appears to be capable of taking care of the situation. As the thermal instability increases, why can't the necessary overturning be accomplished by a great number of fast-circulating cumulus of small diameter rather than a few big thunderstorms which never seem to cover more than about 50 per cent of an area of about 55,000 square miles such as shown on radar at ranges between 20 and 50 miles? The occurrence of convergent wind flow and forced ascent seems to be the answer.

REFERENCES

1. Andre, M. J., "Distribution of Instability Weather Types with Respect to Summer Cold Fronts in the United States." *Bull. Amer. meteor. Soc.*, 30: 228–230 (1949).
2. Austin, J. M., "A Note on Cumulus Growth in a Non-saturated Environment." *J. Meteor.*, 5: 103–107 (1948).
3. Brooks, C. F., "The Local, or Heat, Thunderstorm." *Mon. Wea. Rev. Wash.*, 50: 281–287 (1922).
4. Brunk, I. W., "Pressure Pulsations of April 11, 1944." *J. Meteor.*, 6: 171–178 (1949).
5. Byers, H. R., "Nonfrontal Thunderstorms." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 3 (1942).
6. —— and Battan, L. J., "Some Effects of Vertical Wind Shear on Thunderstorm Structure." *Bull. Amer. meteor. Soc.*, 30: 168–175 (1949).
7. Byers, H. R., and Braham, R. R., Jr., "Thunderstorm Structure and Circulation." *J. Meteor.*, 5: 71–86 (1948).
8. Byers, H. R., and Hull, E. C., "Inflow Patterns of Thunderstorms as Shown by Winds Aloft." *Bull. Amer. meteor. Soc.*, 30: 90–96 (1949).
9. Byers, H. R., Moses, H., and Harney, P. J., "Measurement of Rain Temperature." *J. Meteor.*, 6: 51–55 (1949).
10. Byers, H. R., and Rodebush, H. R., "Causes of Thunderstorms of the Florida Peninsula." *J. Meteor.*, 6: 275–280 (1948).
11. Dinger, J. E., and Gunn, R., "Electrical Effects Associated With a Change of State of Water." *Terr. Magn. atmos. Elect.*, 51: 477–494 (1946).
12. Donely, P., "Summary of Information Relating to Gust Loads on Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1976, 145 pp. (1949).
13. Ficker, H. v., "Über die Entstehung lokaler Wärmegewitter, 1. Mitteil." *S. B. preuss. Akad. Wiss., Phys.-Math. Kl.*, SS. 28–39 (1931).
14. Gunn, R., "Electric Field Intensity Inside Natural Clouds." *J. appl. Phys.*, 19: 481 (1948).
15. Harrison, H. T., and Orendorff, W. K., *Pre-coldfrontal Squall Lines*. United Air Lines Meteor. Dept. Circ. No. 16, 1941.
16. Harrison, L. P., "Lightning Discharges to Aircraft and Associated Meteorological Conditions." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1001 (1946).
17. Helfand, B. B., "Meteorological Conditions Associated with Flight Measurements of Atmospheric Turbulence." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1273 (1947).
18. Koschmieder, H., "Über Boën." *Meteor. Z.*, 61: 244–247 (1944).
19. Kuettner, J., "The Electrical and Meteorological Conditions Inside Thunderstorms." *J. Meteor.*, 7: 322–332 (1950).
20. Malkus, J. S., "Effects of Wind Shear on Some Aspects of Convection." *Trans. Amer. geophys. Un.*, 30: 19–25 (1949).
21. Namias, J., "Thunderstorm Forecasting with the Aid of Isentropic Charts." *Bull. Amer. meteor. Soc.*, 19: 1–14 (1938).
22. —— *An Introduction to the Study of Air Mass and Isentropic Analysis*. Amer. Meteor. Soc., Milton, Mass., 1940.
23. Newton, C. W., "Structure and Mechanism of the Prefrontal Squall Line." *J. Meteor.*, 7: 210–222 (1950).
24. Normand, Sir Charles, "Energy in the Atmosphere." *Quart. J. R. meteor. Soc.*, 72: 145–167 (1946).
25. Press, H., "A Statistical Analysis of Gust-Velocity Measurements as Affected by Pilots and Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1645 (1948).
26. —— "An Application of the Statistical Theory of Extreme Values to Gust-Load Problems." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1926 (1949).
27. —— and Binckley, E. T., "A Preliminary Evaluation of the Use of Ground Radar for the Avoidance of Turbulent Clouds." *Tech. Notes nat. adv. Comm. Aero.. Wash.*, No. 1684 (1948).
28. Press, H., and Thompson, J. K., "An Analysis of the Relation Between Horizontal Temperature Variations and

Maximum Effective Gust Velocities in Thunderstorms." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1917 (1949).

29. Rhode, R. V., "Gust Loads on Airplanes." *J. Soc. automot. Engrs., N. Y.*, 40: 81–88 (1947).
30. Schaffer, W., "The Thunderstorm High." *Bull. Amer. meteor. Soc.*, 28: 351–355 (1947).
31. Schmidt, F. H., "Some Speculations on the Resistance to the Motion of Cumuliform Clouds." *Meded. ned. meteor. Inst.*, (B) Deel 1, Nr. 8 (1947).
32. Simpson, G. C., and Robinson, G. D., "The Distribution of Electricity in Thunderclouds, II." *Proc. roy. Soc.*, (A) 177: 281–329 (1941).
33. Simpson, G. C., and Scrase, F. J., "Distribution of Electricity in Thunderclouds." *Proc. roy. Soc.*, (A) 161: 309–352 (1937).
34. Stommel, H., "Entrainment of Air into a Cumulus Cloud." *J. Meteor.*, 4: 91–94 (1947).
35. Suckstorff, G. A., "Kaltlufterzeugung durch Niederschlag," *Meteor. Z.*, 55: 287–292 (1938).
36. —— "Die Ergebnisse der Untersuchungen an tropischen Gewittern und einigen anderen tropischen Wettererscheinungen." *Beitr. Geophys.*, 55: 138–185 (1939).
37. Tepper, M., "A Proposed Mechanism of Squall Lines: The Pressure Jump Line." *J. Meteor.*, 7: 21–29 (1950).
38. Thompson, J. K., and Lipscomb, V. W., "An Evaluation of the Use of Ground Radar for Avoiding Severe Turbulence Associated with Thunderstorms." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1960 (1949).
39. Tolefson, H. B., "Preliminary Analysis of N.A.C.A. Measurements of Atmospheric Turbulence Within a Thunderstorm—U. S. Weather Bureau Thunderstorm Project." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1233 (1947).
40. U. S. Air Weather Service, *Further Studies of Thunderstorm Conditions Affecting Flight Operations: Turbulence.* Air Weather Service Tech. Rep. No. 105–39, Washington, D. C., 1949.
41. U. S. Weather Bureau, "Thunderstorm Rainfall." *Hydrometeor. Rep.*, No. 5 (1947).
42. —— "A Report on Thunderstorm Conditions Affecting Flight Operations." *Tech. Pap.*, No. 7, Washington, D. C. (1948).
43. —— *The Thunderstorm*, 287 pp. Washington, D. C., U. S. Govt. Printing Office, 1949.
44. Weickmann, H., "Die Eisphase in der Atmosphäre." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 6 (1949).
45. Wexler, H., and Namias, J., "Mean Monthly Isentropic Charts and Their Relation to Departures of Summer Rainfall." *Trans. Amer. geophys. Un.*, 19: 164–170 (1938).
46. Wichmann, H., "Grundprobleme der Physik des Gewitters." *Phys. Forschung*, P. Jordan, Hsgbr., Nr. 1, 118 SS. (1948).
47. Workman, E. J., Holzer, R. E., and Pelsor, G. T., "The Electrical Structure of Thunderstorms." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 864 (1942).
48. Workman, E. J., and Reynolds, S. E., "Electrical Activity as Related to Thunderstorm Cell Growth." *Bull. Amer. meteor. Soc.*, 30: 142–144 (1949).
49. —— "Electrical Phenomena Occurring during the Freezing of Dilute Aqueous Solutions and Their Possible Relationship to Thunderstorm Electricity." *Phys. Rev.*, (2) 78: 254–259 (1950).

CUMULUS CONVECTION AND ENTRAINMENT

By JAMES M. AUSTIN

Massachusetts Institute of Technology

Introduction

Cumulus clouds are the visual evidence of the overturning of the atmosphere which follows the establishment of an unstable state of equilibrium. The unstable state may be created by such processes as the following:

1. The heating of the lower atmosphere through its contact with a warm surface gives rise to the cumulus clouds which appear on a summer afternoon and to the cumulus clouds which often develop when a cold air mass flows over a warmer surface.

2. The cooling of the middle troposphere may produce an unstable condition in the lower atmosphere. The nighttime cumulus which forms over oceans may be attributed, in part, to this process.

3. The saturation of a layer of convectively unstable air results in the development of convective cells. A good example of this process is the cellular pattern of the precipitation in convectively unstable air of tropical maritime origin.

Following the establishment of the unstable condition the atmosphere overturns and thereby transports heat upward. The manner in which an unstable layer of fluid overturns has been the subject of experimental and theoretical investigations.

Bénard has described the convection cells which appear in a horizontal layer of fluid which is heated from below. The experimental work of Bénard, Idrac, Terada, Mal, Walker and Phillipps, and others is reviewed by Brunt [5]. These experiments involved only very thin layers of fluid. The evidence of the existence of convective cells in the atmosphere, other than cumulus clouds, includes the observational work of Durst [8] and Woodcock and Wyman [16], and the measurements by radar and aircraft as reported by Wexler [15] and Byers and Braham [6].

The theoretical problem of the development of convective cells in an unstable medium of small thickness has been analyzed by Rayleigh and others. The theory is reviewed by Stommel [14]. However, a theoretical analysis of the development of cells of the dimensions of cumulus clouds is lacking. Some of the disturbing problems are the applicability of the linearized equations, the variability of the coefficient of eddy diffusion, the compressibility of the atmosphere, and the horizontal wind shear in the vertical plane. Also, the liberation of the latent heat of condensation in the cloud provides a heat source other than the ground. It would appear that the problem of cumulus convection should be analyzed from the standpoint of the development of cellular circulations. In the absence of a complete solution of this problem it has been necessary to make various assumptions as to the important aspects of the process.

Parcel and Slice Methods

Convective currents develop in an atmospheric layer which is in an unstable state of equilibrium because a perturbation then gives rise to an acceleration of air particles away from their original positions. The simplest approach to the convection problem is to consider that a parcel of air can be displaced adiabatically without any density change taking place in the environment. The stability of an atmospheric layer is then tested by comparing the density of a parcel which is displaced from its position of equilibrium with the density of the environment. This technique, which is referred to as the *parcel method*, is reviewed in many meteorological texts. The parcel method states that convective cells appear in nonsaturated air whenever the actual lapse rate of temperature is in excess of the dry-adiabatic rate, and in saturated air whenever the actual lapse rate of temperature is greater than the moist-adiabatic rate.

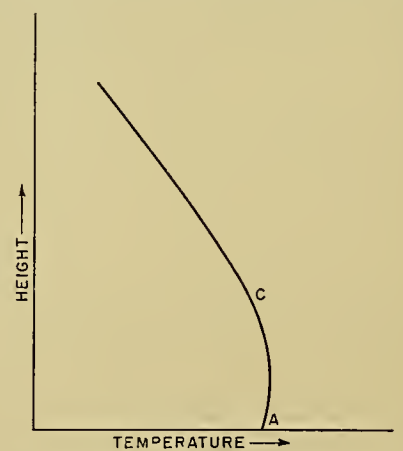


FIG. 1.—An idealized atmospheric sounding.

Cumulus clouds represent the condition where the rising air is saturated while the environment is nonsaturated. The parcel-method analysis then gives cumulus growth provided that (1) the surface layer AC in Fig. 1 is heated so that the convective currents in this layer give condensation at C (that is, C is the convective condensation level), and (2) the lapse rate of temperature above C is greater than the moist-adiabatic lapse rate.

One of the principal defects of the parcel method is the assumption of an undisturbed environment. The *slice method* eliminates the assumption of an undisturbed environment by assuming that there are compensating downward currents which are accompanied by adiabatic heating. This method gives the same stability criteria as the parcel method for the cases of a nonsaturated layer and an entirely saturated layer. However, for the

case of saturated rising air in a nonsaturated environment the two methods differ. The details of the slice method will not be treated here (see the discussion in Petterssen [11]). The principal difficulty in applying the slice method to the cumulus problem resides in the fact that the slice method considers slices of infinitesimal thickness whereas cumulus clouds are of finite extent. When it is assumed that the rising air cools moist-adiabatically there is generally some temperature difference between the cloud top and its surroundings and, therefore, the slice-method analysis cannot be applied beyond the initial stage at the cloud base.

One of the major defects of these two methods is the assumption that the vertically moving currents are isolated from their surroundings and that only adiabatic temperature changes occur within the cloud and its environment. The questionable nature of this assumption is well illustrated by the temperature observations within cumulus clouds. For example, the observations analyzed by Stommel [13] and Byers and Braham [6] failed to reveal moist-adiabatic lapse rates of temperature within the cloud. Instead it was observed that the cloud temperature was only slightly different from that of the environment. Also measurements indicate that the liquid-water contents are only a fraction of those which would be expected from the assumption of the moist-adiabatic ascent of air. The empirical data then demonstrate that there must be continual mixing between the rising current and the environment. This mixing of environmental air into the rising current of cloud air is referred to as *entrainment.*

Concept of Entrainment and Mixing

The concept of entrainment as the turbulent mixing of fluid from an environment into a jet stream has been analyzed by hydrodynamicists. Its application to the cumulus problem is discussed by Schmidt [12] and Stommel [13]. The difficulty of applying hydrodynamical jet-stream theories to the cumulus problem resides in the concept of the jet. In the atmosphere there is no mechanism to eject air into its surroundings and therefore the moving air column, or the jet, must originate from the vertical accelerations which are produced within the air itself. Such accelerations can be deduced from an assumption that the rising cloud air cools adiabatically. However, such an assumption is not very realistic. If air commences to rise so that its temperature decreases at the moist-adiabatic rate in an environment where the lapse rate of temperature exceeds the moist-adiabatic lapse rate, the virtual temperature T' of the rising air is higher than the virtual temperature T of the environment. As a consequence of the assumption that the pressure on the rising particle is equal to the pressure of the environment, the rising particle is accelerated upward and the acceleration is given by

$$\frac{dv_z}{dt} = \frac{T' - T}{T} g, \tag{1}$$

where g is the acceleration of gravity. Since the acceleration is positive the upward velocity of the rising air increases with height. At the same time the density of the rising air decreases with height and is determined by the variation of the virtual temperature T' and by the pressure of the environment. For any appreciable value of $T' - T$ the density decrease is less than the velocity increase. If the outer boundaries of the column of rising air form a cylinder, it follows that more mass is leaving the top portion of any section of the cylinder than is entering the base of the section. This state is impossible because it violates the condition of continuity. Continuity of mass requires that air be brought in through the sides of the cylinder, but this is not possible in view of the assumption that the pressure inside the cylinder is equal to the pressure outside. Alternatively, continuity may be satisfied by allowing the horizontal dimensions of the rising stream to decrease with height. This case does not appear physically real as a model for a cumulus cloud, since it leads to a very narrow current moving with a large velocity. Hence, when it is considered that the jet develops through the accelerations produced by the moist-adiabatic ascent of cloud air, it must be recognized that there has to be a horizontal inflow of air into the rising saturated current in order to maintain continuity of mass within the ascending stream.

Up to the present time a common approach to the entrainment problem has been to consider the extent to which the mixing changes the lapse rate of temperature within the cloud from the moist-adiabatic rate. With this approach three different aspects of mixing between the cloud and its environment should be considered, namely,

1. The horizontal inflow of environmental air into a cloud column which is assumed to grow upward in a number of adiabatic steps.
2. The hydrodynamical entrainment of environmental air.
3. The eddy diffusion of cloud particles into the environment. This aspect of mixing is present even when there is no general rising motion within the cloud.

The analysis has been restricted so far to a vertically moving stream in a stationary environment. In an actual weather situation the horizontal wind speed varies with elevation and it is now necessary to consider the effect of the vertical wind shear on the entrainment. This aspect of cumulus convection has been discussed by Malkus [9]. It is shown that the horizontal translation of an ascending cloud column relative to its surroundings affects the growth and motion of the cloud. For the usual case of an increase in the horizontal wind with elevation, the entrainment is greatest on the upwind side of the cloud while on the downwind side the liquid cloud is detrained from the field of rising motion. Thus the entrainment of outside air is asymmetrical. A further consideration is the horizontal variation through a cloud section of the mixing of environmental air with cloud air even in the absence of a vertical wind shear. This problem has been analyzed for the hydrodynamical jet stream, but it has not been solved for the other aspects of mixing which have been discussed here.

Finally, it is recognized that, as a consequence of the

downward motion, the environment is continually changing while the cloud grows upward and, therefore, it might be expected that the entrainment and mixing at a particular cloud level are changing with time. When this factor is added to the other aspects of mixing it is apparent that entrainment is a complex problem. At this point the question may be raised as to whether the problem of cumulus convection could be more readily solved by an approach which recognized the cloud and its surroundings as a cell rather than one which attempted to treat the cloud and environment separately.

Effect of Mixing on the Cloud Properties

Even though it is not yet possible to determine the precise details of the mixing of environmental air with the ascending cumulus cloud, an insight into the growth of cumulus cells may be obtained by a theoretical analysis of the effect of the mixing on such aspects of the cloud as its temperature and liquid-water content. A graphical technique for the determination of the lapse rate of temperature within the cloud is described below and will be followed by a thermodynamic analysis of mixing.

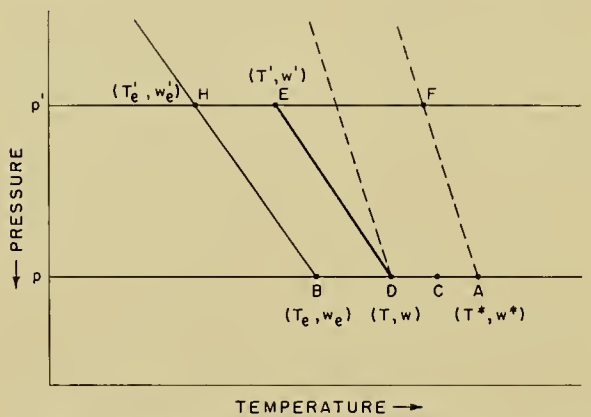


FIG. 2.—Graphical computation of the effect of mixing on the lapse rate of temperature. AF is a moist-adiabatic, BH is the environmental lapse rate, and DE is the lapse rate which arises from mixing.

Graphical Procedure. In the graphical procedure the following assumptions are made:

1. The clouds are formed as the result of heating at the ground.
2. Vertically moving columns of air are subjected to adiabatic temperature changes.
3. The rising air mixes with the air of the environment. The physical properties of the environmental air can be obtained from the radiosonde data.
4. When mixing occurs, the nonsaturated air of the environment becomes saturated at constant pressure by the evaporation of some of the liquid water of the cloud. This is saturation by the wet-bulb process.
5. The lapse rate of temperature within the cloud is obtained from the computed values of the temperature of the cloud top as the cloud grows upward from the condensation level. This assumption ignores the fact that the condition of the cloud is probably changing continually as a result of the influence of the descending currents upon the temperature and humidity of the environment.

The graphical procedure for the determination of the lapse rate of temperature within the cloud is illustrated in Fig. 2. Suppose that at some point in its ascent along the moist-adiabatic a cloud parcel (A in Fig. 2) is permitted to mix with its environment. If the mixing takes place at constant pressure and if some value K is assigned for the proportion of the environment entrained by the cloud, the temperature T_u and the mixing ratio w_u of the cloud air are given by

$$T_u = (T^* + KT_e)/(1 + K),$$
$$w_u = (w^* + Kw_e)/(1 + K), \qquad (2)$$

where (T^*, w^*) and (T_e, w_e) are, respectively, the cloud and environmental variables before the mixing occurs. As a result of the mixing, the cloud point on the thermodynamic diagram has moved from $A(T^*, w^*)$

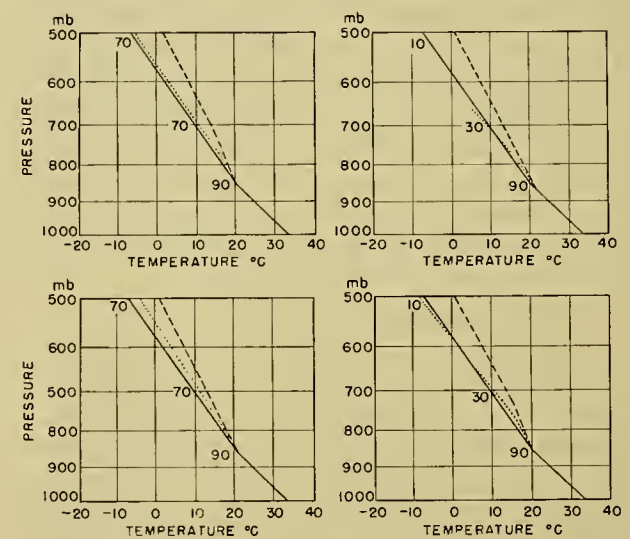


FIG. 3.—Cloud lapse rates of temperature which could arise through mixing. The solid lines, dashed lines, and dotted lines are respectively the environmental, moist-adiabatic, and cloud lapse rates of temperature. The numbers represent the humidity of the environment in per cent. In the two upper diagrams three parts of cloud air mix with one part of environmental air at every 50-mb step; in the two lower diagrams five parts of cloud air mix with one part of environmental air at every 50-mb step.

to $C(T_u, w_u)$, where the cloud parcel is no longer saturated. Let the parcel become saturated at constant pressure by the evaporation of the liquid water of the cloud. The final state of the cloud parcel is then given by the wet-bulb temperature T and the saturation mixing ratio at this temperature w.

Permit the cloud to continue its ascent but now along the moist-adiabatic through D, and at a pressure p' repeat the process of mixing and evaporation. The condition of the air in the cloud at a pressure of p' is then given by the point $E(T', w')$ on the thermodynamic diagram. The lapse rate of temperature within the cloud, between p and p', is given by the line DE.

Under the assumption of the parcel method the lapse rate of temperature within the cloud would be given by AF.

This graphical procedure has been followed for two different moisture distributions and mixing regimes (Fig. 3). The computed values of the lapse rate of temperature within the cloud provide an insight into the effect of mixing upon the growth of the cloud.

Thermodynamic Analysis. As a saturated cloud mass m rises a vertical distance dz, through a pressure change dp, let it mix with a mass dm of its environment. The ascent and mixing may be considered to consist of the following processes: an adiabatic temperature decrease of the cloud mass m; an isobaric cooling of the mass m and isobaric heating of dm; and evaporation of a portion of the liquid water of m in order to saturate dm so that the final mixture is saturated and has a uniform horizontal distribution of temperature, water vapor, and liquid water. The lapse rate of temperature in the cloud, γ_c, may be derived by considering the heat changes occurring in the rising mass. Austin and Fleisher [2] have shown that

$$m(c_p\,dT - RT\,d\ln p) = -c_p(T - T_e)dm - L(w - w_e)\,dm - mL\,dw, \tag{3}$$

and that

$$\gamma_c = \frac{\frac{g}{c_p}\left[1 + \frac{wL}{RT}\right]}{1 + 0.621\,wL^2/c_pRT^2} + \frac{\frac{1}{m}\frac{dm}{dz}\left[(T - T_e) + \frac{L}{c_p}(w - w_e)\right]}{1 + 0.621\,wL^2/c_pRT^2}, \tag{4}$$

where c_p is the specific heat at constant pressure, L the latent heat of evaporation of water, w the saturation mixing ratio at the temperature T, w_e the actual mixing ratio of the environment, and R the constant of the gas equation for dry air. If there is no mixing, $dm/dz = 0$ and the lapse rate of temperature in the cumulus cloud is the same as the moist-adiabatic lapse rate of temperature γ_m. Since $T_e < T$ and $w_e < w$, $\gamma_c > \gamma_m$, that is, the lapse rate of temperature within a cloud which is mixing with its environment is greater than the moist-adiabatic lapse rate.

The cumulus cloud ceases to grow either when the liquid-water content is zero or when the cloud becomes sufficiently colder than the environment so that the vertical velocity is zero. The variation of the liquid-water content w_l can be ascertained by equating the total water content before and after mixing:

$$dw_l = -dw + \frac{dm}{m}(w_e - w - w_l). \tag{5}$$

With the substitution of (5) in (3) an expression is obtained for the variation of the total mass of liquid water,

$$\frac{d}{dz}(mw_l) = \frac{mc_p}{L}\left[(\gamma_d - \gamma_c) + \frac{1}{m}\frac{dm}{dz}(T - T_e)\right], \tag{6}$$

where $\gamma_d = g/c_p$, the dry-adiabatic lapse rate of temperature. Consider the usual situation where $\gamma_d \geqslant \gamma_c$. Since $dm/dz \geqslant 0$, then $d(mw_l)/dz \geqslant 0$ as long as $T_e \leqslant T$, and the cloud does not dry through mixing. Therefore, whenever the lapse rate of temperature of the environment is less than the dry-adiabatic lapse rate, the cloud temperature becomes less than the environmental temperature before the liquid-water content reaches zero. It can be concluded that the factor which controls the termination of the cloud growth is the temperature of the cloud.

Cloud Properties. Because there is little information on the mixing process, equations (3)–(6) are of limited practical use. However it is possible to gain an insight into some of the cloud characteristics by considering the orders of magnitude of the various terms of these equations. The observations analyzed by Barrett and Riehl [3] and Stommel [13] show that the temperature difference between the inside and outside of a cumulus cloud is very small. Figure 4 is reproduced from the

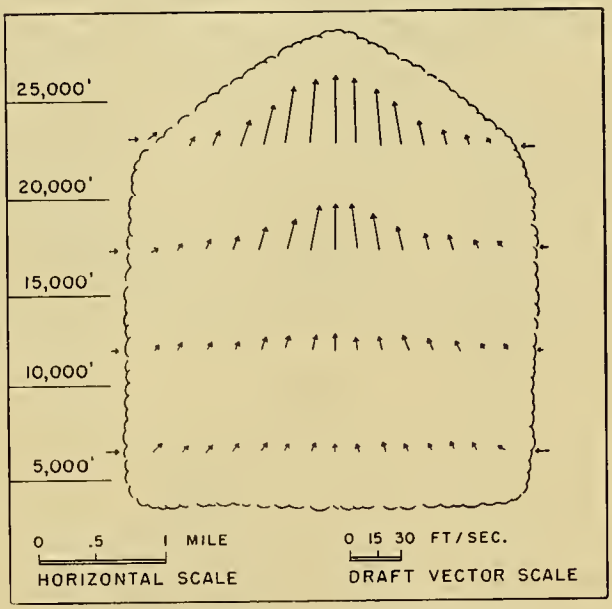


FIG. 4.—Circulation within a typical cell in cumulus stage. Inflow and vertical motion beneath the cloud are light and are not shown. (*After Byers and Braham* [6].)

work of Byers and Braham [6] who arrived at this typical picture from the analysis of extensive observational data. If the variation with height of the average vertical velocity is considered and if allowance is made for the force required to set the entrained air in motion, it follows that a virtual temperature difference of less than 1C is ample to explain the development of the cloud. Even when friction is considered it is probable that an average virtual temperature difference of 1C or less is sufficient to account for the upward force which causes the vertical development of most cumulus clouds. From this empirical information it may be concluded that the lapse rate of temperature within the cloud differs only slightly from the lapse rate of temperature within the environment.

With $T - T_e < 1$C, it follows that $\frac{L}{c_p}(w - w_e)$ is large as compared with $T - T_e$ except for a rare situation of a practically saturated environment. Hence the assumption of $T = T_e$ appears a reasonable one for the discussion of such factors as liquid-water content and the variation of cloud mass with height. With $T = T_e$ and $\gamma_e = \gamma_e$, equation (3) becomes

$$\gamma_e - \gamma_d = \frac{L}{c_p m}\frac{dm}{dz}(w - w_e) + \frac{L}{c_p}\frac{dw}{dz}. \qquad (7)$$

Let $w_e/w = H$, the relative humidity of the environment, and $\gamma_d - \gamma_e = \Gamma$. If Γ and H are considered as constants between elevations Z_0 and Z, the integration of equation (7) yields

$$\frac{m}{m_0} = \left(\frac{w_0}{w}\right)^{1/(1-H)} \exp\left[-\frac{c_p \Gamma}{L(1-H)}\int_{z_0}^{z}\frac{dz}{w}\right]. \qquad (8)$$

The variation of the cloud mass with different lapse rates and relative humidities is illustrated in Fig. 5.

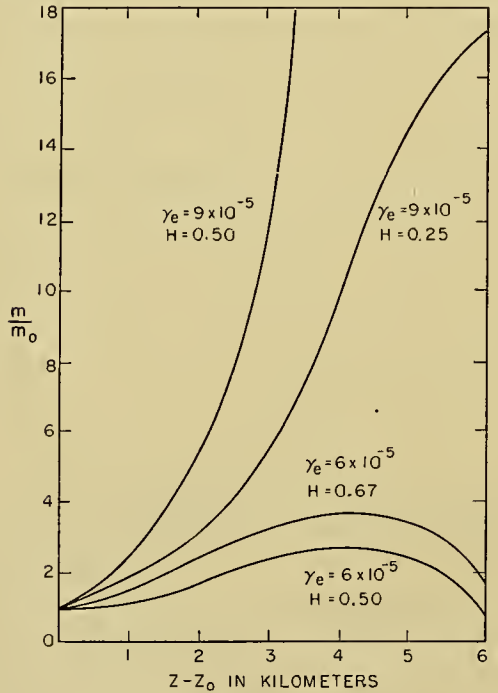


Fig. 5.—Dilution of cloud mass with elevation. m/m_0 is the ratio of the actual cloud mass m, at the level Z, to the fractional mass m_0 which originated at the cloud base Z_0, γ_e is the lapse rate of temperature, and H is the relative humidity. The cloud base is at 900 mb and 20C.

The mass increases more rapidly when the environment has a steep lapse rate of temperature and a high relative humidity. All these curves pass through a maximum, the location of which depends on Γ. These maxima occur at that height where the moist-adiabatic lapse rate becomes equal to γ_e, as can be seen from equation (4) if dm/dz is set equal to zero.

With the same assumptions equation (6) becomes

$$d(mw_l)/dz = mc_p\,\Gamma/L, \qquad (9)$$

thus,

$$w_l = \frac{c_p \Gamma}{Lm}\int_{z_0}^{z} m dz. \qquad (10)$$

The variation of w_l is illustrated in Fig. 6. The liquid-water content and cloud mass vary with Γ and H in a manner summarized in Table I. It is suggested that these general conclusions on the variation of m and w_l also hold for the case where the cloud air is slightly warmer than the environment; that is, the situation where there is a force to accelerate the cloud air upward.

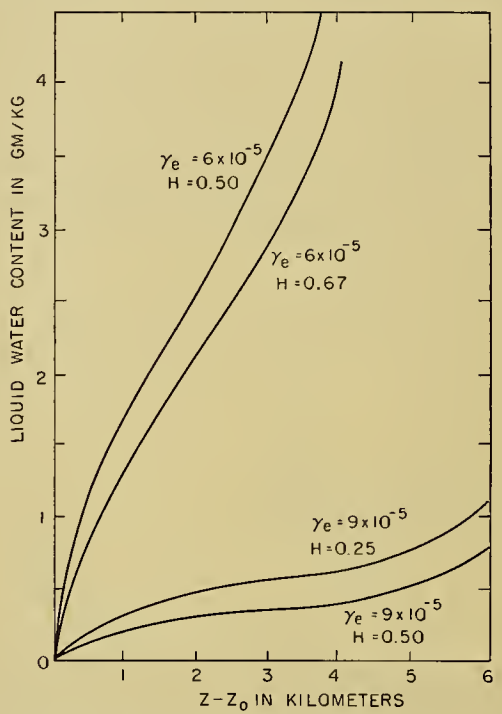


Fig. 6.—Variation of the liquid-water content with elevation for the same conditions as in Fig. 5.

For the application of the above results to the analysis of an actual cloud, consideration must be given to the horizontal variation of the amount of environment mixed with the rising column, the change with time of the lapse rate and the relative humidity of the environment, the change with time of the degree of mixing, and the significance of the internal circulation cells within the cloud. Also it has been assumed that the entire liquid-water content is carried upward, therefore the analysis cannot be applied to cases where precipitation occurs.

Table I. Dependence of Cloud Mass and Liquid-Water Content upon Lapse Rate of Temperature ($\Gamma = \gamma_d - \gamma_e$) and Relative Humidity (H)

	Γ increasing	H increasing
m	decreases	increases
w_l	increases	decreases

The graphs in Fig. 5 may suggest a peculiar cloud

shape since they indicate that the cloud mass continues to increase with height. These graphs should be interpreted as representing, at any elevation, the fraction of the cloud air which has originated at a lower elevation. For example, in the case $\gamma_e = 9 \times 10^{-5}$C cm^{-1} and $H = 0.50$ in Fig. 5, the air at 2 km above the cloud base contains one part of air which originated at the base and five parts of air which have come from the environment. However, this may not necessarily mean that the mass of the cloud at 2 km is six times as great as the mass at the base. In fact, it is probable that not all of the air which reaches a particular level continues to ascend.

Data on Convective Showers

In view of the eddy diffusion to which a cloud is subject, it seems probable that the ideal conditions for the persistence of a cumulus cloud are a high liquid-water content and large horizontal extent. For the development of the vertical accelerations to force the cloud upward it seems desirable that there be a steep lapse rate of temperature. The analysis, as summarized in Table I, indicates that it is not possible to satisfy all three conditions; for example, a condition which gives rise to a high liquid-water content does not appear to favor the development of vertical accelerations. Therefore, it should not be expected that a steep lapse rate of temperature in the environment is necessarily the most favorable condition for cumulus growth. Consequently, attempts have been made to obtain some empirical data on cumulus clouds which would give an insight into the mixing process.

Austin [1] studied the occurrence of convective showers in the vicinity of three radiosonde stations in the eastern United States. It was found that the relative humidity of the environment was as important a consideration for the development of showers as the vertical stability of the air. Chalker [7] confirmed this conclusion and showed that showers occur most frequently in regions where the lapse rate is about 6C km^{-1}, that is, only slightly steeper than the moist-adiabatic lapse rate. These studies were restricted to the occurrence or nonoccurrence of shower-type clouds. It would be desirable to have extensive empirical information on cumulus humilis and cumulus congestus clouds in order to determine more precisely the effect of the humidity of the environment on the cloud growth. However, at present, such information is not readily obtainable.

Prediction of Cumulus Development

Since the unstable state may be created in various ways, as pointed out in the introduction, the problem of forecasting cumulus development can conveniently be divided into two sections.

1. *Convectively Unstable Air.* Petterssen [11] has provided an analysis of cumulus development in convectively unstable air. In brief, if a layer of air is characterized by a decrease of the wet-bulb potential temperature with height and if this layer becomes saturated, an unstable state will exist. As a consequence of the instability, cumulus cells develop in the saturated layer. The prediction of this phenomenon requires first, a determination of the convective state of the air *before* it is subjected to some rising motion which produces the saturation; and second, an estimate of whether rising motion will occur. The convective state is determined from the distribution of the wet-bulb potential temperature. If a deep layer of the air is convectively unstable the forecaster must estimate the likelihood of the air being lifted to saturation. Rising motion may be expected in the vicinity of fronts, along mountain ranges, and in general regions of horizontal convergence near the earth's surface. The best example of the latter process is the horizontal convergence in the vicinity of isallobaric minima.

The principal defect of the forecast procedure is the vague qualitative approach. More information is needed on the manner in which a deep layer of air overturns and on the distribution of rising velocities. Radar observations [10] of precipitation areas have shown that the vertical extent of rain cells, and presumably that of convection cells, varies directly with the degree of convective instability. Hence it should be expected that highly convectively unstable air will give rise to tall cumulus cells of thunderstorm dimensions.

2. *Surface Heating.* Surface heating may arise from the transport of cool air over a warmer surface or from direct radiational heating. A method for the prediction of cumulus development may be applied to either type of heating.

The parcel-method analysis of the equilibrium state leads to a forecast method which involves the so-called positive and negative areas, as illustrated in Fig. 7.

Fig. 7.—The positive and negative area analysis for cumulus convection. The dashed line is a moist adiabatic and the dotted line is a dry adiabatic.

$ABPD$ is the path on an adiabatic chart followed by an air parcel which is lifted through an atmosphere $ALPMD$. The forecast method states that cumulus clouds will develop to shower dimensions provided that

the positive area is greater than the negative area, that the positive area extends beyond the 0C isotherm, and that the surface heating is sufficient to initiate the convection. This method is open to serious criticism since the cloud air does not follow the path *ABPD* and hence the positive and negative areas have no significance as regards the cumulus development. Furthermore, the negative area is zero when the cloud commences to develop.

A more appropriate application of the parcel-method technique is to determine the convective condensation level, *C* in Fig. 8, and to estimate the heating necessary to initiate convection, $T_H - T_A$. If this heating is expected, then cumulus clouds should develop to the level *D*. This method may be criticized on the same theoretical grounds that make the parcel-method analysis untenable. Furthermore, observations show that this technique overestimates cumulus development. Also this method offers no satisfactory explanation for the failure of cumulus clouds to develop whenever there is dry air in the middle troposphere.

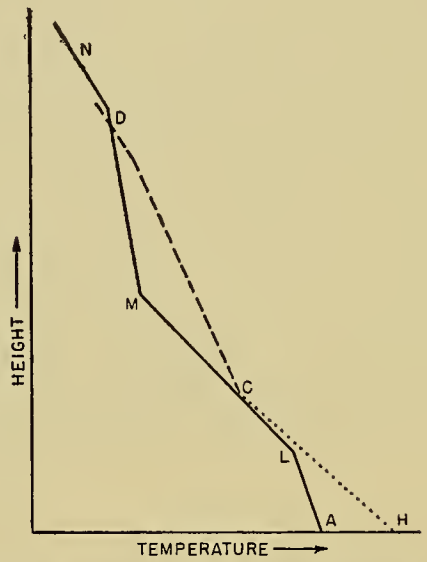


FIG. 8.—The parcel-method analysis of cumulus convection. The dashed line is a moist adiabatic and the dotted line is a dry adiabatic.

Beers [4] has proposed a forecast technique which is based upon the slice method. The method treats layers of finite thickness and, therefore, is an attempt to apply the slice-method technique to a deep layer of air. This method assumes that the cloud air rises adiabatically and hence it is open to question. However it may be possible to modify the method so as to allow for the mixing. The desirable feature of Beers' technique is the treatment of the cloud growth as the development of a circulation cell.

The concept of the entrainment of outside air suggests an alternate procedure for the prediction of cumulus convection through heating which may be summarized as follows:

1. The surface dew-point temperature to be expected about the time of cloud formation is estimated. This determines the convective condensation level. The heating necessary to initiate cumulus growth can be determined by drawing a dry-adiabatic from the convective condensation level to the ground. If it is expected that this heating will occur during the forecast period, some cumulus development may be expected. Since these temperature estimates are based on widely scattered observations, the forecaster should tend to overestimate the dew-point and maximum temperatures. This overestimation is justified as there are probably locations nearby which are more favorable for cumulus development than directly over the isolated reporting stations.

2. If the lapse rate of temperature above the convective condensation level is less than the moist-adiabatic lapse rate, no cumulus development is to be expected. If the lapse rate of temperature is in excess of the moist-adiabatic rate, the degree of cumulus development should be based upon the relative humidity distribution above the convective condensation level. The forecaster may be aided by the statistics given by Austin [1] and Chalker [7].

3. A layer with low relative humidity, or a layer with a lapse rate less than the moist-adiabatic rate above the convective condensation level, definitely disfavors cumulus development.

4. These estimations may be made from radiosonde observations taken prior to the forecast period, but here consideration must be given to the probable change in the lapse rate of temperature and relative humidity during the forecast interval. It is suggested that this estimate be based on the trend as illustrated by a stability chart which depicts the lapse rate of temperature from 850 mb to 500 mb and the relative humidity within the same layer (see Chalker [7]).

Suggestions for Research

The introduction of the concept of entrainment and mixing may be considered an advance in the analysis of cumulus convection. It recognizes that a cumulus cloud does not grow upward as an isolated column of saturated air which cools at the moist-adiabatic rate. However, many theoretical problems remain to be solved. No satisfactory procedure has been offered whereby the degree of entrainment and mixing may be estimated and recent analyses of entrainment have failed to take into consideration the disturbed state of the environment. As a consequence of this descending environment the cloud is continually in a state of change which presents serious theoretical problems if it is desired to know the physical properties of the cloud at a given time. More theoretical research is clearly indicated.

One feature of cumulus growth which requires more consideration is the significance of the surface heating. In the past the tendency has been to concentrate attention on the cloud after it starts to develop and to consider that the surface heating is only important insofar as it initiates the cloud development. However, it appears logical to expect that the degree of the surface heating, beyond that necessary to start cloud growth,

should affect the development of the cloud. The ground may be as significant a heat source as the latent heat released in the cloud. When this surface heating is considered it again appears that an approach like that applied to the Bénard cells may be more fruitful than recent attempts to handle entrainment and mixing. The author recognizes that there are serious theoretical problems in the cell approach. In many respects the question of convection is perhaps as complicated as the cyclone problem.

In conclusion, it appears likely that no major progress will be made with the forecast problem of cumulus convection until more knowledge is gained of the mechanics of cloud growth.

REFERENCES

1. Austin, J. M., "A Note on Cumulus Growth in a Non-saturated Environment." *J. Meteor.*, 5: 103–107 (1948).
2. —— and Fleisher, A., "A Thermodynamic Analysis of Cumulus Convection." *J. Meteor.*, 5: 240–243 (1948).
3. Barrett, E. W., and Riehl, H., "Experimental Verification of Entrainment of Air into Cumulus." *J. Meteor.*, 5: 304–307 (1948).
4. Beers, N. R., "Atmospheric Stability and Instability" in *Handbook of Meteorology*, F. A. Berry, Jr., F. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See pp. 693–711)
5. Brunt, D., *Physical and Dynamical Meteorology*, 2nd ed. Cambridge, University Press, 1939. (See pp. 219–223)
6. Byers, H. R., and Braham, R. R., Jr., "Thunderstorm Structure and Circulation." *J. Meteor.*, 5: 71–86 (1948).
7. Chalker, W. R., "Vertical Stability in Regions of Air Mass Showers." *Bull. Amer. meteor. Soc.*, 30: 145–147 (1949).
8. Durst, C. S., "Notes on the Variations in the Structure of Wind over Different Surfaces." *Quart. J. R. meteor. Soc.*, 59: 361–371 (1933).
9. Malkus, J. S., "Effects of Wind Shear on Some Aspects of Convection." *Trans. Amer. geophys. Un.*, 30: 19–25 (1949).
10. Mather, J. R., "An Investigation of the Dimensions of Precipitation Echoes by Radar." *Bull. Amer. meteor. Soc.*, 30: 271–277 (1949).
11. Petterssen, S., *Weather Analysis and Forecasting*. New York, McGraw, 1940.
12. Schmidt, F. H., "Some Speculations on the Resistance to the Motion of Cumuliform Clouds." *Mededl. ned. meteor. Inst.*, (B) Deel 1, Nr. 8 (1947).
13. Stommel, H., "Entrainment of Air into a Cumulus Cloud." *J. Meteor.*, 4: 91–94 (1947).
14. —— "A Summary of the Theory of Convection Cells." *Ann. N. Y. Acad. Sci.*, 48: 715–726 (1947).
15. Wexler, R., "Cellular Structure of Intermittent Rain." *Ann. N. Y. Acad. Sci.*, 48: 777–782 (1947).
16. Woodcock, A. H., and Wyman, J., "Convective Motion in Air over the Sea." *Ann. N. Y. Acad. Sci.*, 48: 749–776 (1947).

OBSERVATIONS AND ANALYSIS

World Weather Network *by Athelstan F. Spilhaus* 705

Models and Techniques of Synoptic Representation *by John C. Bellamy* 711

Meteorological Analysis in the Middle Latitudes *by V. J. Oliver and M. B. Oliver* 715

WORLD WEATHER NETWORK

By ATHELSTAN F. SPILHAUS

University of Minnesota

Introduction

To solve the problem of covering the earth with a meteorological network, international cooperation must be presupposed even though the present time seems inauspicious. However, meteorology has an enviable record of international cooperation which has withstood two major wars and has been maintained, albeit in a somewhat restricted manner, in times of severe international stress. Cooperation started as a means of exchanging observations between nations. The actual planning of networks has hitherto taken place on the basis of individual national needs. Consequently, an extremely uneven coverage is afforded the meteorologist who desires to study world-scale problems rather than regional synoptic meteorology. Air navigation, which now encompasses the earth, has given a new impetus to the international organization of weather networks. However, even here the planning is directed to the practical needs of air operations and seldom results in a network satisfactory for the study of the general circulation.

That the need for such studies is well recognized is indicated, for example, by the Southern Hemisphere map analysis project being conducted at the Massachusetts Institute of Technology by Willett [8]. Evidence of extensive consideration of the problem is indicated by numerous resolutions of the Conference of Directors at the meeting of the International Meteorological Organization in Washington, D. C., in 1947. Resolutions 21, 36, 109, and 210 [6], dealing respectively with meteorological reconnaissance in areas with inadequate coverage by other means, with various ways of amplifying the information from oceans, with the density of land stations, and with the proper utilization of suitable islands, point out the necessary steps in the direction of establishing a suitable world weather network.

Studies such as those being conducted in connection with the Southern Hemisphere project will reveal in detail the deficiencies in the world network. However, it is the purpose of this paper to look at the problem from a comprehensive point of view.

In general, the world weather network is deficient in the two polar areas and in the equatorial zone. Latitude for latitude, it is more lacking in the Southern Hemisphere than in the Northern; for a given latitude, it is always more inadequate in ocean regions than in land areas. Because of the general distribution of land and sea on the earth, the Southern Hemisphere at higher latitudes presents a tremendous and very difficult problem of weather coverage.

The Gaps

To localize the worst spots of meteorological terra incognita, so as to form the basis of the very long range planning which is needed in establishing the world weather network, certain assumptions can be made. Apart from the antarctic continent, it is comparatively simple, from a technical standpoint, to provide the requisite density of land stations, although the present closeness on land—particularly in equatorial areas north of 10°S—is very far from the 100–150-km spacing recommended in Resolution 109 [6]. The greater problem lies with the other two-thirds of the earth's surface which is covered by the oceans. Here the assumption is made that, for convenience and economy, suitably selected islands and reefs should be fully utilized before the further step is taken of arranging observations from the surface of the deep ocean.

The ideal situation, in which all strategically placed islands are suitably used for meteorological stations, is far from a reality. Because stations are established primarily for national regional synoptic needs, a number of islands which are highly important from a world-network point of view are not fully manned and equipped, meteorologically, since they are far removed from the sovereign nation which is, therefore, little concerned with their weather observations. Examples are St. Helena in the Atlantic which has no upper-air station and Clipperton in the east equatorial Pacific. Nevertheless, the meteorological occupation of some very important remote islands has been encouraging in the years since World War II (Marion and Amsterdam Islands being two recent examples).

To illustrate graphically the areas remaining after continents, islands, and reefs are properly turned to account, Fig. 1 shows a chorometric chart of the world ocean with isochors representing lines of equal distance from land of any kind at three hundred nautical mile spacing (five degrees of latitude). The base map used for this figure is that devised by the author in 1942 [11]. The data were revised by F. E. Lukermann, Jr., Geography Department, University of Minnesota, from a study of the islands missing both from the first chorometric chart published in 1898 [2, 7] and also from the later ones for the Atlantic, Indian, and Pacific Oceans [9, 10]. Because it is feasible to build structures for meteorological purposes on reefs which are either awash or covered by water less than five fathoms deep, these have been regarded in the same sense as islands in Fig. 1. Included in this category are reefs such as the Virginia Rocks (three or four fathoms) in the Atlantic Ocean; Saya de Malha (about two fathoms) in the Indian Ocean; and, in the Pacific Ocean, the Maria Theresa Reef (awash), the Ernest Legouvé Reef

(awash), Harans Reef (with a ship aground), and Filippo Reef (about two fathoms). Islands apparently missing from Schott's charts, which have been added, are, in the Atlantic Ocean, Rockall and Annobón; in the Indian Ocean, Scott's Reef, Rowley's Shoals,

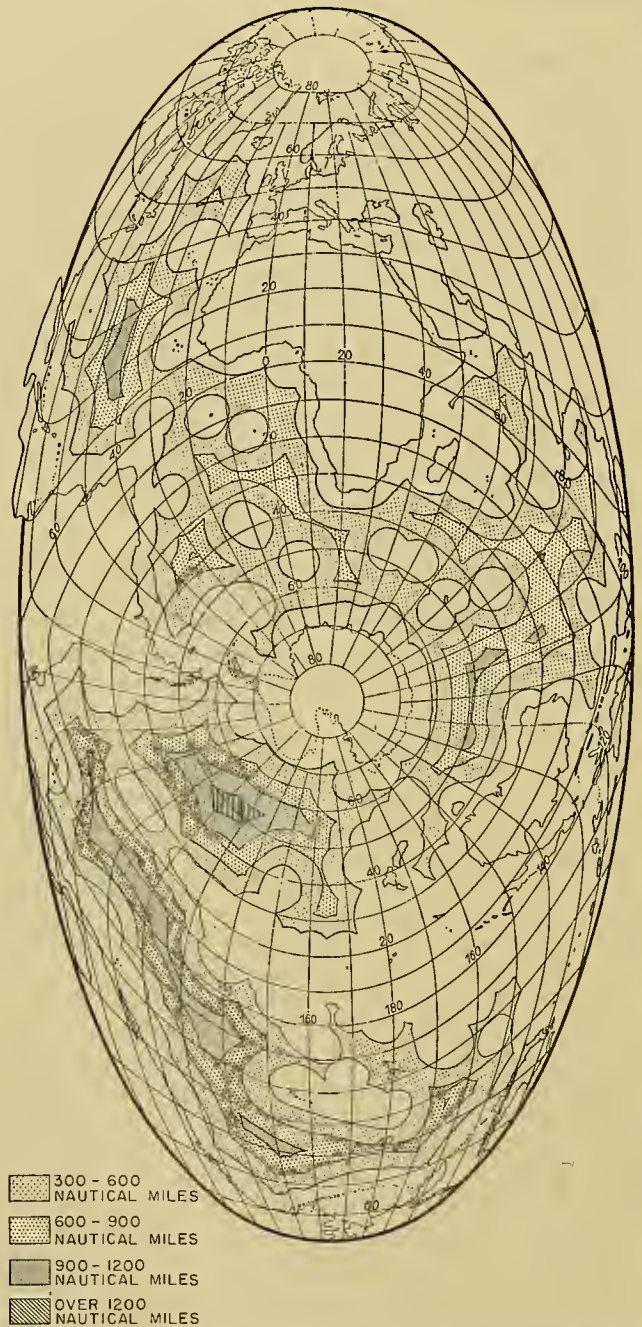


Fig. 1—Lines of equal distance from land at spacings of 300 nautical miles. (Aitoff equal-area projection centered at 70°S, 15°E.)

Bassas da India and Europa Island; in the Pacific Ocean, Palmyra, Kingman's Reef, and Beveridge Reef. Some islands, seemingly included in Schott's maps, but now established as nonexistent, have been eliminated.

Figure 1 shows clearly the regions of the great world oceans which are farthest from land—with two areas in the Pacific more than a thousand nautical miles from land or reef of any kind. It is interesting to note that the North Atlantic, where the most extensive weather-ship program is being operated, has less area distant from land of any kind than the other major oceanic divisions.

Ships' Reports

To cover the areas of the open ocean, there has for many years been a well-directed program enlisting the cooperation of vessels plying the oceans for other purposes. At least for regions traversed by the well-established sea lanes of commerce, ships' reports of this kind are of tremendous value even though their quality is sometimes questionable, and elementary observations, such as those of rainfall, are not taken. An idea of the coverage afforded by this means may be obtained by analysis of the grand total of observations for all months of the period of more than fifty years analyzed in the U. S. Weather Bureau's *Atlas of Climatic Charts of the Oceans* and summarized in Chart 1 of that atlas [14]. This chart shows the observations for all months distributed by five-degree unit areas over the oceans. Figure 2 shows the outlines of the areas in which there were less than five hundred observations in the 50-yr period ending in 1933 (average of less than ten observations per year). The steady increase of shipping of all kinds tends to increase the number of ships' reports. On the other hand, the effect of modern developments (canals, for example) for short-cutting sea routes and the increasing use of aircraft militate against the possibility of there being sufficient ships in these remote areas to provide regular observations. Even the increase in the general amount of shipping in the world occurs along established sea lanes and will not cover the untravelled areas of the ocean. Figure 2, therefore, may be accepted as a fair general indication of the areas which cannot be covered by normal ship-observation techniques. If Figs. 1 and 2 are superimposed, lines can be drawn delimiting parts which are more than 600 mi from land of any kind and which, in the 50-yr period mentioned, had less than ten observations a year. This chart, Fig. 3, gives a good idea of the regions (principally in the Southern Hemisphere) which need to be covered (from a world weather network point of view) with floating stations of some kind, whether they be weather ships or specially designed floating stations, automatic or manned. In these infrequently travelled regions which are far from land, some of the attempts at a solution may be [6, Resolution 36] the use of (1) fishing vessels, (2) aircraft reconnaissance, (3) stationary weather ships, or (4) automatic stations on buoys. To these should be added the less desirable, but also less costly, indirect methods of coverage afforded by sferics, radar networks, and ocean-wave analysis as applied to the location of storms [1]. One example of the gaps that could be filled by the use of fishing vessels is in the Southern Hemisphere where an arrangement for the use of whalers for meteorological purposes might be made. It is to be hoped that cooperation with the International Whaling Association

can be achieved along these lines before the whale becomes extinct. Aircraft reconnaissance has proved exceedingly valuable, particularly for special atmospheric phenomena such as hurricanes and typhoons, and also in its contribution to the knowledge of meteoro-

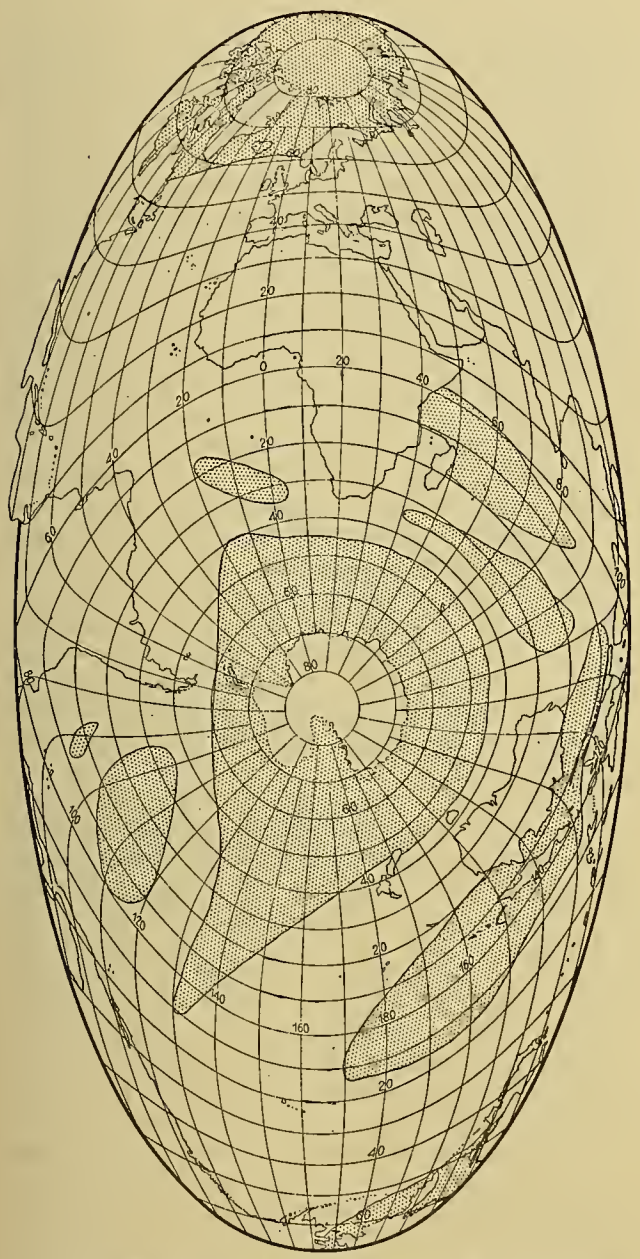


FIG. 2—Ocean areas (shaded) having less than 500 ships' observations in the 50-yr period ending in 1933.

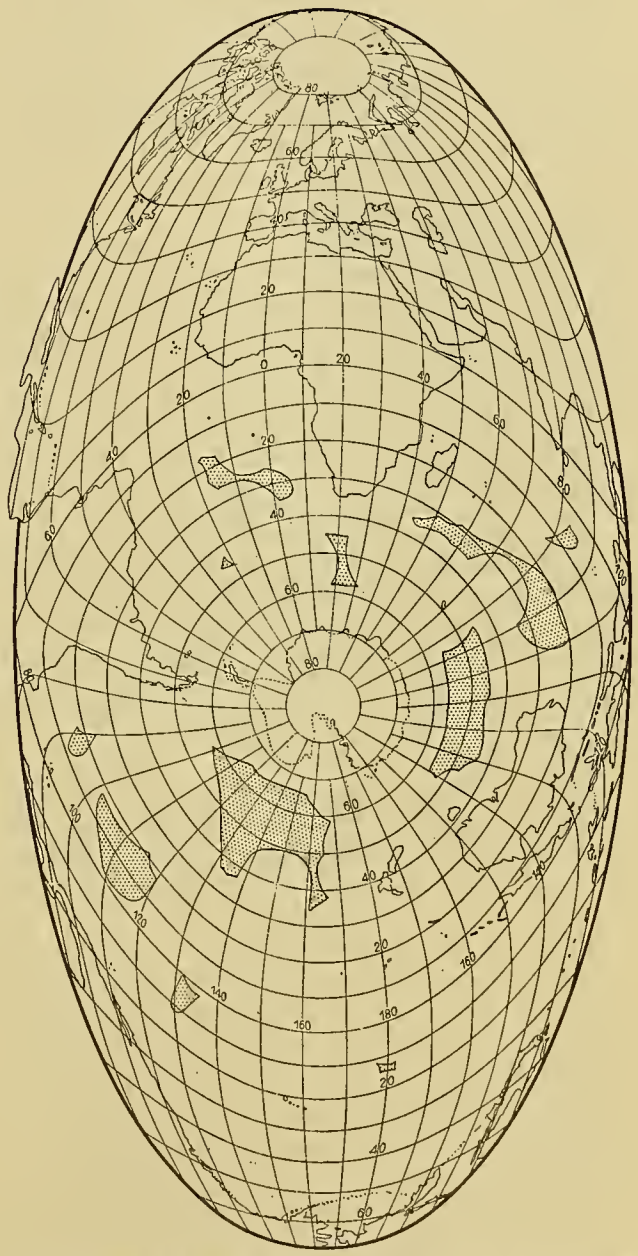


FIG. 3—Ocean areas (shaded) more than 600 nautical miles from land and having less than 500 ships' observations in fifty years.

logical conditions in areas which cannot yet be covered by ordinary observational means. An example of this is the regular weather flight from Alaska to the North Pole. It is important, however, to remember that to be fully effective aircraft reconnaissance of this latter type needs to be coordinated more closely with surface observations. Therefore, as is well illustrated by the North Pole flights, there is still a need for sea-level observations. The development of techniques of measuring winds and other conditions below the aircraft down to the surface gives promise that such techniques will be in routine use fairly soon. The occupation of surface stations, however, still remains essential. In regions very far from land in the Southern Hemisphere, aircraft reconnaissance would not be an economical way of obtaining the information. Stationary weather ships are a most satisfactory solution, except for the excessive cost of their maintenance if ordinary vessels are used for this purpose. The requirement, therefore, is for a specially designed floating vessel, manned or unmanned, which can be suitably anchored even in great depths of water. Work along these lines has proceeded as far as

the introduction of a bill into the Senate of the United States[1] to authorize the Coast Guard to construct an experimental nonpropelled seadrome ocean station. Although there are many questions concerning economy, logistics, and human factors involved in the floating weather station, it is feasible from an engineering point of view. The foregoing analysis shows that the development of a platform for meteorological purposes of this kind is necessary and must be encouraged. Continued development of fully automatic weather stations for use not only on such floating buoys (which could then be of far smaller size) but also on reefs and uninhabited islands is an important parallel endeavor.

Arctic and Antarctic Stations

As with the great oceans, so with the Arctic and Antarctic—the principal problems of filling out the network are not strictly meteorological ones. The prime difficulties in the Arctic and Antarctic are means of ingress and egress and methods of safe and suitable living in extremes of cold. Even if automatic stations are developed for polar regions (and this in itself presents far greater problems than the corresponding ones for temperate zones), the access problem in connection with the maintenance of the stations remains. In the oceanic Arctic, the problems are somewhat different from those in the continental Antarctic. The fact that the Arctic Ocean is predominantly covered with sea ice also presents problems different from those of temperate and tropical oceanic areas. Surface vessels such as ice breakers can penetrate only a limited distance into the ice fields and only at certain seasons of the year.

The most promising methods of getting to and from stations in the Arctic Ocean are by aircraft or submarine. Whether transportation is by aircraft over the ice or by submarine underneath it, methods must be available to surface on the ice. Studies of arctic sea ice will permit this. The submarine may come up between ice floes in open water or, by the use of some kind of boring device, afford access to its occupants through the ice. In the Antarctic, although aircraft are essential, especially designed surface vehicles are also necessary. Caterpillar tractors, designed to operate at extremely low temperatures, together with wanigans, such as are used in Alaska north of the Brooks Range, are of obvious utility. The various transportation methods for the Arctic and Antarctic (aircraft, submarine, and surface vehicles) have one thing in common and that is that none of them have been developed especially for polar operations. Even aircraft operation under the climatic conditions of arctic and antarctic regions is unsatisfactory. With the present intense interest in polar matters in all fields of science, it seems most opportune to stress the development of special vehicles and methods of operation.

With the solution of the transportation problem, or concurrently with its solution, the next problem is one of establishing and occupying semipermanent observing stations in the Arctic. For such stations, saucer-shaped vessels, which would ride up over the ice in much the same manner as did the *Fram*, have been suggested. A more logical approach, however, is to use the ice itself as the floating platform. There is evidence [3] of the existence of ice floes which are many miles across and of the order of a hundred feet or more thick. Such floes would be able to withstand crushing by the general sea ice in the Arctic, which is only 6–12 ft thick, and would be most suitable bases upon which to establish runways and stations. The use of ice as a floating platform would be a start in the direction of using materials in the Arctic itself for the establishment of stations there. It has been remarked that in the excellent stations which the U. S. Weather Bureau has established in cooperation with the Danish and Canadian governments (two above 80°N in the Ellesmere Islands and four at or above 75°N) the supply problem could perhaps be eased if more attention were paid to the utilization of local resources, such as situating the stations on coal deposits which are not far from the existing locations. Coal is not presently employed as a fuel in these operations. Ice, once its engineering properties are known, might be used for building, insulating, water supply, and fire fighting, as well as for many other purposes [5].

Evidence of an intensified attack on the antarctic occupation lies in the Norwegian-Swedish-British Expedition in progress at the present time and in the plans of the Argentine government to establish two more meteorological stations on the antarctic continent [4].

Interim Measures

It is evident that the solutions proposed above to the ultimate establishment of a world network present immense engineering problems and manifold questions of international cooperation which will clearly take many years to accomplish. The understanding of the mechanics of the general circulation of the earth's atmosphere is the "primary problem of meteorology on which scarcely a beginning has been made." It is on the solution of this "that all basic improvement of weather forecasting is now waiting," as Dr. H. C. Willett[2] has said. It is evident that work along these lines cannot await the completion of the world network as envisaged here. Several interim measures have been suggested and have, to a limited extent, already been implemented. For general circulation studies based on extensive synoptic observations, one of the most important things is the use of synoptic aerological data and, as Willett has pointed out:

> The observational basis of all study of the mechanics of the general circulation as a whole has been restricted to the troposphere of the Northern Hemisphere north of 20°N. This is particularly unfortunate when it is realized that probably this is the least important third of the atmosphere

1. U. S. Senate Bill S. 1009, February 17, 1949 (subsequently withdrawn).

2. Private communication: "The Needs in Synoptic Aerological Observations for the Study of the General Circulation," 1947.

dynamically and thermodynamically. The Southern Hemisphere circumpolar vortex is evidently a more intense dynamic phenomenon than that of the Northern Hemisphere and consequently might be expected to be the more important of the two as the seat of disturbances or changes of the general circulation as a whole.

He emphasizes also:

The tropics must contain the principal heat source of the general circulation and also constitute the zone of interaction of the two great hemispherical cyclonic vortices.

Yet it is in the tropics and in the Southern Hemisphere as a whole that our principal lack of aerological data lies and, pending the establishment of a good over-all world network, the distribution of upper-air stations along two or three selected meridians would be most helpful. Two sections which have been suggested by the author and which are largely in effect are (1) a maritime one, from the North Pole through Alaska, the Pacific Islands and New Zealand, to Antarctica; and (2) a continental section from the North Pole through Canada, the United States, Central and South America, to Antarctica [12].

Another measure which would be of great help to the basic theoretical studies of the general circulation is the provision for the right kind of measurements from ships and island stations for treating the heat budget of the atmosphere-ocean complex as a whole. To put it very simply, this budget necessitates knowing the ingoing and outgoing water at the ocean surface (rainfall and evaporation) and the ingoing and outgoing radiation at the same surface. While evaporation can be estimated from sea and air temperatures combined with wind speeds which are currently part of ship measurements, rainfall for the oceans is derived exclusively from the records of island stations. Consideration should be given to the measurement of rainfall on moving ships at sea. Even though the interpretation of the catch of rainfall from a moving vessel presents certain difficulties, it is probable that such measurements would prove far more satisfactory than the present practice of estimating ocean rainfall from island stations. Small islands in extensive oceanic regions usually exert a profound and a very local orographical influence and, therefore, properly interpreted rainfall data from ships in transit may prove far more representative.

The establishment of radiation-measuring stations on properly dispersed islands in the oceans and on a few selected ships seems also to be a desirable measure [13].

Summary

The principal lines along which further work should proceed in order to fill out the world network may be summarized as follows:

1. The development of automatic stations for land and ocean, tropical and polar use.

2. The filling out of the land station network in tropical areas.

3. The establishment of manned or automatic surface stations on suitable islands and reefs.

4. The establishment of upper-air stations on selected islands and reefs. These would be manned, but, at the same time, it would be desirable to initiate research on indirect methods of atmospheric sounding without flight equipment, which would lend themselves ultimately to automatic operation.

5. The development of buoys, which could be anchored in deep water, to carry manned stations or automatic stations for remote oceanic points.

6. The development of methods of getting in and out of arctic and antarctic areas and of maintaining manned stations and automatic stations in those regions.

These are tremendous engineering problems which the humility of the meteorologist may give him pause to urge or to undertake. It will be recalled, however, that the pioneering work on rockets in this country was undertaken with the meteorological sounding of the upper air as its primary objective. Furthermore, all such developments contribute directly to problems common to other geophysical fields and the support of these fields should be enlisted to justify the effort. The other aspect of the world network problem (that of international cooperation) is no less difficult of solution than the technical developments involved. It is, however, merely an extension and growth of the cooperation which is going forward in the International Meteorological Organization, the International Civil Aviation Organization, and in the North Atlantic (and other) weather ship programs. Only by beginning an attack on this immense problem of the world network will suitable data ultimately become available both for synoptic studies of the general circulation and for the kind of information which is needed in the objective weather-forecasting approach promised by the use of electronic computers.

REFERENCES

1. Deacon, G. E. R., "Waves and Swell." *Quart. J. R. meteor. Soc.*, 75:227–238 (1949).
2. deWindt, J., "Sur les distances moyennes à la côte dans les océans," *Mémoires couronnés et mémoires des savants étrangers*, Tome LVII. L'Académie Royale des Sciences, des Lettres, et des Beaux-Arts de Belgique, 1898.
3. Fletcher, J. O., *Floating Islands in the Arctic.* Paper presented at the Alaskan Science Conference of the National Academy of Sciences-National Research Council, Washington, D. C., November 9–11, 1950.
4. Garcia, R., "Discussion on International Co-operation in Obtaining Radio Soundings in the Antarctic with a View to Representing the Desirability of Such Observations to the I. M. O." *P. V. Météor. Un. géod. géophys. int.*, Oslo, 1948. Uccle (1948). (See p. 113)
5. Haynes, B. C., *The Weather Bureau's Arctic Observation Program outside of Alaska.* Paper presented at the Alaskan Science Conference of the National Academy of Sciences–National Research Council, Washington, D. C., November 9–11, 1950.
6. International Meteorological Organization, *List of Resolutions.* Conference of Directors, Washington, D. C., Sept. 22–Oct. 11, 1947. Secretariat, I. M. O., Lausanne, 1948.

7. Rohrbach, C. E. M., "Über mittlere Grenzabstände." *Petermanns Mitt.*, 36:76–84, 89–93 (1890).
8. Rubin, M. J., and Willett, H. C., *Southern Hemisphere Map Analysis Project.* Rep., Mass. Inst. Tech., Cambridge, Mass., July 1, 1950.
9. Schott, G., *Geographie des Atlantischen Ozeans.* Hamburg, C. Boysen, 1926.
10. —— *Geographie des Indischen und Stillen Ozeans.* Hamburg, C. Boysen, 1935.
11. Spilhaus, A. F., "Maps of the Whole World Ocean." *Geogr. Rev.*, 32:431–435 (1942).
12. "Stations and Networks." Specific Recommendation No. 1, p. 10 in *Recommendations for Research in the Pacific Area.* Seventh Pacific Science Congress, Div. of Meteor., New Zealand, February 22, 1949.
13. —— Specific Recommendations Nos. 2 and 3 in *Recommendations for Research in the Pacific Area.* Seventh Pacific Science Congress, Div. of Meteor., New Zealand, February 22, 1949.
14. U. S. Weather Bureau, *Atlas of Climatic Charts of the Oceans.* U. S. Wea. Bur. Publ. No. 1247, Washington, D. C., 1938.

MODELS AND TECHNIQUES OF SYNOPTIC REPRESENTATION

By JOHN C. BELLAMY

Cook Research Laboratories, Chicago, Illinois

Introduction

An ideal station model can be said to be one in which the synoptic observations are rapidly and efficiently plotted in such a way that (1) the analyst or forecaster can ascertain, with a minimum of thought or effort, the complete three-dimensional distribution of atmospheric conditions, and the time changes of that distribution, and (2) all meteorological observations made in a given region during a given time interval are readily available to the analyst or forecaster.

The latter condition is predicated on the fact that the basic station model is determined largely by the characteristics of the communication system which must serve the varied interests and requirements of all users of the meteorological observations. It seems improbable that there will be developed in the near future sufficiently general and accurate thermodynamic models that the number of observations desired by all users will be reduced.

Analysis of Present Techniques

If we use the characteristics of this ideal station model for purposes of comparison, the following inadequacies of present station models are apparent:

1. The manual transcription of all data from teletype reports to the present station models at all receiving stations is obviously neither rapid nor efficient. This is especially obvious when such charts as thermodynamic diagrams, hodographs, cross sections, and nonstandard constant pressure charts are considered. Such charts are not now used extensively because of the excessive manpower required for their plotting and analysis.

2. It is difficult for the analyst to ascertain the complete three-dimensional distribution of atmospheric conditions. Apparently, this is a result of the use of many different kinds of charts, or station models, on many pieces of paper. This difficulty is most serious when one parameter is considered alone. It is still present, however, even when dynamic or thermodynamic models are used to correlate two or more individual parameters. Following is a discussion of the shortcomings of present techniques with respect to the three major dynamic or thermodynamic models now in common use.

The Hydrostatic Equation. The hydrostatic equation can be considered as a dynamic model for correlating observations of pressure, temperature, and height. The common techniques now in use have the following shortcomings with respect to this model:

1. Only a small portion of the continuous pressure-height relationship in the vertical which can be obtained from radiosonde observations is used, since usually the heights of but a few standard-pressure surfaces are available to the analyst. The discarded data are very desirable for the accurate correlation of wind or cloud observations, which are made with respect to height, and radiosonde or aircraft observations, which are made with respect to pressure. These continuous data are also required for accurate aircraft altimetry.

2. Present hydrostatic computations are sufficiently complex that usually they are carried out quantitatively only at the radiosonde observation stations. This complexity limits the efficient correlation of representations of temperatures and heights to special cases and requires excessive experience and memory on the part of the analyst.

3. Present sea-level pressure values cannot be correlated directly with upper-air conditions since in general the particular values used for the reduction to sea level are unknown to the analyst.

The Gradient-Wind Equation. The gradient (or geostrophic) wind equation, in conjunction with the hydrostatic equation, can be considered as being a dynamic model for correlating the wind, pressure, and temperature fields. Present station models are not completely satisfactory for this dynamic model because of their inadequacies with respect to the hydrostatic equation. For example, at present the gradient-wind model can be applied easily only to a few standard constant-pressure surfaces and can be applied only indirectly to vertical representations. These limitations place serious restrictions on possible operational use of pressure-pattern navigational techniques and precise aircraft altimetry, and seriously hinder the assimilation of the three-dimensional distribution of wind, pressure, and temperature conditions.

Air-Mass and Frontal Analysis. Present station models have been designed largely for convenient air-mass and frontal analysis, and are quite adequate for that purpose. However, the efficiency of ascertaining the three-dimensional configuration and detailed characteristics of the air masses and fronts can probably be increased with improved station models.

Only a small fraction of meteorological observations are made available in convenient form for the analysts. Of the upper-air wind observations only the winds at a few predetermined levels and perhaps a few hodographs are plotted. Of the radiosonde observations usually only the observations at a few predetermined pressures and a few complete thermodynamic diagrams are plotted. Of the surface observations usually only the values at 6-hr intervals are plotted. The teletype reports of hourly surface observations are sometimes available, but their form is hardly conducive to efficient assimilation by the analyst. The observations of pressure, temperature, humidity, wind, ceiling, etc., which are

measured continuously in time at the surface, are not even transmitted over the present communication system.

Improved Techniques

Some new techniques and station models which were designed to eliminate most of the shortcomings listed above have been described [1, 2, 3]. Following is a discussion of the basic concepts of these new techniques as they are related to the previous considerations.

It is apparent that some means of automatically plotting the observations in final form is required to eliminate large plotting staffs. From an equipment-design point of view it is in general more convenient to represent the observations by plotting graphs automatically, rather than by printing numbers, on the appropriate maps or charts.

Facsimile communication equipment is already in use for automatically plotting some of the observations (a few thermodynamic diagrams) in graphical form. However, some serious limitations to the transmission of observations by facsimile are:

1. High carrier frequencies are required, thus limiting the usable types of transmission lines and complicating or eliminating the processes of storing, editing, and routing in the communication system.
2. All observations must be collected, probably by means other than facsimile, plotted in a standardized form, and retransmitted.
3. All automatic plotting is limited to a standardized form for transmission so that the individual analyst has little or no possibility of selecting those charts, parameters, or particular values which are best suited for his individual purposes.

A direct-writer type of communication system seems to be ideally suited to the transmission of synoptic observations. In this system, signals representing the coordinates of any desired graph are transmitted. These signals cause a pen to draw the graph, or, if desired, digital numbers corresponding to the desired observation at discrete points can be printed. Such a system has none of the limitations listed for the facsimile system, and it could also be used at least as efficiently as facsimile for the transmission of analysis of weather conditions. No such direct-writer system is as yet sufficiently developed for universal use, but there are no apparent serious technical difficulties to be overcome in its development.

In order to reduce the number of sheets of paper required for complete representations, station models which use the least space should be adopted. In this respect it is advantageous to use graphs, rather than digital numbers, since in a graph but one point is required to represent a given observation. This economy of space is best realized by choosing parameters for the description of the observations so that a minimum of grid, or reference, lines are required for their interpretation. This elimination of grid lines prevents confusion because of their overlapping when graphs for different stations or times are plotted close together.

Three-dimensional representations of the observations can be obtained with graphs arranged according to isometric drawing principles. For example, the height scales of all graphs can be drawn parallel to each other with their origins at points on a map corresponding to the positions of the observing stations. They can then be viewed as, say, telephone poles sticking up into the air. The values of observations at any given height can be represented by lines drawn from the graphs of the observations to the height scale. A three-dimensional illusion is then obtained by considering the lines to be crossbars on the telephone poles, with lengths proportional to the values of the observations.

This isometric technique can also be used to provide useful representations of the continuous time variations of surface observations throughout the region considered. Parallel time scales, with one particular time at the geographical positions of the stations, can be visualized as lying along the ground.

By placing both the time and height graphs on the same map, *all* the observations of a particular parameter or group of parameters made in a given region and time interval can be plotted on one piece of paper. Examples of this technique [3] indicate that it can materially aid an analyst in ascertaining the three-dimensional distribution of conditions as well as the surface time variations of these conditions. These examples also indicate that much analysis in the form of the drawing of isopleths can be eliminated when only a general concept of the spatial distribution of a given parameter is desired.

Graphs drawn according to these isometric principles have the useful property that the relative geographical positions of the observations are precisely maintained. This is seen from the fact that a given value of an observation at a given time or height is represented by a point which is displaced from the origin by the same amount at each station. This property permits convenient analysis of almost any desired conditions. For example, the height of and conditions on any desired constant-pressure surface, isentropic surface, inversion, front, etc., are readily available. Similarly, the conditions in any desired vertical cross section are represented on the isometric maps.

Additional advantages can be obtained by using special parameters for describing the observations in terms of dynamic models. Some parameters of this type are described below.

The Hydrostatic Equation. Useful continuous representations of the pressure-height relationship in the vertical can be obtained by using parameters defined in terms of deviations from arbitrary standard conditions [1]. These definitions amount to defining a standard column of air for use as a barometer. The parameter for describing the intensity of pressure then becomes the height z_p at which that pressure occurs in the standard atmosphere. Heights at which given pressures occur in the actual atmosphere can then be described either by the mean sea-level height z, or by the deviation from the standard height D, defined by

$$D = z - z_p. \tag{1}$$

The use of D eliminates most of the variation of pressure with height and very accurate pressure-height curves (D plotted as functions of z_p) require relatively small space. These curves permit rapid, easy, and accurate correlations between observations measured and plotted with respect to pressure z_p, and observations measured and plotted with respect to height z.

It is convenient to choose the standard atmosphere with which to define z_p to be that which is also used for the calibration of aircraft altimeters. The quantity D then becomes merely the additive conversion factor to be applied to pressure values z_p, as measured with altimeters, to obtain heights z. This choice of definition of the standard atmosphere provides an integration of meteorological and flight techniques which should be advantageous for all concerned.

In terms of the parameters D and z_p, the hydrostatic equation [1] is linearized to the form:

$$\frac{dD}{dz_p} = S^*, \tag{2}$$

where S^* is the virtual specific temperature anomaly defined by

$$S^* = \frac{T^* - T_p}{T_p}. \tag{3}$$

Here T^* is the virtual temperature at any given pressure, and T_p is the temperature at that pressure in the standard atmosphere.

If temperatures are represented by the parameter S, they are readily expressed as change of height D per unit change of pressure z_p. Accurate correlations between any temperature conditions and the corresponding pressure-height relationship are then immediately obvious. This change from the usual exponential form of the hydrostatic equation to a linear form appears to be even more advantageous than the comparable use of decibels in sound, light, or electrical measurements.

It appears that the use of D, rather than "sea-level" pressures, for expressing the values of surface pressure observations [1, 2] has several advantages. Chief among these are more accurate representations of horizontal pressure gradients near the surface of the earth and convenient exact correlations with upper-air pressure-height relationships.

The Gradient-Wind Equation. The use of parameters such as z_p, D, and S permits the continuous representation in the vertical of the wind, pressure, and temperature fields as related in the gradient-wind model. This possibility facilitates the assimilation of the three-dimensional atmospheric conditions. It also permits the convenient analysis of the contours of any desired constant-pressure surface so that the extensive use of pressure-pattern navigational techniques becomes feasible. Such representations would also be very useful for precise aircraft altimetry.

The present barbed-arrow type of representation of the winds does not seem to be suitable for exhibiting the continuous variations of wind with respect to time or height. Another more satisfactory method of representing such wind observations [2, 3] consists of drawing graphs of the east-west and north-south components of the wind as functions of height or time.

Air-Mass and Frontal Analysis. The techniques of isometric graphical representations can probably be applied with advantage to air-mass and frontal analysis. This is especially true of maps which contain all the temperature and humidity observations [3]. Such maps, together with similar maps of pressure, wind, and cloud conditions, should provide convenient means of increasing the ease and accuracy of both detailed and general analysis of air-mass and frontal distributions and characteristics. At present the major inadequacy of the isometric maps for this purpose appears to be the difficulty of plotting the cloud and state-of-weather observations from present teletype reports.

It is apparent that the efficient transmission of all observations can most conveniently be accomplished in terms of the continuous variations of the various parameters. For example, upper-air winds should be transmitted as continuous functions of height; radiosonde temperature and humidity observations and calculated heights should be transmitted as continuous functions of pressure; and surface observations should be transmitted as continuous functions of time. Of present observations, those of clouds and the state of the weather offer the most difficulty for this type of transmission, primarily because of the tremendous amount of such information available. Some techniques have been proposed [3] for the continuous representation of clouds, etc., which, though not yet completely satisfactory, do indicate that very useful results can be expected from this method of approach.

The use of a direct-writer communication system seems to be necessary for the transmission of all the observations since automatic segregation of particular observations from the great mass of data would undoubtedly be required. This segregated plotting would be feasible with direct-writer systems either in terms of graphs, proportional lines, or digital numbers, as desired by each individual analyst.

Conclusion

The inadequacy of present techniques of representing synoptic observations can be traced to three sources: the use of inadequate parameters with which to describe the observations, the use of inadequate methods of plotting these parameters on maps or charts, and the use of inadequate communication systems. It appears that most of these shortcomings can be eliminated with the following techniques: the use of parameters defined in terms of deviations from standard atmospheric conditions, the use of isometric graphical representations of the observations, and the use of a direct-writer type of communication system.

These new techniques are independent of each other, and could be adopted individually as opportunity or desire permits. For example, all radiosonde observations could be reported and plotted with present techniques in terms of z_p, D, and S. In fact, the parameters

D and z_p are now used for aircraft observations and reports. Similarly, thermodynamic diagrams are now being transmitted by facsimile. This transmission could probably be improved by altering its form to that of an isometric map. It might also be advisable to transmit all upper-air wind observations in a similar fashion with the present facsimile facilities. Finally, in view of the flexibility of direct-writer equipment in comparison to facsimile equipment, the former should be developed for the communication of weather information.

REFERENCES

1. Bellamy, J. C., "The Use of Pressure Altitude and Altimeter Corrections in Meteorology." *J. Meteor.*, 2:1–79 (1945).
2. —— *Graphical Representations of Meteorological Observations: Preliminary Report.* Dept. Meteor. Univ. Chicago, Mimeogr. Rep., 1947.
3. Cook Research Laboratories, Chicago. *Graphical Representation of U. S. Weather Observations.* Red Bank, N. J., Watson Laboratories Contract No. W28-099-ac-394, 1949.

METEOROLOGICAL ANALYSIS IN THE MIDDLE LATITUDES

By V. J. OLIVER

U. S. Weather Bureau, Washington, D. C.

and M. B. OLIVER

Washington, D. C.

Introduction

Fundamentally there are two diverse analytical approaches to the treatment of data representing synoptic meteorological situations. One is the research approach. Here time is of no great consequence in the selection of the processing techniques to be considered in the analysis. The other approach is that of the synoptic meteorologist constrained by the exigent demands of time to select that part of the vast array of existent raw data and charts which may be assimilated expeditiously. The particular elements chosen must combine to give, with the greatest dispatch, the most accurate delineation of the meteorological complex. Limited time is, in effect, the crux of the problems present in daily synoptic analysis.

The analytical techniques selected for daily synoptic practice stem in reality from the methods formulated in the course of meteorological research, although it is frequently necessary to modify these methods to attain a maximum economy of the time and effort of synoptic analysts. Selection of analytical procedures must always be made with an eye toward comprehensiveness combined with operational simplicity. With these criteria as a basis the authors will attempt to present here a critique of the principal synoptic analytical practices.

In view of the significance of the whole problem of analysis it is well not only to review the techniques of analysis used at present and in the past, but also to attempt to appraise the meteorological research currently being pursued, estimating the measure to which each relevant scientific disclosure fulfills the fundamental requirements of daily synoptic analysis. Although an attempt will be made here to review the present and past analytical techniques in some detail, it is feasible to consider but a few examples of meteorological research with the purpose of evaluating the applications of such research to daily synoptic practice.

During the past twenty years, weather map analysis in the United States has undergone two major changes in technique. The first was the result of the adoption by United States meteorologists of the frontal technique of separating air masses of different properties. This classical system of analytical procedure, first suggested by the Norwegians [4], has been covered elsewhere in papers by Bergeron and Bjerknes. Their methods, now utilized by most meteorologists, satisfactorily delineate a large portion of the meteorologically significant motions of the atmosphere and ascribe, at least tacitly, the related weather phenomena to these causative motions. With the establishment of a network of upper-air soundings, the second major change in analytical procedure evolved: upper-air analysis, as an adjunct to Norwegian frontal technique, augmented the scope of meteorology by rendering the analysis systematically three dimensional.

Today the application of electronics to meteorology, including such adaptations as radar and sferics, portends a third progressive development in the field of analysis. The implications of a portion of these studies will be examined subsequently. But here it will suffice to say that the changes induced by electronics are far from reaching their full fruition, although exceptional advances have been made along these lines in the few short years since the use of electronics became widespread.

Paralleling the advances in the technical phases of analytical procedures, there has been a trend toward a new type of organization for meteorological analysis, an organization which permits of the more productive use of time in the field stations by eliminating duplication of effort. This trend has culminated in the rise of analysis centers and the transmission of charts by facsimile reproduction. It seems likely that the future will see an even greater specialization of meteorological personnel and a concomitant concentration of analytical work. Eventually we may even see in concrete form the now visionary picture of a fully automatic analysis center, in which the observational data are fed directly into a machine which will sort and evaluate them, with the resultant analysis proceeding without an intermediary into a vast electronic and mechanical "forecaster." The first rudimentary approaches to this type of analysis are being made by Panofsky [15] at New York University. But the distance to the ultimate goal is so great that in this paper we shall confine our discussion to meteorological centers operated by human beings. Granted even this limitation, we find that the importance of analysis centers cannot be fully actualized until there are improvements not only in the field of meteorology, but also in the field of weather communication.

Up to the present time the facsimile method of transmitting analyses has been the only technique of communication that has been put into operation which obviates the need for personnel to plot the incoming information at the receiving station. Such a method is commendably in line with the trend toward centralization and economy of human effort, but the particular facsimile method in use at this time has various drawbacks. In particular, it is too slow, and it cannot be used to transmit color. Moreover, the data as repro-

duced by the receiving mechanisms are so blurred as to preclude the transmission of meteorological reports from each of the stations of our present-day observational network. Since such methods of communication as automatic writing and electronic scanning and transmissions already exist and have been brought to a high level of efficiency and refinement in some of their applications, it seems likely that meteorologists will in the near future have available some improved communication system adaptable to detailed transmission work. Bellamy [3] has been working on this problem and his suggestions regarding it are discussed in an adjoining article.[1] The whole problem of the transmission and physical representation of meteorological data is pressing and should be an object of concentrated meteorological research.

Given an adequate method of transmission, with analysis centers located in strategic communication centers, even more specialized functions can be performed by the analysis centers. In addition to analyses, elaborate prognostic charts including information concerning precipitation and cloudiness along with isobars and isotherms could be sent out to the individual field stations to be reproduced there automatically. Local meteorologists at the field stations could apply objective techniques of forecasting to the various charts and data received, with the purpose of prognosticating the weather in detail for their respective limited areas.

In the past, one of the difficulties which has arisen in connection with almost all analysis centers is a lack of confidence on the part of the men in the field in the analyses and prognostic charts sent out from the centers. The best way to correct this is, of course, to see that the central analysts are drawn from the ranks of the very ablest meteorologists, and that they work without the undue haste necessitated by meeting ill-advised transmission deadlines. Needless to say, these expert meteorologists would make use of all the latest refinements in the techniques of applied meteorology.

We shall now consider some of these techniques more specifically. As has been pointed out, our present analysis is based largely on the concept of fronts originated by the Norwegian meteorologists. There are, however, certain types of weather which seem to lie somewhat beyond the scope of the classical frontal models. The sort of phenomenon we have in mind is the prefrontal squall line (or instability line as it is now generally known).

The Prefrontal Squall Line

The prefrontal squall line is a line of showers or thunderstorms, which often appears in the warm sector of a cyclone, extending in a line roughly parallel to the cold front. It is usually not more than five hundred miles ahead of the front and is most often noticed between one hundred and three hundred miles ahead of the cold front. Meteorologists at first believed this squall line to be either the principal cold front which some previous analyst had moved too slowly, or a weak cold front previously dropped, which, if it had been retained and moved along with the speed indicated by the baric field, would be at the present location of the squall line.

During the past few years the independent reality of prefrontal squall lines has become generally accepted. Airline meteorologists [9] and others have brought this phenomenon to the attention of meteorologists as a whole and have ascertained rather conclusively that the prefrontal squall line does not represent a misplaced cold front, as had previously been thought. At most analysis centers an attempt has been made to include these squall lines in the analysis and to use for the detection of these lines the criterion of persistence in time and space of a line of showers. Such a standard for detection may be depriving the forecasters of most of the prognostic value of the concept of prefrontal squall lines in that it fails to provide a means for the recognition of the precursors of the squalls. Because of the lack of knowledge of the mechanics of the formation and structure of these phenomena and the want of adequate observational techniques for their detection, there is much to be desired in the analyses. Recent investigations, however, give promise of an imminent improvement in this part of the analysis program.

These investigations have followed two avenues of approach. On the one hand, Fulks[2] and Williams [24] have constructed detailed descriptive models of the actually observed distributions of temperature, moisture, and atmospheric pressures associated with squall lines. On the other hand, a new theory of squall lines has been formulated, made possible by the introduction into meteorology by Freeman [8] of the concept of pressure jumps, the atmospheric analogy to the "hydraulic jump." Tepper [22], combining data of the types presented by Williams and the concept of pressure jumps in the atmosphere, has developed the theory that squall lines are caused by pressure jumps. He has demonstrated how squall lines and their concomitant weather can be initiated and propagated by pressure jumps and has further proposed that the intersection of two pressure-jump lines has properties which favor the formation of tornadoes. Currently, Tepper is testing this hypothesis by means of empirical data. The results of this work should make possible rapid improvement of our understanding of, and ability to forecast, squall lines and tornadoes.

An article by Freeman[3] describing in detail the results of this type of inquiry into the nature of prefrontal squall lines is included elsewhere in this Compendium. Here we shall restrict our discussion to only a portion of the new results emerging from this research, that is, to those properties of the pressure jump which are pertinent to the problem of analytical procedure.

Although still in a very early stage of development,

1. "Models and Techniques of Synoptic Representations" by J. C. Bellamy, pp. 711–714.

2. Consult "The Instability Line" by J. R. Fulks, pp. 647–652 in this Compendium.

3. Consult "The Solution of Nonlinear Meteorological Problems by the Method of Characteristics" by J. C. Freeman, pp. 421–433.

the investigations of pressure-jump phenomena show such promise that it seems timely to consider the steps which would have to be taken to revise our present procedures of atmospheric analysis with the purpose of taking full advantage of what we already know about this phenomenon. The first revision of synoptic practice that is indicated by this work would be at the observing level, namely, the development of the necessary instrumental and observational techniques for detecting the passage of a pressure jump at each meteorological station. Secondly, we would be obliged to codify and transmit the data pertinent to pressure-jump passages from the observation point to the forecast and analysis centers. In all probability it would be necessary to make use of the code for special observations to report the position of the pressure jump when it is initially noted. Subsequently its position could be included in our three- and six-hourly coded reports.

Finally, it would be essential that we decide upon conventions for detecting pressure-jump lines on our synoptic charts. This would entail, first, a symbol for entry on our synoptic charts to represent the pressure changes associated with pressure jumps and, secondly, a set of rules for enabling analysts to locate, uniquely, pressure-jump lines on our synoptic charts. The method of graphical representation of pressure, temperature, wind, and moisture data, suggested by Bellamy [3], appears to be well suited for showing the data pertinent to the detection of pressure jumps, while at the same time it lends supplementary emphasis to the other portions of the data used in the type of air-mass analysis customarily performed.

Bellamy's method for the graphical representation of data warrants the attention of all meteorologists since it constitutes, in point of fact, a convenient and practicable means for applying to analytical procedure the disclosures of many divergent investigations in the field of modern meteorological research. We have already touched upon the usefulness of Bellamy's method apropos of pressure jumps. Turning our attention now to an entirely different topic, namely, the meteorological significance of the results of recent radar studies, we find that again Bellamy's graphical plotting method may profitably be applied.

Use of Radar as an Analytical Tool

The advent of radar had sudden dramatic repercussions throughout the field of physical science. In meteorology A. Bent and R. Wexler [23] were among the first to apply this new scientific tool, utilizing radar in the investigation of the structure of precipitation clouds. Subsequently others have conducted further research along these lines.

Recent radarscope movies, produced by A. C. Bemis at the Massachusetts Institute of Technology, where this work has been carried on for several years, virtually compel us to adopt some new type of cloud analysis. These radarscope pictures refute the classical model of an active cold front attended by a rather solid line of cumulonimbus with occasional holes between the cells, frequently replacing this model with the picture of the cumulonimbus centers themselves, as well as the breaks in the clouds, organized in space in a series of rows, roughly parallel to each other and nearly parallel to the atmospheric flow above the frontal surface. Similarly, the disclosure that there are, as a rule, parallel bands of precipitation associated with warm fronts constitutes a marked departure from the accepted model. Bellamy's cloud graph, with height above the surface as the ordinate and time as the abscissa, permits inferences concerning the horizontal distribution of frontal clouds.

The same sort of graph has been used to represent the data obtained by turning a radarscope upwards to investigate the distribution in the vertical of clouds associated with fronts. At the time the radar studies were made, pilots and most meteorologists connected with airplane operations had already come to disagree with the model of the clouds associated with an approaching frontal system which is found in most textbooks: the familiar picture of cirrus gradually thickening and lowering, merging into altostratus, and finally stratus, with a solid cloud deck extending from the cirrus level down to the lowest cloud base. Radar studies have confirmed the findings of those who have maintained that the cloud pattern over a warm front consists of several distinct layers of clouds, generally not merging except over limited areas. Radar photographs of the vertical structure of clouds are now being made at the Signal Corps Laboratories in Belmar, New Jersey, under the able direction of Dr. Michael J. Ference. One of these pictures, illustrating the cloud deck in advance of a warm front from a coastal storm in the eastern United States, is reproduced as Fig. 9 (p. 1220) in the article by Dr. Ference in this Compendium.

In view of these findings, a systematic investigation of our classical model of cloud structure should be undertaken to determine details concerning the apparent wave pattern and layer structure. Certainly these changes should be brought to the attention of pilots and all those who are responsible for the meteorological training of pilots. We need, too, to perfect and put into operation the instruments needed to make observations of both horizontal and vertical cloud patterns so that such data can be transmitted in our regular synoptic weather codes. Once these data have been received in a weather station, Bellamy's method or some other method of graphic representation can furnish a practicable basis for the analysis of clouds.

Upper-Air Analysis

Since the structure of clouds is intimately connected with the configuration of upper-flow patterns, the revision of our model of cloud structure implies consonant changes in the technique of evaluating the data on upper-level charts to include the analysis of vertical motions. Although research meteorologists have already tackled this problem, unreliable data have forestalled its solution. For this reason we shall restrict ourselves to the consideration of the methods of analysis of upper-level charts in general use today.

Several diverse techniques of analysis exist, differing primarily in their method of attaining a verisimilar upper-level baric pattern over regions of sparse reports. In such areas upper-level contours are not uniquely determined by the wind and height data now available. Thus some sort of systematic extrapolation is necessary.

One such system is the technique of differential analysis. Approximately ten years ago the basic principles underlying this analytical method were formulated in the United States by Rossby and Starr and put into practice in analysis by Willett and Namias in extended-range forecasting. During World War II Petterssen [16] and the meteorologists of the United Nations who collaborated with him in England fully developed this technique into a powerful analytical tool in conjunction with forecasting high-level winds for bombing operations. Another form of differential analysis had previously evolved in Germany under the leadership of Scherhag [20].

The German method of differential analysis differs from the others in that it is based upon the partition of the atmosphere into layers chosen in such a way that the thickness of any one layer is approximately equal to that of adjacent strata. Below 500 mb, Scherhag's technique is virtually identical with Namias' method, which will be treated in detail below. For the very high atmosphere, Scherhag has compiled statistical tables relating the thickness of each of his selected strata with the height above sea level of the lower limit of the layer. By noting the difference between the statistical values for the thickness of the various layers and the observed thicknesses, where radiosondes are operated, Scherhag is able to extend the application of differential analysis up to the high stratosphere. Herein lies the significance of Scherhag's development. In this age of stratospheric aviation with its concomitant demands for wind forecasts at very high levels, Scherhag's method constitutes a systematic technique for extrapolating data throughout the stratosphere to obtain a logical, internally consistent, baric pattern.

For mid-tropospheric levels the best method devised for extrapolation of upper-air data is a combination of vertical extrapolation from individual reports, plus a horizontal distortion of normal thickness lines. This distortion must conform with all available reports at the level in question, the surface frontal configuration, the previously observed thickness pattern, the thermal winds, and the major atmospheric circulations at sea level and aloft. Such a method of differential analysis, described by Namias [12], is now used by the Weather Bureau for drawing the Northern Hemisphere daily charts for 700 mb and appears equally applicable to all levels below the tropopause. This technique makes use of monthly normal thickness maps, drawn at biweekly intervals on a transparent plastic material, which depict the normal thickness between the 1000- and 700-mb levels.

The actual procedure as developed by Namias is to superimpose upon the completed surface map the appropriate normal thickness chart. On top of both is laid the 700-mb chart, plotted on a tissue paper base, since the translucent tissue paper allows the analyst to see the lines on the two charts beneath it. Finally, the 700-mb chart is covered by a transparent plastic sheet on which the thickness lines for the current map are to be drawn.

Before the analysis begins, the following data are plotted on the 700-mb chart:

1. At all radiosonde stations—the 700-mb temperature, wind, and height, and the thickness from 1000 to 700 mb;
2. All available winds at 700 mb (lower levels if none are available for 700 mb);
3. At ships located in a region where the meteorological situation permits an accurate estimation of the lapse rate—the estimated thickness between 1000 and 700 mb;
4. In regions of sparse data—the thermal winds between gradient level and 700 mb.

With these data available, the analysis begins. First, the analyst draws lines of constant thickness in the regions where data are scanty. (Such regions now include the greater part of the oceanic areas of the Northern Hemisphere, Siberia, and most of southern Asia.) This is done by distorting the normal thickness lines to fit all the upper-air data, the thermal winds, and the circulation patterns of the surface chart, with due regard to the continuity of the mean temperature field. An example of an observed thickness pattern and the associated surface frontal and pressure fields is presented in Fig. 1. After the thickness lines are completed, the height of the 700-mb surface is determined by adding the thickness to the 1000-mb height. The latter is obtained by converting the sea-level pressure to the 1000-mb height. In actual practice the analyst computes the height of the 700-mb surface at the intersections of latitude and longitude lines, using a ten-degree spacing of both longitude and latitude. Thus a grid of 700-mb heights is obtained.

With this grid in the regions of infrequent reports, and with the radiosonde data elsewhere, the analyst can, with assurance, draw in the contours on the 700-mb chart. This method has the advantage of insuring horizontal as well as vertical consistency and combines expedition of analysis with maximum accuracy over land as well as over oceanic areas.

At present, however, this method has one serious shortcoming: lack of sufficient data precludes the construction of normal thickness charts in the upper layers of the troposphere and in the stratosphere. There is thus an empirical ceiling on the efficacious use of this method. Above 500 mb, Scherhag's system of extrapolating data for the construction of baric charts still is applicable, as is another method, based upon a study conducted by Riehl and LaSeur [19] of high tropospheric lapse rates. The general technique is little different from the long-standing practice of assuming a lapse rate at a point and thereby determining the pressures aloft at that point. For instance, the height of the 700-mb surface can be attained with reasonable accuracy by assuming a moist-adiabatic lapse rate up to 700 mb above a ship reporting a shower, but the choice of an appropriate lapse rate

in more complicated synoptic situations and at high tropospheric levels is not so easy. What Riehl and LaSeur did, essentially, was to determine an appropriate lapse rate between 700 and 300 mb for each typical synoptic flow pattern. Then using the 700-mb chart as a guide and the proper, calculated, lapse rates, Riehl and LaSeur were enabled to obtain reasonable estimations of upper-atmospheric pressures up to 300 mb. One decided advantage of their technique is that the lapse rates, being statistically determined, are obtained independently of the mean temperatures in the layer from 700 to 300 mb. Since, however, these two quantities are so closely related, mean temperatures may be used as a check on the accuracy of the extrapolated hypsometrical values of the 300-mb surface, contributing in no small way to the value of the subsequent analysis.

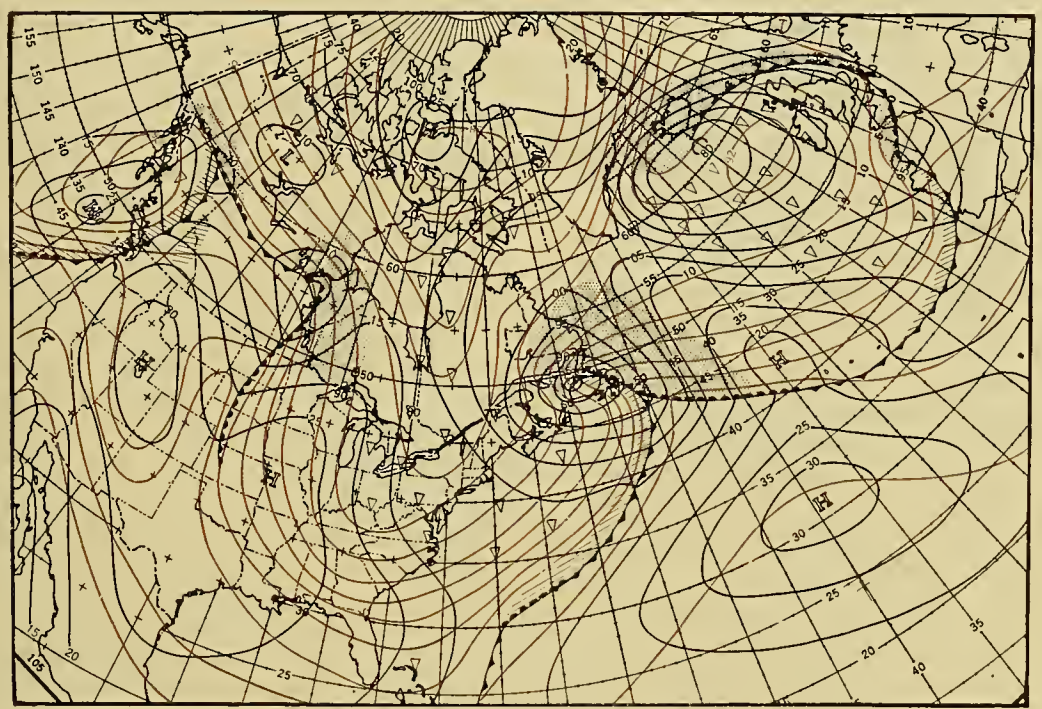


FIG. 1.—Example of observed thickness patterns in the vicinity of several different frontal systems. Red lines are 1000–700-mb thickness; black lines, sea-level isobars.

There is still another approach to the problem of the analysis of regions of exiguous data, based not on horizontal nor vertical extrapolation, but rather on extrapolation with respect to time. When the surface data are particularly scant, vertical extrapolation is possible only at widely separated points and is therefore incapable of providing sufficient data to insure a unique analysis. Likewise, under these particular circumstances the technique of extrapolation based on the horizontal distortion of thickness lines is of but meager productivity. When these are the circumstances, reliance upon extrapolation at a point with respect to time produces the optimum analytical results.

The quintessence of the information derived from the technique of time extrapolation is the structure of the pressure system passing a single point and, in addition, the direction in which that pressure system is moving. Notwithstanding the very definite limits of the spatial range throughout which the deductions drawn by this method can be validly applied, an analysis, which for practical purposes is unique, can be obtained for a considerable region about each point where the extrapolation with respect to time is carried out. This area for which the extrapolation is valid varies with the synoptic situation. Details of the interpretations of the changes of various data with time are found in the studies of single-station analysis made at the University of Chicago in the early 1940's [14]. More recent applications of this sort of extrapolation for daily analysis have been made in such diverse regions as the Aleutians, where time cross sections were used to aid in analyzing the vast unpeopled expanses of the Pacific Ocean, and in the tropics. Riehl [18] recently published a paper elucidating the use of extrapolation with respect to time in the solution of problems of tropical analysis.

In the field of analytical research, time cross sections are tools which have shed light on countless perplexing questions. In research, one particular class of time cross section has enjoyed widespread and manifold uses [10]. This consists of a graphical representation of some element (*e.g.*, pressure) plotted on a diagram in which time is used as an ordinate and a spatial unit (*e.g.*, longitude) is the abscissa. In this hypothetical case, lines could be drawn on the graph portraying the position of troughs and ridges with respect to time. The divergent uses to which this sort of analysis has been put attest to its value in clarifying significant meteorological relationships.

Coordinated Use of Sea-Level and Upper-Air Analysis

Having once arrived at a satisfactory analysis of both sea-level and upper-level charts, we come to what is perhaps the greatest problem facing the forecaster today: the coordination of the sea-level data with the upper-air data. Most meteorologists are now familiar with both the sea-level and the upper-level charts, and routinely take both into account before rendering judgments upon the development and motion of storm centers and their associated manifestations of weather. Despite this fact there still appears to be an appreciable difference between the scope of the synoptic meteorologist's general knowledge of weather processes and the part of this information which he actually applies in analysis or forecasting. It appears that many forecasts miss their mark not because of our over-all lack of understanding of the factors which control the changes in the weather, but because the coordination of all the various pertinent details at all levels of the atmosphere in a relatively short time interval is a task too difficult to accomplish satisfactorily with our present system of map display and data representation.

If we try to picture what is needed in order to coordinate with facility the meteorologically significant features of the higher strata of the atmosphere with those at the surface, we shall arrive at something like the following:

First and most important, all the various charts depicting conditions at the surface and at upper levels should be the same size. This proviso, though simple, is vital if the forecaster is to be able to compare one level with another quickly and accurately.

Second, the charts for all levels should be drawn on semitransparent paper in order that one may be superimposed upon that for an adjacent level to facilitate examination of the changes of meteorological features with height and to insure internal consistency in analysis. Furthermore, it is obvious that we should use the same units for data at all levels if the maximum ease in vertical synthesis is to be obtained. Consider, for instance, the disordered array of units in current use: centigrade and Fahrenheit temperatures; Beaufort scale, knots, miles per hour, and meters per second denoting wind velocities; altitudes measured in feet, meters, or dynamic meters; and pressure analysis hampered by the use of isobars on sea-level maps and contour lines on upper-level charts. The system of units proposed by Bellamy [2] is of a unitary type. As has been mentioned, he has also proposed a system for the graphical representation of data, which obviates the need for converting numerical values from one system of units to another. By means of simple transparent overlays, the numerical values of the data in any desired units may be read directly from the graphs.

Although complete standardization of all units is not feasible at this time, we may even now progress towards this goal by adopting a single system of units for pressure. In particular, the unfortunate duality of units with isobars drawn on the surface map and contour lines on the upper-level charts can be abolished without complexity by replacing the sea-level isobars with the 1000-mb contours; the observational stations could report the height of the 1000-mb surface as well as the sea-level pressure.

At present, in our attempts at vertical synthesis of atmospheric elements, we labor under a difficulty other than that occasioned by a multiplicity of units; we refer to the unfortunate time interval which elapses between surface and upper-air observations. As a result of this lack of synchronization, many of the fundamental charts upon which analyses and forecasts are based are, in effect, invalid. Rectification of the observational program must inevitably precede sound coordination of upper-level charts with surface data.

A further argument in favor of using uniform map scales and units plus transparent or semitransparent map bases is that they make possible the construction of a chart which we believe to be most effective in bringing out the relationship between surface and upper-air circulations. This is simply the sea-level chart with upper-level contours traced on it, preferably using for the upper-level contours some color differing from that of the sea-level isobars. This method seems to be the most satisfactory for relating the patterns at upper levels with the weather and pressure changes occurring at sea level [13]. Such a chart can be used to especial advantage by those who must issue forecasts or complete an analysis of the synoptic situation in a very limited time. When time is pressing, analytical comprehensiveness is all too frequently sacrificed as long as the surface and upper-level charts are kept separate, for even when both are carefully studied independently, the possible inferences to be drawn from the changes of flow with height can be assimilated only by a careful comparison which consumes more time than is now available to most operational meteorologists.

If this general concept for the coordination of surface and upper-air data is accepted up to this point, the problem then arises as to the selection of the upper-level chart most productive from the standpoint of the number of utile inferences deducible from the combined analysis. This set of upper-level contours depends to some extent on the primary purpose to which the charts will be put. If we consider that forecasting the appearance of the sea-level and upper-level pressure charts 24–48 hours in the future is the primary problem, then we should choose for analysis the upper-level chart considered to be most significant for the investigation of steering, cyclogenesis, and anticyclogenesis. At present many meteorologists prefer the 500-mb chart over all others for this purpose. It would therefore be proper to superimpose on the surface map the 500-mb contours at analysis centers whenever pressure-pattern prognoses constitute the chief forecast problem.

On the other hand, if the primary problem is to forecast the cloudiness and precipitation for the coming twenty-four to forty-eight hours, then the upper-level chart chosen should be the one best suited for the detection of advection and the motion of the moisture layers which are most intimately associated with the observed cloudiness and precipitation.

The problem of such short-range precipitation fore-

casting has been under consideration by many investigators during the past few years [5, 17]. The three factors which have been found to be most important for precipitation forecasts are the orientation and curvature of the flow patterns at the altostratus level (about 700 mb), the location of areas of pressure fall at sea level, and the extension of warm-air advection near the 850-mb level. Since the detection of advection between 700 mb and the surface can be accomplished at a glance all three of these charts the relation between the two sets of isopleths and the weather and cloud data throws into full relief the following information: the prineipal activating mechanisms of precipitation and cloudiness, the areal extention of warm and cold air advection, the movement and spatial distribution of pressure-fall areas on the surface map with respect to the steering flow aloft, the relationship between moisture sources and the orientation of the upper flow, the likelihood of vertical

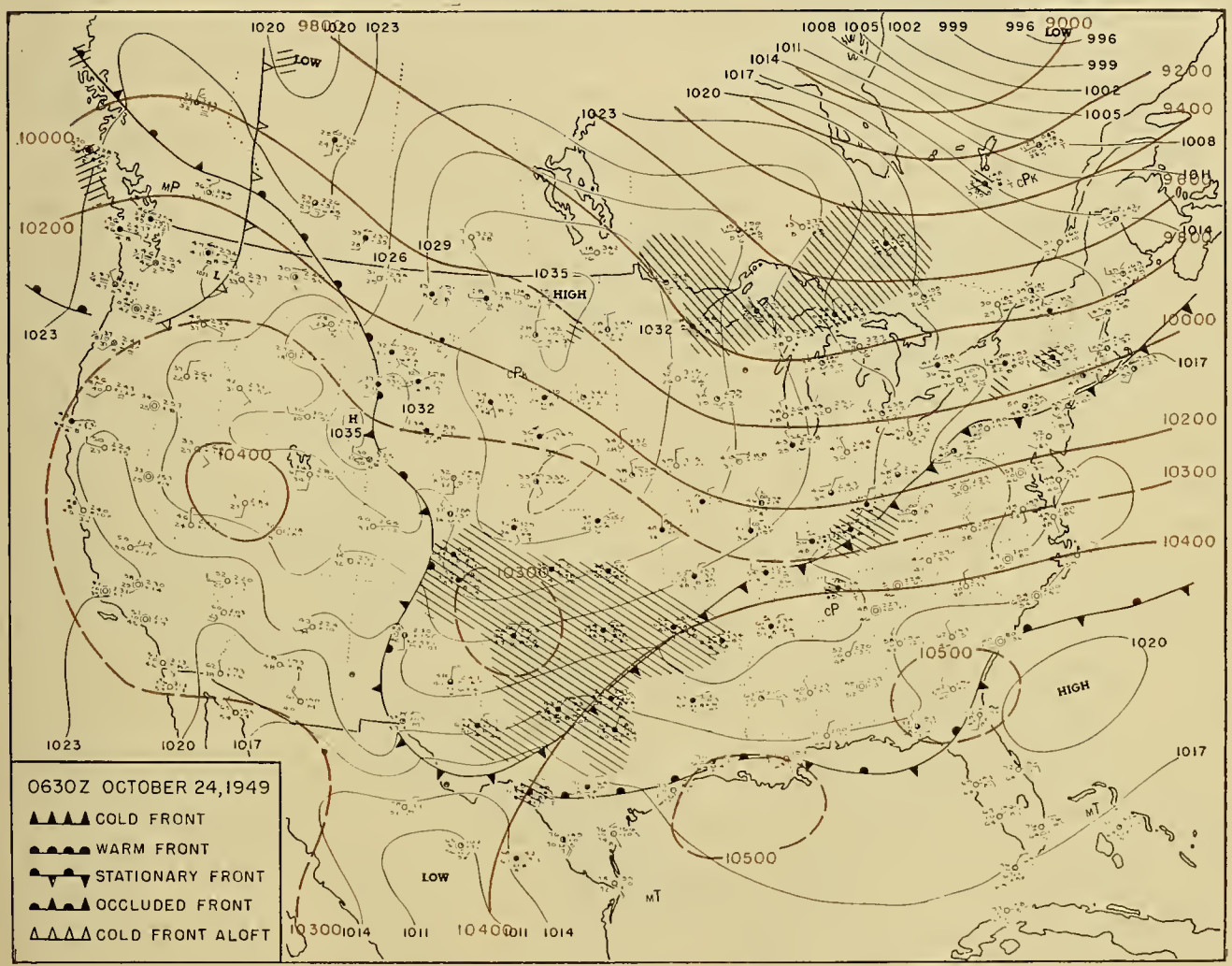


Fig. 2.—Combined sea-level and 700-mb analysis for the United States at 0630Z, October 24, 1949.

when the charts of these two levels are superimposed, and since the 700-mb contours represent the air flow at the level where most of the altostratus rain clouds are centered, the 700-mb chart superimposed on the sea-level chart is considered the most useful combination in giving a comprehensive view of the weather prospects in stations where issuance of 24- and 48-hr forecasts to the general public is the prime function. The authors feel that this is such an important subject that some examples of the combined analysis of surface and upper-level charts are presented here in detail.

These examples include three consecutive sea-level charts with the corresponding 700-mb contours superimposed upon them, constructed for 24-hr intervals. On displacement of an air mass above a front, and the areas along fronts where the meteorological properties of the atmosphere are most conducive to wave formation.

The most striking feature on the first of these examples (Fig. 2), is the large area of rain in the southern Great Plains States. The strong easterly flow indicated by the sea-level isobars in this area leads to the tentative assumption that the rain is caused by topographic upslope motion of the cold air mass as it moves toward the higher land areas of Colorado, western Texas, and western Kansas. In view of the configuration of the upper flow, however, it seems quite evident that the warm moist tropical air from the Gulf of Mexico is

moving northward and then northwestward over the advancing and deepening cold air mass. The explanation of the precipitation then, clearly brought out by the two combined charts, must be the enforced elevation of the tropical air, rather than orographic lifting of the cold air mass.

Other than that in the southern Great Plains, the only area of precipitation of any large extent is that indicated in the vicinity of Lake Superior. Here the obvious causative factor is the well-known heating effect of the Lakes. The important point here is the concurrence of cyclonic curvature of the sea-level isobars and the 700-mb contours with the area of shower activity. The intimate connection between the curvature of the flow pattern and the occurrence of such instability phenomena as the "Lake effect" stands out pellucidly by virtue of this particular method of superimposition of charts.

Turning next to the portion of the cold front extending from Tennessee northeastward to Massachusetts, we note two things: (1) that the flow from the west is stronger aloft than it is at the surface, indicating no lifting of the upper air mass by the eastward-advancing portion of the surface front and therefore little or no frontal weather; (2) that east of the Appalachian Mountains, the portion of the front which is moving slowly southward is parallel to the upper winds and should therefore produce lifting of the air above the frontal surface. In the area to the rear of this portion of the front some cloudiness and precipitation are in evidence. This is the only area where the combined analysis suggests the possibility of frontal precipitation directly behind the front.

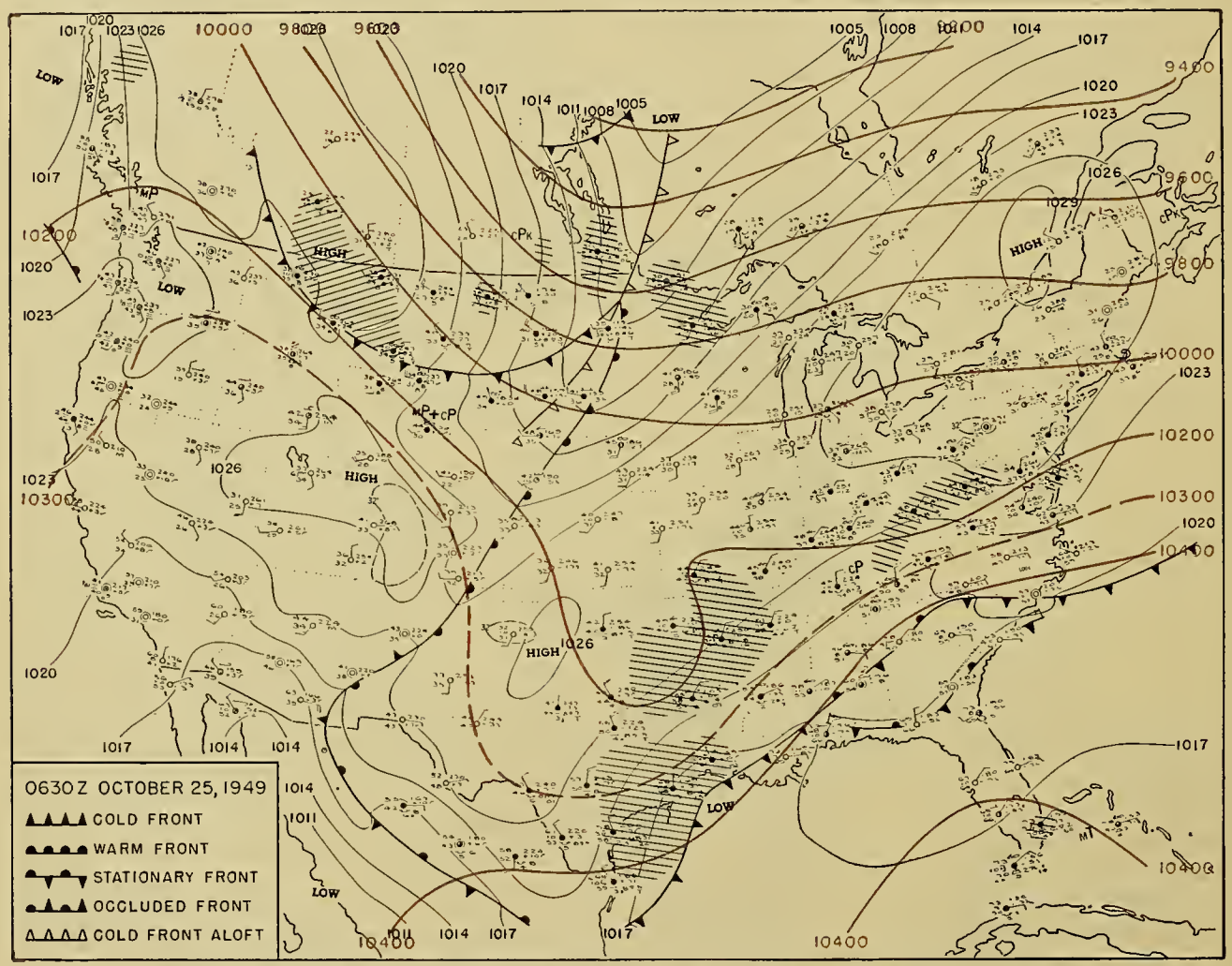


FIG. 3.—Combined sea-level and 700-mb analysis for the United States at 0630Z, October 25, 1949.

The front in the southeastern United States, with a wave along the coast of South Carolina, is in a position which often causes the weather along the coast to deteriorate suddenly. In this example, however, the presence of a closed high-pressure cell aloft near South Carolina clearly indicates that the front is lifeless and that the forecaster need have no worry lest it develop in the future.

Looking at the analysis again for the purpose of detecting the areas of warm- or cold-air advection, we come upon striking evidence that strong cold-air advection is occurring through Iowa, Illinois, Indiana, Ohio, and most of the area north of these states. Strong

advection of warm air is indicated only in eastern Montana. The sea-level pressure tendencies are rising throughout most of the area of cold-air advection and falling over the area of warm-air advection, as is usual [1]. Thus the distribution of advection may be appraised by only a cursory investigation of the combined sea-level and 700-mb charts.

This combination likewise promotes the ease with which the principle of steering may be applied to a specific meteorological situation. Examining Fig. 2 from the standpoint of steering, we can infer swiftly and with moderate assurance that the northern portion of the high-pressure area located in the northern Great Plains States should move southeastward and then eastward to near New York, while the portion of the cold-air dome which has moved south of the strong belt of westerly winds, into Texas and Oklahoma, will have to remain trapped in that region. The isolated low aloft over Texas is moving slowly eastward as is indicated by the sea-level pressure falls east of it and sea-level pressure rises west of it, but it seems clear that since it is now out of the westerlies it should not move far in the next 24 hr, and that any wave activity it might induce on the cold front could not move along the front to the northeast until after the passage of the rapidly moving high-pressure center in the northern Great Plains States.

The intimate connection between the upper-level circulation and the attendant weather is again dramatically illustrated in Fig. 3, the combined analysis for 24 hr later. Here in the South Central States we see that the precipitation and cloudiness from the weak wave in the Gulf of Mexico extends only as far north and west as central Missouri, while northern Missouri remains perfectly clear. Likewise the circulation around the upper-level center over Oklahoma shows that the moisture from the south is carried only to the same region in Missouri but no farther. The reason for the distribution of cloudiness and precipitation is not at all apparent from the surface map alone, while from looking at the upper-level map alone one would remain unaware of the actual distribution of weather. It is the combination of the two sets of data that presents the complete, pellucid picture of the synoptic situation.

If we examine the combined analysis over the southeastern states to elicit the explanation of the widespread cloudiness to the north of the stationary front, we find that the orientation and curvature of the upper-level contours with respect to the surface pattern is again rewarding. Considering first the orientation of the upper-level contours in this region, we note that the upper-level flow is essentially parallel to the surface front. This indicates that there is no downslope flow over the frontal surface, a circumstance which greatly enhances the likelihood of postfrontal middle clouds. In this particular instance, the existence of clouds in the area to the rear of the front becomes a certainty, since the upper-level flow emanates from a region of ample moisture supply.

Let us consider next the curvature of the upper-level contours over the South Atlantic coast. Directly under a portion of the anticyclonic flow we find on the surface map a wave on the front in South Carolina. Since the formation or perpetuance of a wave is unlikely beneath such a flow aloft, this wave should not be of much synoptic significance unless the upper-level contours above it change their characteristics of curvature, orientation, or both.

Over West Virginia the weak cyclonic curvature of the upper contours in conjunction with a certain amount of warm-air advection, seems adequate to have caused precipitation in that region. Again the surface and the upper-air analyses considered separately fail to give a lucid explanation for the distribution of cloudiness and precipitation; with both charts examined in combination, the causative factors stand out.

Let us continue our examination of the combined analysis in Fig. 3 with a view toward determining areas of temperature advection in the layer between the surface and 700 mb. The salient features of the field of advection are the strong warm-air advection from Iowa northeastward, the strong cold-air advection in North Dakota and Montana, and the cold-air advection in southeastern Texas, all so plainly revealed as to obviate further comment. Similarly the significant features of the steering pattern are apparent. The contours at the upper level indicate that the frontal system in the northern Great Plains States will continue to move rapidly eastward and that the trough associated with this system will in all probability overtake the slowly moving trough in Oklahoma. The wave located south of Louisiana is being steered to the northeast by the increasingly strong southwesterly winds over it. At the same time the combined analysis reveals the interesting probability of the development of this wave. Several circumstances favor this development: the wave on the surface is close to the upper-level center, with the flow above it cyclonically curved; there is air-mass contrast across the front; and the solenoidal field is being intensified by increasing cold-air advection to the rear of the wave with pronounced warm-air advection in advance of it.

The combined analysis illustrated here successfully expedites the analysis by making it easy to evaluate all these factors with the use of but one chart.

Figure 4 affords an even more vivid illustration of the use of combined analysis for bringing out the well-established but all too frequently unnoticed relationships between the upper-level flow and the concomitant sea-level patterns and weather phenomena. In Fig. 4 the difference between the two cold fronts in the central and southeastern portion of the country with respect to the extent of postfrontal middle cloudiness again correlates nicely with the characteristics of the upper flow over the frontal surface: Where the upper flow is parallel to the front and has cyclonic vorticity, there is rain; where the upper flow, increasing as it is with height, is nearly perpendicular to the surface, front the skies are clear. Along the Divide in Montana where the front is being overrun by westerly and northwesterly winds aloft, there is no precipitation. In this area the sharp anticyclonic turning of the upper-level flow suggests

that the front should remain inactive as far as cloudiness and precipitation are concerned.

A glance at Fig. 4 shows that the cold high has moved from northwestern Montana to northeastern Kansas, in accordance with the steering indicated by the upper-level flow.

It is the ease with which a truly comprehensive picture of the synoptic situation may be gained which commends the method of combined analysis presented here. With other methods the same results may, of course, be obtained in time. But the authors feel that the saving of time and energy effected by this method is of sufficient significance to warrant its inclusion in the standard operational procedure used for analysis.

Before leaving the problem of how best to facilitate the comprehensive, three-dimensional visualization of a synoptic situation, one further point should be considered: the total area of the meteorological map base being used must be so restricted in extent as to insure ease in superimposition together with ease in comprehension. A map whose areal dimensions are greater than about three feet square is difficult to handle and is too large in scope to see all at once. Meteorologists who are required to make two- or three-day forecasts need maps covering considerably more area of the earth's surface than is needed for shorter-period forecasts. If this additional area is achieved by increasing the size of the map, the map soon becomes so large that it is difficult to perceive the relationship between changes on opposite sides of the world. Such concepts as the spread of energy downstream from a newly formed storm or the creation of new troughs whose location is indicated by constant vorticity currents remain inaccessible to the synoptic meteorologist unless his working charts are small enough so that he can study a large area simultaneously.

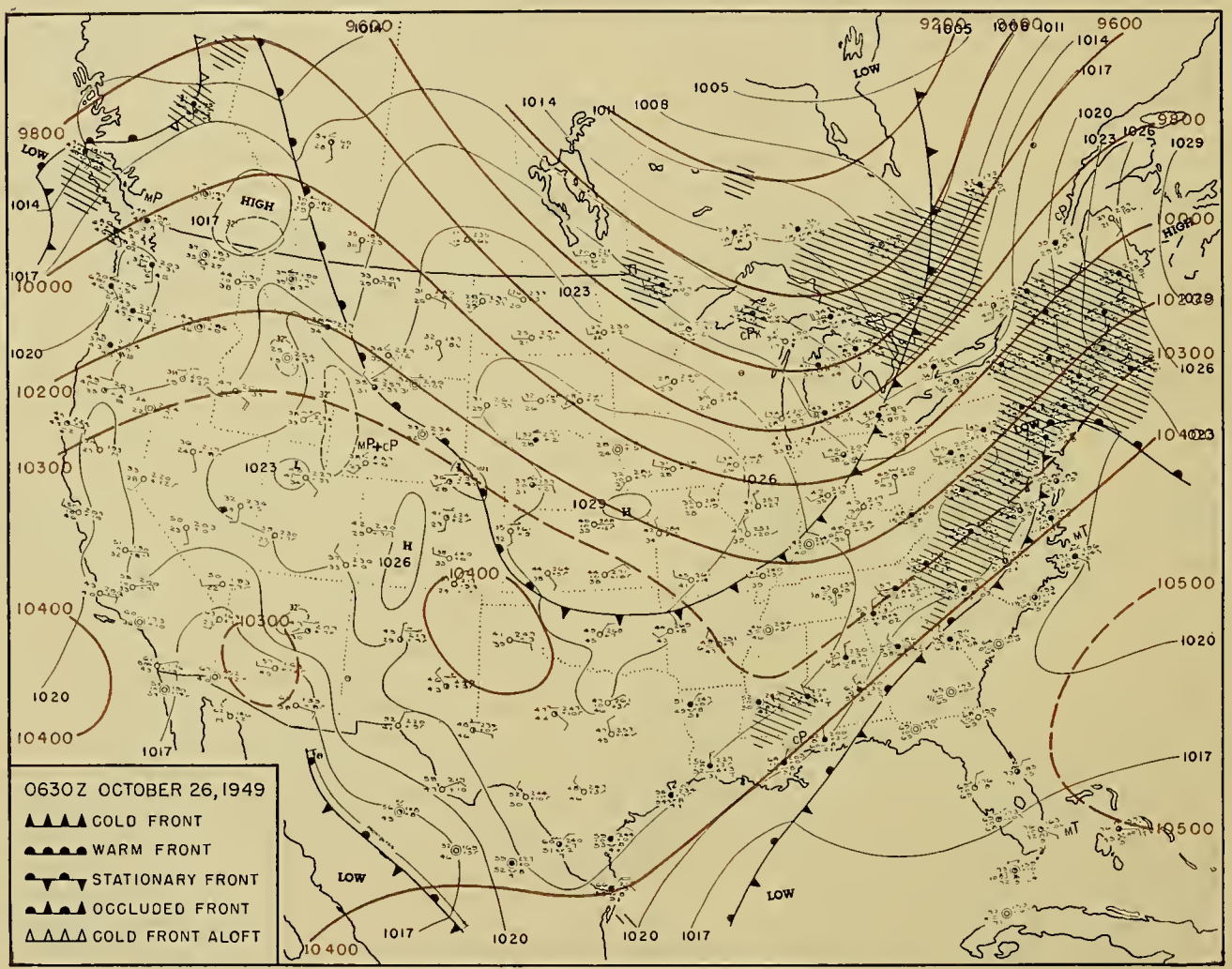


FIG. 4.—Combined sea-level and 700-mb analysis for the United States at 0630Z, October 26, 1949.

Although small-size, large-area maps are needed by the forecaster for predictions of two to three days or longer, the analyst who is responsible for a detailed analysis of a limited portion of the earth's surface, or the forecaster concerned with a local forecast, usually needs a different map base. This latter base must have a scale large enough so that the data from all the observational stations in the region may be conveniently represented on the map, since detailed data are necessary for the analyst to be able to locate pressure jumps, nascent waves on fronts, and poorly marked fronts.

Detailed analysis and study is likewise a requisite for short-period local forecasts. This whole question of the selection of a meteorological base map is not one which can be solved theoretically once and for all time; it is a recurrent practical problem, whose solution is contingent upon the specific conditions under which the map will be used. This specificity, which is characteristic of so many meteorological perplexities, renders the problem by no means unimportant.

Moisture Charts

Another equally specific, equally significant, difficulty is that of devising a chart, or charts, which will portray graphically those factors which are conducive to showers other than those associated with fronts or pressure jumps. Such a chart (or charts) would be the *sine qua non* of local forecasting, particularly in regions of high summertime precipitation. In an attempt to solve this problem, sundry types of energy diagrams have been developed along with techniques for computing the likelihood of showers. Empirical results have been disappointing to such an extent that it appears probable that our present methods for evaluating the energy due to thermal instability fail to take into account all the factors involved in the release of energy as shower activity. This is corroborated by the recent studies on entrainment of air into the sides of increscent cumuliform clouds. The findings of these investigations suggest that the occurrence of showers is not often due to pure thermal instability, pure air-mass showers being very rare. Indeed, evidence has been disclosed that a widespread convergent wind flow, a pressure jump, a mountain range, or some other type of trigger action is necessary to give well-developed shower activity. The recent studies of the thundershower activity in Florida and its relation to the convergent effect of the sea breezes from two edges of the peninsula afford a good illustration of how important such effects are on air-mass shower activity [7]. Certainly, this is a subject on which more research is sorely needed.

Although our knowledge of the relationship between the vertical energy distribution and subsequent shower activity is imperfect at present, there appears to be a more readily applicable relationship between the horizontal moisture distribution and the resultant showers. In the days before there were reliable upper-air soundings, an analysis of clouds and precipitation was the main clue to the distribution of moisture. But after the establishment of the network of suitably accurate radiosondes, many experiments were made to determine the type of chart (or charts) which would be best suited for portraying the horizontal distribution of moisture in the atmosphere and for following its successive variations.

At the present time the horizontal distribution of atmospheric moisture is usually studied by means of isopleths of dew point on either the 850-mb or 700-mb chart, or on both. The basic disadvantage of this procedure is that air particles frequently undergo vertical motions rendering it impossible to be sure that one is following from moment to moment the same parcels of air. This, of course, makes various invalid inferences very tempting. But even with the assumption of quasi-horizontal motion, this method for the analysis of atmospheric moisture is far from being completely satisfactory. Its value is impaired by the fact that the patterns produced by the dew-point isopleths tend to be indistinct and fragmentary on both the 850- and 700-mb charts. Through the study, however, of these particular isopleths, considerable valuable information may be obtained.

There is evidence that it is in the plateau and mountain regions of western North America that the most fruitful application of this procedure for moisture analysis can be made, a somewhat surprising disclosure since most analytical methods give their best results over oceans or over plains of low elevation. In the areas of these western plateaus and mountains we find that the 700-mb configuration of isopleths of dew point seem closely correlated to shower precipitation all during the warm season. In fact, many meteorologists feel that the wind and moisture patterns at 700 mb may be used alone to give satisfactory forecasts for showers in these mountain areas. Over relatively flat areas where the trigger action of mountains is lacking, more information is needed for the prediction of showers. In particular, it would be of benefit to know the location of areas where upslope or downslope motion of moist parcels of air is occurring.

Isentropic Analysis

This need for a more inclusive chart, designed to depict not only the distribution of atmospheric motion but also proximity of each air parcel to saturation, the location of areas where there are upslope or downslope movements, and the major flow patterns of the air, is met by the isentropic chart [11]. Besides its important property of inclusiveness, the isentropic chart is the chart best suited for following the motion of air particles since the particles of an isentropic surface are conservative for all adiabatic changes. The patterns formed by lines of equal condensation height, or pressure, on an isentropic chart are therefore less fragmentary and more persistent than the configurations formed by the isopleths of dew point on a surface of constant pressure. In short, on an isentropic chart the distribution of moisture is simply and vividly apparent and the analyst can, in general, validly follow a particular air particle in its day-to-day convolutions. Thus the isentropic chart is a potent addition to our tools for preparing optimum analyses.

Shortly after the upper-air network had become well established, the isentropic chart was introduced and tried out in all the meteorological centers of the United States. At that time it was the practice to transmit to the field stations, along with the other upper-air data, the data for three different constant potential temperature surfaces. The purpose of transmitting the data for three isentropic surfaces was to allow the local meteorologist to select for analysis the one isentropic surface most useful to his own locale.

During this period of experimentation the most im-

portant application of the isentropic chart proved to be its use in forecasting cloudiness and precipitation, including the evasive summertime showers. The chart was also found to be of especial value in predicting the occurrence of precipitation along the West Coast [6], where frontal analysis taken by itself is inadequate for this purpose.

Unfortunately the years of experimentation drew to a close with only a small portion of the meteorologists in the United States familiar with the applications of the isentropic chart. Subsequently the isentropic chart was abandoned by the Weather Bureau and the data for its construction were no longer transmitted over the teletype network. This was due in part to certain difficulties in communication, but to a larger extent it was due to the time factor inherent in the analysis itself. The isentropic chart is more difficult to draw correctly than are most of our other upper-level charts. For inexperienced men the analytical time consumption proved to be an unsurmountable obstacle to the realization of the usefulness of the chart. Experienced forecasters and analysts, however, who had become familiar with the isentropic chart generally felt, and still feel, that a long step backwards was taken when the isentropic analysis was abandoned.

With the present trend towards the concentration of analysis in large centers, with the increased accuracy of radiosondes, and with the advent of improved communication facilities, the time seems ripe for the reinauguration of the isentropic chart. It would be entirely possible for the analysis center to transmit to field forecast centers three complete isentropic charts, based on the same data that were previously transmitted from the observational network. With the advent of a means for transmitting colors, these isentropic charts could include not only streamlines, some sort of contour lines, and condensation height or pressure lines, but they could include, as well, moist tongues shaded in red, dry tongues in blue, and saturated areas in green in conformance with the conventions observed previously in the construction of isentropic charts. Thus the forecaster would be provided with a simple efficacious tool for studying the moisture distribution of the atmosphere. Moreover, instead of having one "homemade" isentropic chart, plus the raw data for two other constant potential temperature surfaces, he would be equipped with carefully executed analyses for all three surfaces, enabling him to investigate the changes with respect to height of all the data on the isentropic chart.

Such an investigation may be facilely accomplished through the use of another form of isentropic analysis devised by Starr [21]: the relative-motion chart. The aim of such a chart is to supplement the information given by streamlines at a single isentropic level. The relation between streamlines and contour patterns indicates whether there is upslope or downslope motion on the isentropic surfaces provided the wind near the isentropic level in question increases with height, as is generally the case. The function of the relative-motion chart is to give an approximation of the change in wind with height at the isentropic level being considered. Relative-motion charts are easily prepared, given the isentropic data. They could be drawn and transmitted from analysis centers as an additional tool for the use of the local forecaster. Such arrangements would benefit not only the synoptic meteorologist, but the research man as well, an important consideration since in the field of research, isentropic analysis is acknowledgedly of significant value.

Conclusion

In various specific contexts in this discussion of analytical techniques we have mentioned the value of the use of color in augmenting the clarity of analyses. Despite the habitual use of color on meteorological charts, its utilization has tended to be haphazard and accidental rather than systematized. To a surprising extent even simple uses of color have been neglected in the United States. The psychological effect of color and its use as an aid to analysis should in the future receive the attention it warrants.

Another neglected problem is that of the ideal arrangement for an analysis center or forecast station. As far as the authors can discover, the results of comprehensive studies of this subject have not been incorporated into actual practice. The optimum height of working surfaces, lighting, best selection of furniture, and the arrangements of furnishings and charts are considerations which lie outside the field of meteorology proper. But seemingly minor factors such as these (or such as the use of color mentioned above) may contribute appreciably toward better analysis. In order to improve forecasts, these problems must be solved along with those of a more strictly meteorological nature.

But beneath these specific analytical problems, and precluding their solutions, lie the broader barriers to rapid advancement in the field of meteorology as a whole. One of these fundamental obstacles is the confusion occasioned by the plethora of theories and operational practices, attending the sudden multiplication of reliable meteorological instruments, which has presented for the scrutiny of the forecaster an entirely new and unprecedented multiplicity of data. The other basic obstacle is the enigma of the processes involved in the general circulation of the atmosphere. Once a valid theory of this general circulation has crystallized, order should begin to appear in the present chaos of our secondary meteorological theories. However, until such time as this is accomplished, it behooves those of us concerned with analysis to follow up whatever lines of research seem promising, searching through meteorological theory for effective applications to analysis and seeking to better our analytical techniques through experimentation and critical discussions participated in by meteorologists all over the world. But despite our concern with the details of analysis, we must never lose sight of their dependent relationship to the larger and more fundamental concepts of meteorology.

REFERENCES

1. Austin, J. M., "Temperature Advection and Pressure Changes." *J. Meteor.*, 6:358–360 (1949).

2. Bellamy, J. C., "The Use of Pressure Altitude and Altimeter Corrections in Meteorology." *J. Meteor.*, 2:1–79 (1945).
3. —— *A Proposed System for Obtaining Graphical Representation.* Progress Reps. Nos. 1 to 4, Cook Res. Lab., Cook Electric Co., Chicago, 1948.
4. Bjerknes, J. "On the Structure of Moving Cyclones." *Geofys. Publ.*, Vol. 1, No. 2 (1918).
5. Brier, G. W., *Predicting the Occurrence of Winter Time Precipitation for Washington, D. C.* U. S. Weather Bureau, Washington, D. C., 1945.
6. Byers, H. R., *General Meteorology*, 2nd ed. New York, McGraw, 1944.
7. —— and Rodebush, H. R., "Causes of Thunderstorms of the Florida Peninsula." *J. Meteor.*, 5:275–280 (1948).
8. Freeman, J. C., Jr., "An Analogy between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5:138–146 (1948).
9. Harrison, H. T., and Orendorff, W. K., "Pre-coldfrontal Squall Lines." *United Air Lines Meteor. Dept. Circ.* No. 16 (1941).
10. Hovmöller, E., "The Trough-and-Ridge Diagram." *Tellus*, Vol. 1, No. 2, pp. 62–66 (1949).
11. Namias, J., *An Introduction to the Study of Air Mass Analysis*, 1st ed. Milton, Mass., Amer. Meteor. Soc., 1935.
12. —— *Extended Forecasting by Mean Circulation Methods.* U. S. Weather Bureau, Washington, D. C., 1947.
13. Oliver, V. J., and Oliver, M. B., "Forecasting the Weather with the Aid of Upper-Air Data" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, N. R. Beers, ed., pp. 813–857. New York, McGraw, 1945.
14. —— "Weather Analysis from Single Station Data." *Ibid.*, pp. 858–879.
15. Panofsky, H. A., "Objective Weather Map Analysis." *J. Meteor.*, 6:386–392 (1949).
16. Petterssen, S., *Upper Air Charts and Analysis.* NAVAER 50-IR-148, Washington, D. C., 1944.
17. Rapp, R. R., "On Forecasting Winter Precipitation Amounts at Washington, D. C." *Mon. Wea. Rev. Wash.*, 77:251–256 (1949).
18. Riehl, H., "On the Formation of Typhoons." *J. Meteor.*, 5:247–264 (1948).
19. —— and LaSeur, N., "A Study of High-Tropospheric Lapse Rates with Application to the Construction of 300-Millibar Charts." *J. Meteor.*, 6:420–425 (1949).
20. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose.* Berlin, Springer, 1948.
21. Starr, V. P., "The Construction of Isentropic Relative Motion Charts." *Bull. Amer. meteor. Soc.*, 21:236–239 (1940).
22. Tepper, M., "A Proposed Mechanism for Squall Lines: The Pressure Jump Line." *J. Meteor.*, 7:21–29 (1950).
23. Wexler, R., "Radar Detection of a Frontal Storm 18 June 1946." *J. Meteor.*, 4:38–44 (1947).
24. Williams, D. T., "A Surface Micro-study of Squall-Line Thunderstorms." *Mon. Wea. Rev. Wash.*, 76:239–246 (1948).

WEATHER FORECASTING

The Forecast Problem *by H. C. Willett* 731

Short-Range Weather Forecasting *by Gordon E. Dunn* 747

A Procedure of Short-Range Weather Forecasting *by Robert C. Bundgaard* 766

Objective Weather Forecasting *by R. A. Allen and E. M. Vernon* 796

General Aspects of Extended-Range Forecasting *by Jerome Namias* 802

Extended-Range Weather Forecasting *by Franz Baur* 814

Extended-Range Forecasting by Weather Types *by Robert D. Elliott* 834

Verification of Weather Forecasts *by Glenn W. Brier and Roger A. Allen* 841

Application of Statistical Methods to Weather Forecasting *by George P. Wadsworth* 849

THE FORECAST PROBLEM

By H. C. WILLETT

Massachusetts Institute of Technology

INTRODUCTORY REMARKS

The Unsatisfactory Progress of Weather Forecasting as a Science. Probably there is no other field of applied science in which so much money has been spent to effect so little real progress as in weather forecasting. Today the nations of the world spend many times what they did forty years ago to obtain the necessary observational data and to prepare weather forecasts. But the advance in weather-forecasting skill has not kept pace with the increased effort and attention devoted to forecasting. The variety of weather forecasts, as to time range and detail, as to the elements forecast, and as to the elevations and geographical areas for which forecasts are prepared, has been greatly increased in proportion to the increase in the observational synoptic data, to meet more exacting demands. Much has been done to meet these demands by the development of special types of forecasts, frequently forecasts of elements which were not even observed forty years ago. But in spite of all this great expansion of forecasting activity, there has been little or no real progress made during the past forty years in the verification skill of the original basic type of regional forecast, of rain or shine and of warmer or colder on the morrow, the kind of forecasting which first received attention.

The question naturally arises, Why has this great expansion of forecasting activity contributed so little to the basic advance of the science? There are a number of factors to which this lack of progress may be attributed. These factors vary somewhat in relative importance from country to country, but probably conditions in the United States (which has experienced perhaps the greatest multiplication of forecasting tools and techniques) may be taken as essentially typical of the entire field. On this assumption the principal reasons for the failure of forecast skill to keep pace with forecast practice are:

1. The expanding demand for trained forecasters. Since World War I, but increasingly during the thirties and with the outbreak of World War II, there has been a vast expansion in the number, the variety of service, and the trained-personnel requirements of weather-forecast centers. This greatly increased demand for forecasters has come from government weather bureaus, from commercial aviation, and from the military. Since the vastly increased demand for trained forecasters was not anticipated, the proper training of large numbers of forecasters was not provided for. Neither has sufficient salary incentive been provided (in proportion to the exacting nature of the work) to attract into this field, or to hold, the best-qualified men. As a result, thousands of meteorologists have been pressed into forecasting service, many of them after only one or two years of training, and many of them unqualified by interest or temperament for this work. One further circumstance which has been most unfavorable to the improvement of forecasting is that by and large there has been in operation no objective verification procedure by which the forecasting ability of these many inexperienced forecasters could be compared one with another or against any standard, as a basis for the elimination of the less competent forecaster.

As recently as twenty years ago in the United States the few official government forecasters were assigned to this duty only after many years of practice forecasting in objectively verified competition with the official and other practice forecasters. After assignment to forecast duty their records as official forecasters were continuously checked in the same manner. It is a sad commentary on the scientific status of weather forecasting that even today the methods remain so empirical and so dependent upon experience and subjective interpretation that the development of the best forecasters still requires years of experience and the right temperament and interest. The primary problem of weather forecasting remains that of removing the science from this subjectively empirical category and of making it scientifically objective.

2. Multiplicity of forecasting tools and techniques. The availability of rapidly increasing amounts of observational synoptic data, particularly of aerological data, has led to a great amount of experimentation with the use of various charts and coordinate systems for the synoptic presentation and analysis of the new observational material as an aid to the forecaster. Unfortunately this has led to the introduction into forecasting practice of a multiplicity of charts and diagrams which are largely redundant in that they present the same basic information in a variety of forms, the relative merits of which remain completely untested, and the variety of which leads only to a confusion of forecasting techniques and principles that interferes with the clear-cut formulation of any simple set of prognostic rules or criteria, such as is gained from long familiarity and practice with one minimum standard set of synoptic charts. At the present time routine forecast practice can undoubtedly be greatly benefited by the universal acceptance of a minimum standard set of simple synoptic charts, even though the selection of the standard aerological charts might not be the most effective that could be made. This selection of standard charts should be based on some real attempt at objective verification of their relative prognostic merits, a procedure which is equally necessary for forecasting techniques and for forecasters.

3. Failure to assess the essential forecast problem correctly. Probably the primary reason that more prog-

ress has not been made in recent years towards the improvement of basic weather forecasting lies in the failure to assess the essential nature of the problem correctly, with the consequent misdirection of much money and effort. Practical weather forecasting has been until very recently so exclusively a matter of extrapolation into the near future of current weather and weather trend, on the basis of synoptic experience, empirical rules, and statistical probabilities, that most thinking about weather forecasting is patterned in the same mold. Great amounts of money and effort have been expended in increasing the number and content of synoptic observations, in experimenting with new forms of synoptic representation, and in statistical or synoptic analysis of the data. But most of this effort has been expended without any planned attack on the basic problems of meteorology in the vain hope of finding short-cut empirical or synoptic forecast rules or statistical relationships which might radically improve weather forecasting. Forecasting progress has not been realized in proportion to this misdirected effort, and possibly it has even been hampered to some extent by a surplus of information which is poorly planned and inefficiently utilized.

The Scope and Essential Nature of the Forecast Problem. The ultimate practical goal of most meteorological research is to improve present knowledge and understanding of atmospheric processes to the point where accurate scientific weather forecasting becomes possible. The all-inclusive scope of the problem is evident from its essential character, comprising as it does primarily three questions, each of which must be answered in turn. They may be stated, in their simplest terms, as follows:

1. What is the weather? This question requires a knowledge of the distribution of the meteorological elements which constitute the weather and whose changes reflect the weather processes. Such knowledge is required both as the starting point or base from which to forecast the future weather, and as a means of verifying or evaluating past weather forecasts and forecasting techniques. All research bearing on the measurement and recording of the weather elements throughout the atmosphere is directed toward answering this question. It is in this technical field that meteorological research has been advanced out of all proportion to the utilization of the data which are obtained.

2. Why is the weather? This question requires the quantitative physical explanation, in scientific terms, of all weather phenomena and their causative or formative processes. This is the basic part of the forecast problem which weather-forecasting research has tended to neglect in the past, in comparison to the money and effort which have been expended upon the unsystematic accumulation and the routine synoptic or statistical analysis of observational data. It is, however, primarily on progress in seeking an answer to this question that the future improvement of weather forecasting depends.

3. What will the weather be? The right answer to this question constitutes correct weather forecasting. A satisfactory answer requires a comprehensive understanding both of what the weather is and why the weather is. It has been the attempt to bypass the second question, in the hope of finding short-cut empirical synoptic or statistical means of determining the future weather directly from the present weather, without understanding either, that has militated against progress in solving the basic forecast problem.

THE PRESENT PRACTICE AND PERFORMANCE OF FORECASTING TECHNIQUES

Types of Weather Forecasts—Time Range, Content, and Performance. Since most weather-forecasting practice consists essentially of the rather crude extrapolation of current weather patterns and tendencies into the future, it is axiomatic that the accuracy and justifiable detail of the forecasts decrease rapidly with increasing range. Consequently weather forecasts are prepared in a great variety of form and detail, depending upon the degree of detail and accuracy which is required by the specific purpose of the forecast. The forecasting techniques which are used vary greatly with the type (time range and detail) of the forecast. Therefore to discuss the practice and performance of forecasting techniques, it is really necessary to consider briefly the requirements and expected accuracy of the different types of forecasts. It must be emphasized, however, that the complete lack of any uniform objective verification of weather forecasts precludes the possibility of any reliable evaluation or comparison of the performance of the different forecast types and the respective forecasting techniques. All estimates of forecasting skill in the following discussion are expressed on a basis of 50 per cent verification by chance as representing pure guess work or zero skill. For the most part the figures represent rough comparative estimates of forecasting skill and are not based on actual extended series of numerical verification of forecasts.

Weather forecasts are most conveniently classified according to time range and basic character of the forecast in the following four categories:

1. Short-range forecasts, for periods up to eighteen hours from the issue of the forecast. Forecasts of this type have been developed almost entirely during the past twenty-five years, primarily in response to the demands of aviation, both civil and military, although short-range forecasts of local conditions of frost, snow accumulation, and icing of roads have also received increasing attention. Short-range airways forecasts require a high degree of accuracy, both in timing and in local detail of all elements affecting flight operations, particularly of terminal conditions. These elements include wind direction and speed up to an elevation of from ten to fifteen thousand feet, the state of turbulence of the atmosphere, ceiling heights, horizontal visibility, and the occurrence of condensation forms (notably when the danger of icing, fog, or thunderstorms is present). Because this is a relatively new type of forecasting, frequently of elements which previously were not forecast at all, because the usual extrapolation techniques are particularly suited to short-range forecasting, and because great effort has been expended in meeting the

sudden demand for accurate forecasts of this type, this is the one field of forecasting in which great improvement in forecasting skill has been demonstrated in the past twenty years. Although no verification figures for this type of forecasting are available, it is probable that the best forecasting organizations in this field perform in the 90–95 per cent range on a 50 per cent probability basis of verification. This performance does not represent any basic advance in weather forecasting, but merely the happy results of a sudden concentration of effort on the application of extrapolation techniques to very short-range developments.

2. Forecasts of the ordinary daily range, for periods of from twelve to forty-eight hours in advance. Forecasts of this type usually express in some detail, for specific geographical areas, the expected day-to-day sequence of all the aspects of weather which materially affect human activity and well-being. The forecasts are disseminated by press, radio, wire, and special warning systems to the public and to many special interests, both civil and military, including aviation, agriculture, shipping, manufacturing, merchandising, and utilities. It is this type of forecasting on which the greatest effort has been concentrated since synoptic forecast methods were first introduced nearly a century ago and in which, paradoxically, basic progress has been almost nonexistent in recent years. Forecasting activity has been vastly expanded, but the quality of the forecasts has not been basically improved. The accuracy and skill of this type of forecast as practiced by the best forecasters must diminish with increasing time range during the 12- to 48-hr period. The skill verification may be estimated to decrease through the range 90–70 per cent during the period.

3. Extended daily forecasts, for periods including from the second to the sixth or seventh day in advance. Forecasts of this type at the present time tend to assume a dual character, a combination of the ordinary daily forecasts on the one hand, and of the long-range forecasts on the other. In the first role they strive to extend to five or seven days in advance the same detailed forecast of the day-to-day sequence of weather that is offered by the ordinary daily forecasts up to two days in advance, while in the second role they specify the mean or average character of the 5- or 7-day period ahead in terms of the expected anomalies of temperature, precipitation, and possibly other elements, such as sunshine. In the first role these forecasts are expected to serve the day-to-day planning of the entire life of the community in the same manner as, only more in advance than, the daily forecasts, while in the second role they are useful primarily to agriculture, irrigation, flood control and power interests, and to the fuel industry. Forecasts of the extended daily variety have been issued for many years, but the principal recent advances in extended weather forecasting have come in the development of statistical-synoptic techniques for the forecasting of large-scale weather patterns and mean weather conditions. The forecasting of the day-to-day weather sequence shows negligible skill beyond the fourth day in advance, probably decreasing slightly in the 55–50 per cent range. It can be stated categorically that no claim to significant skill in forecasting the day-to-day sequence of weather more than five days in advance has been acceptably demonstrated, nor is it likely to be in the near future. The mean anomaly features of the more reliable extended weather forecasts today verify in the 70–75 per cent range for temperature, but little better than 60 per cent for precipitation. Precipitation, relative to temperature, becomes increasingly difficult to forecast as the time range is extended.

4. Long-range forecasts, for all periods extending beyond one week in advance. Forecasts of this type can deal justifiably only with mean conditions or period anomalies, usually of pressure, temperature, or precipitation. There is no legitimate basis whatsoever for extending forecasts of the day-to-day weather sequence more than one week in advance, frequent unverified claims to the contrary notwithstanding. Forecasts of the longer-period anomalies of temperature and precipitation, whether for weeks, months, seasons, or whole years, can be of very considerable value, insofar as they have real merit, for a wide range of interests, notably for planning effective agricultural production or merchandising campaigns, for the effective utilization of irrigation, hydroelectric and flood control facilities, and for long-range military planning. Unfortunately, however, no forecasts of this type have demonstrated any clearly significant degree of skill in verification. Probably under most favorable conditions the verification lies in the 55–60 per cent range.

Forecasting Techniques or Aids—Their Merits and Limitations. The following definition and evaluation of weather-forecasting techniques as they are practiced today can be only approximate at best, because of the multiplicity of techniques which are in use, the infinite variety of combinations in which they are applied by different forecasters for different and even for the same type of forecast, and the lack of any real effort towards the objective evaluation of these numerous forecasting aids. In the following discussion an effort is made to classify the individual forecasting techniques or aids, insofar as they are distinctive, as to their essential character; to indicate something as to their usual application in the four principal types of forecasts (time range); and to comment qualitatively on their merits and limitations.

Since the following classification of forecasting techniques is based on the essential character of the technique, rather than on the forecast type (time range) to which it is applied, and since the practice of any type of forecasting frequently involves the combination of a variety of techniques, the grouping by this classification is suited primarily to an evaluation of the techniques individually rather than in combination. It is quite impossible in a general discussion of this kind even to enumerate the many combinations in which these forecasting techniques are applied, much less to evaluate these combinations. But in general the skill verification that is obtained by combining forecasting techniques is of the same general order as that of the

best single technique. If the techniques are essentially different, then when their prognostic indications are inconsistent, the forecaster's judgment must enter into any selection between them. That judgment is likely to be prejudiced in favor of the one or two techniques which are believed best on the basis of past performance.

In the following classification of forecasting techniques all of those classified as *extrapolation* in type employ exclusively the current and past weather patterns, including usually trend and in a few cases acceleration patterns (first- and second-order time derivatives of the weather elements). Since most of these extrapolation techniques are employed for forecasts of more than one type (time range) they are classified by their essential nature, rather than by the type of forecast for which they are used, as *statistical* (dealing with the time sequences of weather in limited areas), *synoptic* (dealing with geographical weather patterns), or *mathematical* (dealing with the fundamental equations of motion, continuity, and state).

In contrast to the extrapolation techniques of forecasting, those techniques classified as *physical* are concerned primarily with conditions which are expected to impose modifying influences on the future development of the current weather pattern, influences which are not in evidence in the past development, hence cannot be considered in any procedure involving extrapolation only. These physical processes are further subdivided between those whose effects are specifically related only to the current weather pattern, hence are of significance only to the short-range or daily type of forecast, and those which are more permanent in their influence, affecting any weather pattern for some time to come, and which are therefore of primary significance for the extended or long-range type of forecast.

1. *Extrapolation Techniques*—based on the existing state and trend of the weather, exclusive of modifying physical factors.

a. Statistical techniques of extrapolation—based on the time sequence of change and association of the weather elements individually in restricted localities. The purely statistical, as opposed to synoptic, forecast aids are usually applied to the long-range forecasting of the mean weekly, monthly, seasonal, or even annual anomalies of pressure, temperature, or precipitation.

Statistical methods are occasionally applied to daily forecasts, usually to supplement the empirical climatological information of a forecaster faced with the problem of forecasting for a region with which he has had relatively little experience. For this purpose statistics can be prepared, on the basis of past records, to indicate the probability of any meteorological occurrence, or combination of occurrences, to assist the forecaster in choosing between alternative possibilities, or to decide the degree of severity which an anticipated condition is likely to attain. However, by virtue of long experience in forecasting for a given region the forecaster usually develops a strong sense of the probability of any weather occurrence in his district, and his judgment of the potentialities of a given synoptic situation is likely to stand him in far better stead in estimating its probable development than are the statistical probabilities. Consequently, purely statistical methods can make at best only a very dubious contribution to forecasting of the short-range or daily type, particularly in middle and higher latitudes where the weather is so erratically variable from day to day. Although climatology and the statistical probability of both time and space sequences of weather change may well be developed further for the guidance of the forecaster, it can be assumed with reasonable certainty that no amount of effort expended in this direction will lead to any radical improvement of the shorter-range types of forecast.

For the extended and long-range types of weather forecasting the statistical extrapolation techniques assume greater relative importance with increasing range, as their applicability and performance are relatively independent of the time interval to which they are applied, whereas the usual short-range synoptic extrapolation techniques rapidly become ineffective as the time range is extended. The statistical extrapolation techniques which are applied to the longer-range forecasting of mean pressure, temperature, and precipitation anomalies at selected points are quite simple in principle, being based essentially on persistence and trend, linear lag-correlation (regression), and periodic analysis. It is quite obvious that none of those statistical extrapolation techniques can be expected to forecast the exceptional or unusual weather occurrences that are of greatest practical importance, because they are all keyed to the most probable occurrence.

Pure statistical extrapolation techniques are identical, whether the length of the unit period involved is the day, the week, the month, the season, or the year. Linear regression equations, based on linear correlation, may be computed from past records to give the most probable numerical anomaly of any meteorological element for any selected station or district, on the basis of the values of the preceding one, two, or any arbitrarily selected number of periods. The simplest example of this technique is to compute the probability of simple persistence or reversal of the sign of an anomaly from one calendar month to the next. Usually such lag correlations are low, but occasionally in restricted areas and for certain seasons the correlation is high enough to have real forecasting significance. This method of statistical extrapolation is assisted by the fact that in many regions there exists high contemporary seasonal correlation between pressure, temperature, and rainfall, so that the forecast anomalies of these elements must be mutually interdependent in a meteorologically consistent manner.

An elaboration of the lag-correlation technique of statistical extrapolation is furnished by the use of lag correlation between current or past weather anomalies in one region, or in one branch of the general circulation, and subsequent anomalies in another region or another branch. This technique, which is capable of almost unlimited experimental application, has been used by many investigators, but it is only occasionally that

really significant correlations are found. It has been used widely by Baur [1] in Europe, by Wadsworth [16] to check contemporary and lag relationships between the positions and intensities of a number of the semi-permanent "centers of action" of the monthly mean Northern Hemisphere pressure maps, and most notably by Walker[1] in his classical study of weather relationships over the globe and in particular of the forecasting of the seasonal monsoon rainfall in India.

It is doubtful whether the highest degree of skill attained at the present time by forecasts based on the use of statistical extrapolation by correlation (as illustrated perhaps by some of Walker's seasonal forecasting by multiple regression equations derived from the Southern Oscillation) exceeds 60 per cent. This figure probably represents the best that is attainable by purely statistical extrapolation techniques in the absence of any physical understanding of the underlying cause and effect relationships.

Experience shows that without some real physical basis, the prognostic regression equations derived by random correlation almost invariably deteriorate significantly in their performance when applied to the forecasting of data other than that from which they were derived. This deterioration is doubtless caused in part by the fact that the few significant correlation coefficients which are used in the regression equations are usually selected from a relatively large number of predominantly insignificant coefficients computed on a random "trial and error" basis. Hence the selected coefficients do not possess the normal probability of significance. The deterioration is also partly caused by the fact, which is observed both statistically and synoptically, that not only the general circulation pattern itself, but also its pattern of variation, is subject to considerable change over a long period of years.

It is undoubtedly true that the area in which this technique of forecasting by statistical extrapolation may be applied can be extended by further statistical research, but it is doubtful if the present top level of performance will be exceeded appreciably. In other words, there is at present no reason to hope for any basic improvement of long-range weather forecasts by the routine statistical extrapolation technique under discussion.

A second technique of long-range forecasting by statistical extrapolation is based on the statistical analysis of long-term homogeneous records of pressure, and less frequently of temperature or precipitation, at selected stations, in the effort to establish some type of periodicity or repetition in the variation that can be used for extrapolation. A large amount of harmonic analysis or other techniques of smoothing of such records has been done in order to resolve the variation into simple component periods. A great variety of such periods, ranging in length from a few days to at least forty-six years, have been rather dubiously established. In particular the longer periods have been tentatively established as fractional or integral multiples of the sunspot period, notably by Abbot and Clayton. Another purported feature of the long-period pressure records at certain stations, first pointed out by Weickmann, is the occurrence at irregular intervals of so-called *symmetry points*, on either side of which the time graph of pressure varies in reverse sequence.

There is considerable doubt as to the physical reality or statistical significance of most of these periodic or symmetry characteristics which are supposedly established in time series of meteorological data. It has been shown that in some cases smoothing techniques impose on the data periodicities which do not exist in reality. Forecasts based on the periodicity and symmetry techniques of statistical extrapolation have proved significantly even less successful than those based on the correlation techniques discussed above. There is no evidence to indicate that a skill verification of even as much as 55 per cent has been attained by the periodicity techniques. Baur [2] has made an extensive statistical study of the periodicity and symmetry techniques of extrapolating pressure and other weather data for extended forecasting purposes. He came to the conclusion that these techniques do not merit the expenditure of additional effort. He produces statistical evidence that the great variety of periodicity and symmetry patterns which are derived statistically from time series of weather data probably are merely an incidental result of serial correlation in the data, and as such are presumably of no greater prognostic significance than the persistence itself. Certainly it can be concluded that no important improvement of weather forecasting is to be expected from further development of this technique of statistical extrapolation.

b. Synoptic techniques of extrapolation—based essentially on the extrapolation of synoptic weather patterns. The great bulk of weather-forecasting practice, both past and present, is of the daily, and more recently also of the short-range type. This common forecasting practice is usually based on a combination of techniques, rarely on one exclusively. During the earlier years the forecasting procedures were almost exclusively those of synoptic extrapolation. Since World War I the physical techniques discussed below (p. 739) have come into extensive use, and in recent years the techniques of mathematical extrapolation, discussed under *c* below, have also been utilized to a very limited extent to supplement other methods. But by and large it remains true today that common forecast practice is based primarily on methods of synoptic extrapolation, secondarily on physical considerations, and only slightly on purely statistical or mathematical extrapolation. It is this combination of synoptic-extrapolation techniques supplemented by certain physical considerations which currently effect a top skill verification of 90–95 per cent on short-range forecasting, for which extrapolation techniques are ideally suited, and of 70–90 per cent on the daily type of forecasting up to forty-eight hours in advance.

Except for the occasional use of statistical probabilities to compensate a forecaster's lack of practical experience, as mentioned previously, the entire short-

1. See *Mon. Wea. Rev. Wash.*, Supp. No. 39, pp. 1–26, 1940.

range and daily forecasting routine as commonly practiced today depends upon the effective synoptic presentation of current and past weather patterns, which serve as a basis of either synoptic or mathematical extrapolation, and of any physical forecasting techniques. These synoptic weather patterns usually present the two-dimensional distribution of the weather elements in selected planes. The primary patterns are those of the geographical distribution of the weather elements in horizontal or quasi-horizontal planes. A great variety of such weather charts may be plotted, starting usually with the sea-level or surface map. The upper-level charts may be plotted for selected heights, for selected isobaric surfaces, for selected isentropic surfaces, or for the thickness between selected isobaric or isentropic surfaces. Several upper-level charts are usually prepared, the highest one frequently being located near the top of the troposphere or the base of the stratosphere.

The synoptic charts for the selected levels are frequently supplemented by vertical cross sections through the atmosphere, which follow a line of upper-level observation stations, or at times by the use of individual point soundings through the troposphere. The essential purpose of the entire assemblage of synoptic charts, which, particularly at upper levels, may vary greatly as to number, form, elevation, weather data plotted, and geographical extent of the area covered, is to present as completely and effectively as practicable, in the area which is deemed necessary, the distribution of the weather elements in the form of two-dimensional patterns. It is the regular time sequence of a complete set of such synoptic charts, particularly the sequence of pressure and pressure-change patterns, to which are applied the various techniques of synoptic and mathematical extrapolation, and physical reasoning, by means of which the future, or prognostic, weather pattern is derived.

It is possible to enumerate here only the more frequently applied synoptic techniques of extrapolation, and to comment briefly on their potential effectiveness. This type of forecasting has been practiced on surface weather maps since the beginning of synoptic weather forecasting nearly a century ago, while its application to upper-level charts is a development almost exclusively of the past quarter-century, except that upper-level wind observations have been used to some extent to supplement surface reports since the early 1900's.

The simplest extrapolation procedure consists of the linear extrapolation, or continued displacement at the same speed as that in the preceding period or periods, of the various features of the sea-level weather map, such as isobaric or isallobaric centers, isobaric trough or ridge lines, or frontal discontinuities with the attendant frontal phenomena and air-mass weather characteristics. The linear extrapolation technique is applied also to change or acceleration tendencies such as the deepening or filling of pressure centers and troughs or ridges, changes in the direction or speed of movement of pressure patterns, fronts and air masses, and frontogenesis or frontolysis. The approximate effect on the pressure field of even the second-order time derivatives of the pressure may be anticipated by the deepening or filling of isallobaric centers, and by the proper use of 3-hr pressure tendencies corrected for the normal diurnal trend of pressure.

Since the upper-level pressure patterns are uniquely related to the sea-level pattern and the temperature distribution, and since the upper-level charts present the large-scale dynamic state of the atmosphere more clearly than does the surface chart, there has been a strong tendency in recent years, with the increase of aerological data, to treat the 700- or 500-mb charts as the primary forecasting tools and the surface charts as secondary. The same type of extrapolation procedure as outlined above for the sea-level map can be applied effectively to the simpler upper-level pressure patterns to obtain prognostic upper-level charts, which then serve as a basis for the prognostic sea-level chart. Particularly useful for this type of upper-level synoptic extrapolation, although not as widely used in practice as they deserve to be, are certain kinematical principles formulated by Rossby [10] relating the deepening and filling, and the movement of upper-level troughs and ridges of the isobaric pattern, to the phase relation of the corresponding isothermal pattern. In addition to these synoptic techniques some of the more recent mathematical extrapolation techniques discussed below are properly applicable only to the upper-level flow patterns.

Vertical cross sections through the atmosphere, notably meridional cross sections, also present synoptic information that should be most useful to physical diagnosis and prognosis of the weather, but up to the present time their incorporation into daily forecast practice has been limited primarily to the current and prognostic description of flight weather conditions along selected air routes.

The synoptic extrapolation technique of forecasting has been placed on a relatively routine basis by the method of kinematical extrapolation which was developed by Petterssen [7]. This method consists essentially of a quantitative determination of the instantaneous speed of displacement, in any selected direction, and the tendency toward deepening or filling, of characteristic features of the current horizontal pressure field which are being effected by the current pressure-tendency field. The calculated trend of movement, or of deepening, is assumed to continue unchanged. The method can be applied equally well to sea-level or to upper-level pressure patterns. In their complete form Petterssen's formulas do include acceleration terms which quantitatively evaluate the effect of the current change of trend of development. Unfortunately, however, these complete formulas are both time consuming and difficult to apply because the synoptic data inevitably are inadequate for their objective evaluation; hence so much is left to the judgment of the individual forecaster that the method loses its principal advantage —that of being objectively and routinely applicable by the inexperienced forecaster. Consequently, Petterssen's forecast method as it is usually applied is a quantitative method of routine extrapolation, which

is based on the expectation of a continued uniform trend of development of the synoptic weather pattern. In this character it is particularly susceptible to the basic weakness of all of the extrapolation techniques of forecasting—gross inaccuracy if extended beyond a very limited time range under rapidly or erratically changing conditions.

All of these synoptic extrapolation techniques, which in the past have constituted almost the sole basis of short-range and daily forecasting practice, and which even today are supplanted only occasionally and to a minor degree by techniques of mathematical extrapolation or physical reasoning, are limited by a ceiling of potential performance, which probably is reflected very closely by the best current forecasting. This ceiling is imposed by the failure, and presumable inability of this synoptic extrapolation procedure, even when supplemented by the best synoptic experience, to anticipate systematically or evaluate reliably the indecisively erratic or exceptionally abnormal changes of the synoptic weather patterns. The question may legitimately be raised, To what extent and over what periods of time is the succeeding weather pattern determined by the preceding pattern? or To what extent may factors quite external to the lower atmosphere play a determining role? But it appears certain that if either synoptic or mathematical extrapolation techniques are to lead to any basically substantial improvement of weather forecasting, it will be only on the basis of a much better physical understanding of the mechanics of the atmospheric circulation processes than is represented by the present-day models.

In recent years there has rapidly developed an extensive application of synoptic extrapolation techniques to extended and long-range forecasting, for periods of five or ten days, a month, or even a season in advance. Instead of applying the extrapolation techniques to the weather patterns on a sequence of daily synoptic charts, making use of 24- or 12-hr changes and the 3-hr tendencies, five-day, weekly, or monthly mean charts present the weather patterns which are extrapolated by means of weekly or half-weekly and monthly or half-monthly changes, and 24-hr tendencies. The techniques of extrapolation are not essentially different; basically only the time scale is changed. The maps which are used must be geographically extensive, covering at least half (meridionally) and preferably the whole of the Northern Hemisphere, thus making possible the analysis and characterization by circulation indices, such as the zonal westerly index, of the state of the general circulation as an integrated whole [14]. The synoptic extrapolation techniques are more effectively supplemented by statistical aids in this extended type of forecasting than in the daily forecasting, notably by lag-correlation statistics and by the comparison of prevailing with normal tracks of cyclones and anticyclones.

This type of extended synoptic extrapolation (by use of mean charts), supplemented by statistical and in some cases by physical considerations, forms the basis of the five-day and the experimental monthly forecasts disseminated by the Extended Forecast Section of the U. S. Weather Bureau [6], of the 5- and 10-day and longer-range forecasts of Baur [1] in Germany, of the composite-chart technique of the Multanovski-Pagava school in Russia, and of many others.

The statistical and synoptic extrapolation techniques as applied to extended and long-range forecasting are definitely limited in their potentialities. These techniques perform best for short-range forecasting, showing decreasing skill with increasing range. However, their performance in extended and long-range forecasting of mean weather conditions over extended periods goes far beyond anything that is attainable by the extrapolation of daily weather patterns. Probably 75 per cent for temperature and 65 per cent for rainfall represent approximately the top skill performance which can be hoped for as an average in the forecasting of mean conditions by these methods for one week in advance on the basis of our present limited understanding of the general circulation. Doubtless the skill margin of these potential ceiling performances should be cut in half for monthly forecasts, and possibly to one-third for seasonal forecasts.

Frequently the extended forecasts of the mean weather patterns are used as a guide by which, and framework within which, to extend the prognosis of the day-to-day sequence of weather patterns beyond the two- or three-day limit which is justifiable by the daily forecasting techniques. There is certainly some justification for making such an extension, possibly to five or six days beyond the current date, but beyond that, even though mean conditions can be forecast with some degree of skill, no detectable skill in forecasting the day-to-day sequence of change has been demonstrated, nor does it seem likely to be in the absence of some radical advance in the science of weather forecasting. This advance can come only from an increased physical understanding of the mechanics of the general circulation, not from additional statistical or synoptic manipulation.

The use of analogues represents a routine method of forecasting by synoptic extrapolation which has been rather extensively employed and which can be, but is not necessarily, applied quite independently of other techniques. This method consists essentially of classifying the large-scale synoptic weather patterns by their significant features in some manner such that past maps or conditions which are essentially similar to the current pattern may be extracted readily from files. The weather sequences which followed in the past under similar conditions, at the same season of the year, are then used as a guide to the prognostication of the current sequence. This extrapolation can be made by applying the statistical analysis of a number of similar synoptic patterns in the past (Baur's method), or by selecting from past patterns the single closest fit to the current pattern, and following in detail the weather sequence which resulted in that case (Krick's method). Usually analogue classification and selection is based primarily on sea-level maps, but upper-level charts may be given any desired weight. The method has been extensively

applied to daily forecasting, and to extended or long-range forecasting based on mean maps.

Analogue forecasting has not proved to be notably successful. It is better suited to daily than to extended forecasting, but even in the daily range the extensive effort in the United States to supply analogue information to all forecasters during the war years does not appear to have won many supporters for their use. For the more extended forecast ranges the method suffers the inevitable disadvantage of any routine technique of synoptic extrapolation—that of complete inelasticity in meeting rapidly and erratically fluctuating conditions. Essentially the method fails at two points:

(1) Our knowledge of the mechanics of the general circulation is completely inadequate to the task of establishing any analogue classification of the large-scale circulation patterns which is of basic prognostic significance. In the absence of such knowledge the establishment of sufficient analogy between two synoptic weather patterns to justify the use of the subsequent weather sequence in one case to forecast that of the second case in detail requires a basic similarity not only of the large-scale sea-level synoptic patterns, but also of the upper-level patterns and of the preceding sequence of development by which the analogous patterns come into existence. Needless to say such strict specifications can only rarely be satisfied in the selection of an analogue.

(2) In the absence of a prognostically determinative analogue classification, the selection of an analogue is of little or no assistance to the experienced forecaster. As with the use of statistical aids in short-range forecasting, the experienced forecaster will usually do better by judging the synoptic indications of each individual case in the light of his past experience. In extended or long-range forecasting, the case for analogues is even weaker. In other words, the use of analogues has even less to offer towards any real improvement of weather forecasting in the present state of our physical knowledge of the problem than do the other techniques of synoptic extrapolation.

c. Mathematical techniques of extrapolation—based on various manipulations of the equations of motion and continuity. Accurate weather forecasting by mathematical computation is an ultimate objective for the attainment of which nearly every meteorologist hopes, but as a practical reality it appears today to be quite as distant as when Richardson [8] made his classical contribution to the problem in 1922. Richardson failed completely to derive, from the theoretical equations, satisfactory forecasts even of the short-range (6-hr) changes of the meteorological elements. This failure was doubtless caused in part by his efforts to deal with all of the variables at once, which complicated his calculations to a point where he was unable to identify the sources of his errors, but also by the further fact, since proved by Charney [4], that his unit time interval (6 hr) was far too large, relative to his space grid, to permit a reasonable (convergent) solution of the equations.

Interest in the numerical prediction of weather has been greatly stimulated by the recent development of high-speed computing devices, which places the feasibility of lengthy numerical reckoning on an entirely new basis. In spite of initial optimism, however, as to the potentialities of these computing techniques, it is generally recognized at present by those who have been working on these methods that no radical advance in practical weather forecasting in the near future is probable. One difficulty lies in the great mass of observational data that must be treated in order to provide in time and space a sufficiently extensive and dense network of observations to compute the future state of the atmosphere a day or more in advance. However, that is a technical problem that doubtless can be overcome. A second difficulty lies in the magnitude of random local variations of the weather elements, variations which are large compared with the permissible tolerance of observational errors. Possibly this difficulty also can largely be overcome by smoothing techniques. But the principal difficulty lies in the fact that computing devices are not brains. They must be told what do to. At present our understanding of the mechanics of the general circulation is quite unequal to this task; there exists no practical conception of how the large-scale circulation processes work, to serve as a physical or theoretical basis of computation. Furthermore, as pointed out above, it is not even known to what extent the future state of motion of the atmosphere is determined by the preceding state, or to what extent internal or external energy sources need be taken into consideration. Hence the mathematical extrapolation techniques run up against the same obstacle as do the less rigorous extrapolation techniques—need of a better understanding of the mechanics of the general circulation. At present, high-speed computing machines may be very useful to test the applicability of physical or theoretical models to the large-scale atmospheric processes in nature, but certainly they do not in themselves offer any particular hope of solving the basic problem of weather forecasting, which is to acquire a better understanding of the irregularly fluctuating circulation patterns. Their practical usefulness will probably increase as this knowledge is gained.

A number of computational extrapolation techniques have been applied to practical weather forecasting, techniques which are based on various manipulations of the equations of motion and the equation of continuity in which certain simplifying assumptions are made in their practical application to the atmosphere. The most promising of these forecast aids are those based essentially on the tendency equation of Bjerknes [3], and on Rossby's use of the vorticity principle [9] to compute upper-level wave motions and air-particle trajectories.

Numerous efforts, involving the tendency equation, have been made to compute the fields of horizontal convergence and vertical motion in the middle troposphere from the observed wind field and to compare the thermal advective effects on the observed temperature field with the dynamic effects. Synoptic patterns computed in this manner are correlated with observed pressure changes, that is, cyclogenesis and anticyclogenesis,

and with the hydrometeors, in an attempt to find a numerical basis of extrapolation for forecasting of the short-range or daily types. The success of all such attempts has been notably indifferent, to the extent that none of them has found general quantitative application in practical daily forecasting. This lack of practical application is caused partly by the large amount of work involved and partly by the lack of encouraging results thus far obtained. This failure to obtain good results doubtless stems from the same difficulties which cause the failure of the more elaborate computational techniques: notably insufficient quantity and accuracy of synoptic observations, difficulty in smoothing out local disturbances, and insufficient physical understanding of the atmospheric processes. For these same reasons, equally little can be hoped from these efforts at present towards the solution of the forecast problem.

Rossby's vorticity technique has been applied to both the daily and the extended forecasting of flow patterns in the middle troposphere with enough demonstrable success so that it has received some routine application, notably by the Extended Forecasting Section of the U. S. Weather Bureau [6]. It has proved useful as an auxiliary forecasting tool under certain conditions, but the limitations which are imposed upon it, in part by the restriction of its use to clearly defined wave-patterns aloft, and probably more fundamentally by the necessary assumption of horizontal (nondivergent) air flow, are in themselves sufficient guarantee that this method cannot contribute significantly to the solution of the basic forecast problem.

2. *Physical Techniques of Forecasting the Weather*—based on a consideration of the physical factors which may actively modify the existing state and trend of the weather as it progresses. It was mentioned above under the discussion of synoptic extrapolation techniques that the usual short-range and daily forecasting procedure entails the use of a few physical considerations at least as a supplement to the purely synoptic extrapolation. In the same manner, certain physical factors frequently form the basis of some of the statistical extrapolation techniques in extended or long-range forecasting.

In short-range and daily forecasting the most effective use of physical considerations to supplement or modify the synoptic extrapolation techniques has developed from the Norwegian polar front theory, or as more commonly designated now, from air-mass and frontal analysis. The study of the physical properties of air masses, in particular the vertical distribution of moisture and the vertical stability, together with the physical factors which modify them, has proved most useful for the type of detailed local forecasting which is so important to aviation. The use of energy diagrams such as the adiabatic diagram, the tephigram, or the aerogram to indicate the physical probability of local convection, and the consideration of topographic influences, such as upslope or downslope motion, surface heating or cooling from land and water surfaces to change the air-mass properties, and the turbulence and mixing characteristics produced by wind and rough terrain—all of these are physically modifying influences which are normally considered in routine short-range forecast practice.

Likewise the physical concepts of frontogenesis and the accumulation of available potential energy supplied by air-mass convergence, of cyclogenesis by the development and occlusion of wave disturbances with all of the attendant weather cycle and sequences of hydrometeors, of the modification of frontal structure and condensation forms by orographic barriers and coast lines, all of these are physically modifying influences, the consideration of which has added much to short-range and daily forecasting.

However, the prognostic potentialities of air-mass and frontal analysis and related physical techniques of forecasting appear to have been largely realized. Probably some further refinement of forecasting detail may be effected along these lines, but no radical advance can be hoped for from this quarter. In fact, it can be asserted that the contribution of frontal and air-mass analysis to scientific weather forecasting has fallen far short of that which was hoped for in the early days of the development of this new school. The reason for the failure of frontal and air-mass concepts to solve more weather problems probably lies to a considerable extent in the following observational or hypothetical facts:

a. The utter complexity of atmospheric conditions which are not even approximately represented by such concepts as homogeneous air masses and frontal discontinuities.

b. The fact that cyclogenesis as a process probably rarely if ever closely approximates the ideal Bjerknes wave-cyclone model.

c. The fact that the primary impulse or drive of large-scale cyclogenesis and anticyclogenesis frequently originates far outside of the developing center, hence cannot be identified or anticipated by local conditions.

The physical factors which have been assumed to exert either modifying or controlling influence over the extended or long-period anomalous fluctuations of the large-scale weather patterns, and hence to require either secondary or primary consideration in extended or long-range forecasting, may be classified as follows:

a. Continental (orographic barriers, coast lines, and extensive snow cover).

b. Oceanographic (anomalous surface current flow, temperature, and polar ice conditions).

c. Extraterrestrial (sun, moon, and planets).

Of the continental factors mentioned, obviously the topographic features vary only during geological time, hence their influence is expressed in the seasonal normal patterns. It is only as the circulation pattern is markedly anomalous by reason of some primary disturbing factor that topographic influences find expression as secondary anomalous characteristics of the circulation pattern, notably in the windward and lee effects of a major orographic barrier on an abnormally strong cross-wind system. Likewise an extensively anomalous condition of continental snow cover doubtless contributes to a minor degree to the persistence or intensity of an anomalous weather pattern, but much statistical analysis has failed to establish any signifi-

cant correlation between snow cover and subsequent anomalous conditions of the large-scale weather patterns. Hence these continental factors are at best only of minor secondary importance in extended or long-range forecasting.

A great amount of statistical work has been performed by Helland-Hansen, Walker, C. E. P. Brooks, and many other investigators[2] in the effort to correlate anomalous conditions of sea-surface temperature and polar ice with contemporary, preceding, and subsequent anomalies of pressure (atmospheric circulation), air temperature, and rainfall. The upshot of most of this work can be summarized essentially as follows:

a. Anomalous conditions of sea-surface temperature and polar ice are explained in general by present and past anomalies of atmospheric wind and temperature.

b. Sea-surface temperature anomalies and cooling by melting ice are small in amplitude compared to air-temperature anomalies and show little tendency to persistence or displacement except as they are maintained by anomalies of the atmospheric circulation. They cannot conceivably, from any quantitative consideration, be the primary cause of contemporary or subsequent air-temperature anomalies.

c. Some significant correlation has been found between spring ice conditions in the Greenland Sea and Barents Sea regions and subsequent summer, autumn, and winter weather in northern and central Europe. Since the sea-surface anomalies are utterly inadequate to account for the subsequent weather anomalies, it appears that both must be related to some primary factor of weather control which is best expressed in the highly anomalous state of the general circulation by which extreme ice conditions are initially produced.

Certainly it is reasonable to conclude, granted the correctness of the above statements, that a further study of oceanographic influences is not a promising or direct line of attack on the basic problem of extended or long-range weather forecasting.

Solar control, either direct or indirect, of the normal seasonal features of the world weather patterns is axiomatic, but the question as to the extent to which the anomalous fluctuations of the general circulation are influenced or controlled by irregular solar activity is a highly controversial one. Literally scores of investigations, statistical, synoptic, and theoretical, have been directed toward one or another aspect of this problem. Relationships of world weather patterns to sunspots or solar constant have been investigated by Abbot, Baur, Clayton, B. and G. Duell, Hanzlík, Helland-Hansen, Köppen, Kullmer, Schell, Simpson, Tannehill, Walker, Willett, and many others. Direct solar effects on the higher atmosphere, notably on temperature, circulation, distribution of ozone, and ionization, have been observed or theorized about by equally many investigators, including among others Craig, Dellinger, Dobson, Götz, Haurwitz, Hulburt, Maris, Mimi, Shapley, Stetson, and Wulfe.

In spite of the great amount of effort which has been expended on this problem it remains unproved today whether solar activity plays a primary, secondary, or insignificant role in the irregular fluctuations of the atmospheric conditions of the troposphere (world weather patterns). There is no question whatsoever about the occurrence of numerous direct effects of sudden solar disturbances in the higher atmosphere, but any such effects in the lower troposphere, if they exist, are so indirect, masked, and complex that they have not been statistically confirmed. There is an imposing amount of statistical evidence for long-term fluctuations of the world weather patterns, particularly in the tropics, but also in the higher latitudes, that roughly parallel the single or double sunspot cycle. Since the parallelism is either not uniformly close or not consistent in phase over long periods of time, the evidence indicates that sunspots themselves are not a satisfactory index of the disturbing solar influences. It can be stated without qualification that up to the present no attempt to base extended or long-range weather forecasts directly on any solar index, usually sunspots or solar constant, has attained any significant success. Probably no primarily statistical attack on this problem can be expected to accomplish more at present.

Assuming that there is an ultimate control or direction of the major anomalous fluctuations of the general circulation by solar activity, then our failure to establish the reality or nature of this control is readily explained by our almost complete ignorance in the following three areas:

a. The physical nature and intensity of the disturbing solar influences which enter the outer atmosphere.

b. The qualitative and quantitative effects which these influences produce in the higher atmosphere.

c. The essential mechanics of the general circulation, including any possible mechanism of interaction between the troposphere and the higher atmosphere.

Probably the most representative opinion of meteorologists who have recently dealt with the broader aspects of the anomalous solar-weather relationships [1, 2, 17] can be expressed approximately as follows: Solar activity undoubtedly exerts some guiding influence, possibly even directing control, on the extended and long-range anomalous fluctuations of the general circulation. This influence or control is so indirect and so complex in the sum total of its manifestations that little progress is to be expected from any further primarily statistical attack on the problem. This problem calls for a major program of research in solar and atmospheric physics, a program which must be an integral part of any major attack on the basic problem of weather forecasting. Such a program must be directed at all three of the basic areas of ignorance noted above.

RECOMMENDATIONS FOR THE IMPROVEMENT OF WEATHER FORECASTING

Possible Immediate Technical Improvements. The general problem of the improvement of weather forecasting must be considered from two rather distinct angles. On the one hand there is the question of possible accomplishment in the immediate present or near future

2. See *Mon. Wea. Rev. Wash.*, Supp. No. 39, pp. 27–57, 1940.

within the framework of our present basic knowledge of atmospheric circulations and with our present forecasting techniques. On the other hand there is the question of the ultimate approach to true scientific weather forecasting by means of extended research in the basic problems of dynamic meteorology and the development of new objective forecasting techniques.

It is emphasized in the preceding discussion that important progress in the basic problem of weather forecasting is not to be achieved by mere extension, elaboration, or refinement of any of the present great variety of forecasting techniques. On the other hand, it is quite certain that some improvement of accuracy and uniformity of forecasting performance beyond that now attained can be achieved on the basis of our present basic knowledge and forecasting techniques. The possibility of the technical improvement of present forecasting procedures may be considered at three points:

1. *Weather Observations.* Our knowledge of present and past weather conditions, on which are based both the forecast of the future weather and the verification of past forecasts, depends almost entirely upon the analysis of synoptic weather observations. Obviously it is important that these observations be adequate for the uses to which they are to be applied. By and large it may be stated that observational techniques are far more advanced than forecasting techniques in meteorology, hence primary emphasis in meteorological research for forecasting purposes should definitely be shifted from the former to the latter. However, meteorological observations are far from being all that they should be. In this connection it can be stated that attention should be devoted primarily to the reliability and distribution of weather observations, rather than to increased accuracy, elaborateness, or density of observations in favored local areas. The greatest need, for the improvement of general or extended forecasting as well as for forecasting research, is for simple reliable observational techniques which can be applied uniformly to large areas of the earth's surface and to the upper atmosphere, to fill in the extensive global gaps which still exist in available synoptic information. This calls for instruments which are as foolproof and free of calibration difficulties as possible, and the further development of automatically recording and transmitting weather-station equipment for use in relatively inaccessible areas. Any possible simplification of raob and rawind techniques is highly desirable.

More elaborate or exact weather-measuring devices are not justified for general use because the small-scale random variability of weather conditions renders exact measurement of no significance. For example, useful as it would be to observe the field of vertical motion in the atmosphere, exact measurements of this quantity, even if they could be made, would be of little value because the local variations are very large relative to the prevailing motion.

It is only for short-range forecasting of local conditions for special purposes that it is possible to justify a dense network of special observations, such as the radar technique for the measurement of condensation forms. Insofar as it appears advisable, detailed study of local forecast problems and problems of atmospheric physics can be set up with full equipment in selected limited areas, but at the present time it is unlikely that such studies will contribute as much to the solution of the basic problem of general weather forecasting as will adequate observations to eliminate the great gaps in available synoptic information in many parts of the world. The primary inadequacy of our present forecasting knowledge lies in the field of the dynamics of the general circulation rather than in the field of atmospheric physics.

The proper implementation of an adequate world-wide network of synoptic weather observations is completely dependent upon the existence of a strong international meteorological organization which can specify uniformity in instrumentation, observational techniques, hours of observation, and the coding and transmission of synoptic data, even to the distribution of international funds to support such a program. There is no possibility of realizing the necessary world-wide network of synoptic weather observations except as it is internationally financed according to the ability of the different nations to support it.

2. *Standardization of Forecasting Techniques.* It is evident from the discussion of forecasting techniques (see p. 733) that there is practiced today a great variety of such techniques, many of which are overlapping, essentially redundant, or of very limited applicability, and for which little or nothing in the way of rigorous objective evaluation has been attempted. It was noted that the essential basis of almost all short-range or daily forecasting consists of the empirical extrapolation, in the form of two-dimensional prognostic charts, of the current two-dimensional synoptic weather patterns, influenced by certain physical factors such as topography and the modification of air-mass properties or of frontal structure, and occasionally aided by the use of some climatological statistics. The three-dimensional distribution of the weather elements, both current and prognostic, is expressed by the selection of suitable two-dimensional sections for analysis and prognosis. With increasing range of weather forecasts to the extended and long-range categories, the forecasting techniques are based increasingly on the extrapolation of mean synoptic rather than instantaneous synoptic patterns and on the use of statistical aids.

The greatest improvement of weather forecasting which can be expected within the framework of our present basic knowledge almost certainly is to be effected by the standardization and simplification of the multiplicity of techniques which are employed for the prognostic extrapolation of synoptic weather patterns, particularly in the short-range and daily forecast categories. This applies both to the techniques of extrapolation and to the preliminary presentation and analysis of the synoptic data. It is possible to present synoptic weather data in an infinite variety of forms, as regards the two-dimensional sections which are selected for analysis, the combinations of weather elements which

are selected for presentation, and the units or quantities in which the selected elements are plotted or combined. Most of the infinite possible variations of presentation and analysis of synoptic weather data appear to have been introduced into weather-forecasting practice at one point or another. The horizontal or quasi-horizontal distribution of the weather elements is on occasion represented in constant-level charts, isobaric, isentropic, or tropopause contour charts, isobaric or isentropic thickness charts, isopycnic contour charts, etc. Most of these charts, together with corresponding change charts, are prepared at more or less arbitrarily selected levels, contain varying combinations of the weather elements, and are subjected to variable analytical techniques. Similarly the vertical cross-section charts are plotted along selected meridians, selected latitude parallels, selected airways, or selected lines of raob stations; they are prepared alternately only for the lower flight-levels or through the tropopause, and they are analyzed for a great variety of combinations of the meteorological elements and quasi-horizontal surface intersections. Likewise the vertical structure of the atmosphere as obtained from raob soundings at selected points is represented by a great variety of energy diagrams, notably by the adiabatic diagram (Neuhoff or Stüve), the tephigram, the aerogram (emagram), the equivalent-potential temperature diagram (Rossby) and a number of other less frequently used permutations of the basic thermodynamic diagram.

What have been the consequences of this great profusion of synoptic-analytical tools? They have been essentially disadvantageous, for the following reasons:

a. There is extremely little gained, in proportion to the effort expended, by the preparation of a large number or great variety of presentations of the same synoptic data. It is true that specific requirements for different types of forecasts may best be met by some variation in the presentation of synoptic data, but this desirable variation is comparatively small and has little bearing on the great variety of synoptic tools which have been developed, and which are largely redundant, merely slicing the same information in different ways.

b. A great amount of time and effort, which might better be turned to more constructive use is consumed in the nonproductive manipulation of the elements, units, and coordinates used in the synoptic presentation and analysis of meteorological data. The same criticism of waste of time can be made of much of the effort to base weather forecasting on the statistically most probable or normal sequence of change or movement of the weather pattern.

c. The principal positive harm which this multiplicity of synoptic tools causes is that of confusing the forecaster. It is quite impossible to assimilate and integrate mentally a synoptic picture which is presented in too great complexity and on too many charts. Furthermore, synoptic and forecasting techniques vary so greatly from one forecast center to another that any forecaster is likely to be confused as soon as he leaves his own bailiwick or the system to which he has become accustomed Where forecasting is as empirical and as dependent as it remains today upon the forecaster's experience and his mental awareness of similarities and dissimilarities of synoptic patterns, this confusion and lack of precision and clarity in his mind with respect to comparative significant details of the synoptic patterns can be very demoralizing. This one factor, coupled with the relatively short forecasting experience of many forecasters today, probably accounts largely for the widespread failure of practical forecast performance in recent years even to begin to keep pace with the increase of available synoptic information.

The remedy for the present confusion of techniques for the presentation, analysis, and extrapolation (prognosis) of synoptic weather patterns obviously lies in elimination, simplification, and standardization. This is an immediate problem which should be faced on the basis of our present knowledge. The long-range problem of increasing our basic understanding of atmospheric circulations and of developing correspondingly scientific forecast techniques must be considered quite apart from the immediate problem of simplifying and standardizing present routine forecast practice.

The mere accomplishment of such simplification and standardization will doubtless be of far greater significance to the performance of short-range and daily weather forecasting than will the selection of one particular system of analysis and prognosis rather than another. However, any such program calls for a real effort to devise an objective system of verification by which forecasting skill can be rated and synoptic tools evaluated. The principal emphasis must be placed on simplification, to select from the great variety of synoptic charts and analytical techniques now in use a minimum number of charts and a technique of analysis which will present as clearly and comprehensively as possible the essential features of the changing synoptic pattern. This presentation must vary somewhat with the type of forecast desired and with the broad climatological zones, but such necessary variation should be kept to a minimum, and standardized at least to the extent that forecasters performing similar work in similar climatic zones can speak the same language to one another, which is far from being the case today. This standardization should not at all discourage practicing forecasters from doing forecasting research, particularly on local or regional problems, but it does mean that the performance and practical application of such research should always be supplementary to, and not a substitute for, the standard prognostic procedure, at least not until it has been officially tested and approved as part of that procedure.

Obviously, to be really effective, such a program of standardization of synoptic analysis and prognosis must be developed and applied on an international scale, by a strongly centralized international meteorological organization, and it should be coordinated, by this same organization, with the similar program for weather observations which is outlined above.

Probably less is to be gained in extended and long-range forecasting by standardization and simplification of the synoptic techniques of analysis and prognosis

than in the shorter-range categories, but a very real need also exists in those fields for the evaluation of synoptic techniques and the simplification of prognosis by the elimination of the less effective synoptic tools. It is highly probable, however, that relatively more is to be gained in these longer-range forecast categories by further statistical studies of the behavior and operation of the general circulation than in the shorter-range categories where statistics now contribute relatively little. On the other hand it is quite clear that statistics alone cannot contribute fundamentally to the basic solution of the problem of long-range weather forecasting. It is exactly in this field that basic research in the thermodynamics and mechanics of the general circulation, guided by synoptic observational statistics, is most necessary at the present time.

3. *The Selection and Training of Weather Forecasters.* It must be recognized that weather forecasting as practiced at present and probably for many years to come is essentially empirical and esoteric in nature. As a consequence of this fact, the practicing weather forecaster, whose primary occupation is the routine preparation of weather forecasts of the usual types, must draw extensively on a fund of highly esoteric knowledge which has been gained by long first-hand experience with weather forecasting of the type and in the region with which he must be concerned. Scientific theory plays little part in practical weather forecasting at the present time. Consequently, any forecast service which is concerned primarily with the regular issuance of ordinary types of weather forecasts of maximum accuracy will benefit notably by adhering rather strictly to certain principles with regard to the selection of professional forecasting personnel, in particular as follows:

a. The empirical and esoteric nature of most weather forecasting today places a high premium on the natural aptitude of the individual forecaster. This natural aptitude is invariably indicated by a strong liking for and sustained interest in weather forecasting for its own sake over a long period of time, and can be verified only by the routine objective verification of competing practice weather forecasters over a considerable period of time. It is a mistake to assume that merely because a man has passed a good professional course in meteorology creditably and has learned a certain amount of theory and practice he will be interested in forecasting or successful as a weather forecaster. In many cases he will not compare as a practical forecaster with a relatively untrained meteorologist who is absorbingly interested in practical weather forecasting. Under the pressure of wartime demand many meteorologists who were by aptitude and interest entirely unsuited for forecasting duty were pressed into it. Quite the contrary is true of the meteorologist whose primary concern is to do forecasting research. In that case the primary requisite is an extensive scientific background, including both theoretical training and an interest in and some first-hand experience with techniques of synoptic analysis and prognosis.

b. It is highly important that the forecaster's training and synoptic experience be such that he can assimilate a maximum amount of practical specific information of significance to the prognostic judgment of synoptic weather patterns. This can be accomplished only if he concentrates his attention on a minimum number of basically significant synoptic charts and prognostic techniques, rather than by dispersing his attention over an elaborate array of synoptic and prognostic tools. This is the problem of the simplification and standardization of forecasting techniques which was discussed above.

c. Weather forecasting is the exacting work of a specialist; it should be a full-time, long-term job. It may very well be combined with or alternated with related research in cases where the forecaster is properly qualified, and in any case ample opportunity must be given the forecaster for complete relaxation from the pressure of prolonged forecasting responsibility. In no case should responsible forecasting duty be combined with or alternated with any unrelated exacting job such as administrative or directive work, nor should an important forecasting assignment ever be made a temporary duty to which a man is assigned for a limited period, perhaps for two or three years, after which he is transferred to something entirely different. In this respect the military services are frequently particularly remiss.

d. The incentive to hold the good forecaster in his job should always be present. Specifically, in civilian (civil service) forecasting agencies the rating of forecasting positions should be at least as high as, if not higher than, that of comparable administrative or other positions which might be open to the forecaster. Obviously, forecast accuracy should be the first concern of any weather bureau. Furthermore, the forecaster should be subjected continually to an objective skill verification of his performance in competition with other official and practice forecasters, and there should not be any obstacle to the demotion of a forecaster to other work in case of a poor showing in competition over a long period of time (at least one year as a minimum). In the military weather services it should never be to the disadvantage of a successful forecaster to continue in his work. In other words, the meteorologist and responsible forecaster should always be a civilian technical expert, not a commissioned officer. Officers must take the responsibility for decisions based on the weather forecast, but in no case should the responsible weather forecaster be a commissioned officer whose chance of promotion is impaired by his becoming or permanently remaining a weather expert.

The remarks outlined above indicate a few measures which should contribute to the standardization and moderate improvement of present forecasting performance. However, the fundamental problem of achieving a reasonably scientific basis for weather forecasting will not be solved by any such routine elaboration of present techniques as that just outlined. If this fundamental problem is to be solved, that solution will be achieved only by basic research into the unsolved problems of dynamic meteorology, particularly into the thermodynamics of atmospheric circulations, and the

application of the knowledge thus gained to the development of entirely new scientific forecasting techniques. Obviously, it is quite impossible at this time even to suggest how such a solution is to be achieved, but the following remarks are offered as one meteorologist's opinion as to the direction from which the problem should be approached.

Approach to the Problem of Scientific Weather Forecasting. It is evident from the preceding discussion of the essential limitations of the present forecasting techniques that the primary inadequacy which they share stems from the fact that they are all essentially extrapolation techniques which, by one method or another, statistical or synoptic, qualitative or quantitative, attempt to derive the future or prognostic weather pattern from the current and past patterns. None of them is based on adequate thermodynamic concepts either of the operation of the basic drive of the general circulation, or of the energy sources or transformations which are involved in sudden accelerations or changes of trend in the development of the weather pattern. They all fail at the point of anticipating such frequent accelerational developments before the process is well in progress, when it is too late for anything but a very short-range forecast. The basic problem of forecasting research is to derive some quantitative physical model of the general circulation which fits the statistical and synoptic facts of weather observation.

There are two rather opposite points of view from which this primary forecast problem is most frequently approached. They represent two quite distinct philosophies of the basic character of the longer-period anomalous fluctuations of the world weather patterns. One method is to make an intensive study, by means of a dense network of complete and frequent synoptic observations within the area of the investigation, of the detailed structure and of the dynamic and thermodynamic activity of the atmosphere within a limited region from which complete data can be obtained. The purpose of such an investigation is primarily to determine the mechanics of operation of the dynamic and thermodynamic processes by which limited cellular circulations of the atmosphere are locally generated, transported, and dissipated. Intensive, relatively localized synoptic studies of this type tend to be favored because of the comparative ease of obtaining adequate observational data for restricted areas. It is more or less tacitly assumed in such studies that the entire developmental process of the local circulation is sufficiently self-contained so that it can be essentially explained by conditions within the limited field of observation [5]. If this line of approach to the basic forecast problem is selected, the choice implies that scientific forecasting can be realized at best only as a short-range accomplishment, that the general circulation in its entirety exists only as the integration of a number of individual cells or centers of action which operate independently and more or less at random such that the entire system lacks any unifying principle or control which makes its character or behavior over extended periods of time either distinctive or predictable.

The opposite line of approach to the basic forecast problem is to study the general circulation extensively, preferably over all of the globe from which even an approximate picture of the circulation pattern can be obtained. The time scale is likewise extended to deal with charts at 24-hr intervals or with mean charts for even more extended periods of time, to obtain a comprehensive picture of the large-scale fluctuations of the general circulation in its entirety over relatively long periods. The essential purpose of such an extended investigation is primarily to determine the mechanics of operation of the dynamic and thermodynamic processes by which the entire general circulation is maintained in its continuously and irregularly fluctuating pattern of intensity and form. Investigations on this scale tend to be avoided because of the relative inadequacy of observational data over large parts of the earth's surface, and the large amount of work involved in obtaining and processing the data for analysis. It is more or less tacitly assumed in such studies that the general circulation in its larger features, at least on either hemisphere, operates essentially as an integrated unified system which is functionally interrelated in all its parts.

If this second line of approach to the basic forecast problem is selected, the choice implies that scientific forecasting must be based primarily on a better physical understanding of the operation of the general circulation as a whole, and that the development and movement even of the secondary cellular circulations, and of all the attendant weather phenomena, the primary concern of the short-range forecaster, depends primarily upon the dynamics or thermodynamics of the hemispheric flow pattern and cannot be explained or anticipated on the basis of only the regional conditions in the sector of the development. This point of view indicates that the greatest improvement in weather forecasting is to be expected in the extended or long-range rather than in the short-range categories.

There are a number of indications, based on a great number of statistical, synoptic, and theoretical investigations in meteorology and in related fields, that this second approach to the basic forecast problem is essentially the correct one. This fact does not at all imply that nothing is to be gained from synoptic or theoretical studies of the secondary or tertiary atmospheric circulations, but it does imply that if this type of study is restricted to a limited sector of the earth's surface, factors of primary importance to the weather development probably will be overlooked. The primary evidence for the essential correctness of the global rather than the regional line of approach to the forecast problem may be summarized briefly as follows:

1. The failure of many years of intensive study of regional weather patterns to evolve either a physical model or a theory of cyclogenesis and anticyclogenesis (pressure changes) of any real practical value in forecasting regional changes of the synoptic weather patterns. Gross inadequacy of world-wide observational synoptic data has prohibited any corresponding study of the general circulation in its entirety.

2. The global extent of the persistent and systematically changing anomalous characteristics of the general circulation patterns, notably in their high-index (zonal, poleward displacement) versus low-index (cellular, equatorward displacement) features. The global, or at least hemisphere-wide, character of these basic index fluctuations is shown by significantly higher contemporary correlation between circulation indices for the entire Northern Hemisphere than between indices for the individual continental or maritime meridional quadrants of the hemisphere. The statistical improbability of this fact of observation makes it strong evidence against the primary independence of the individual regional cellular developments of the general circulation, but it is entirely consistent with recent studies of energy propagation in the atmosphere [11].

3. The occurrence of markedly similar anomalous fluctuations of the general circulation pattern between high- and low-index characteristics, fluctuations which are in phase in the Northern and Southern Hemispheres, and which increase in amplitude with the length of the period of oscillation. This fluctuating index character of the general circulation has paralleled the secular trends of climate during observational time, the greater changes of climate during historical time, and the extreme fluctuations between glacial and interglacial climates during geological time. The only single physical factor of climatic control which can possibly be made to account for this entire spectrum of synoptically similar patterns of climatic change is that of variable solar activity, presumably in the ultraviolet, and particle emissions which parallel the sunspot activity that is observed to be as erratically variable as the anomalous fluctuations of terrestrial climate [17].

4. The extensive evidence for the world-wide reaction of weather patterns to irregular solar disturbances, as shown by the anomalous fluctuations of monthly, seasonal, annual, and longer-period distributions of pressure, temperature, and rainfall in response to major phase changes of the sunspot cycle, and of the monthly mean distribution of pressure in response to anomalies of the monthly mean solar pyrheliometric values [1; 2; 18, pp. 31–86].

5. The fact that the theoretical investigations which have been most successful in the physical interpretation of the observed characteristics of the general circulation have without exception been based on some global or hemispheric mechanism of operation of the general circulation [9, 11, 12, 13, 15].

If it is accepted that the best line of approach to the basic problem of weather forecasting is by the study of the general circulation of the earth's atmosphere in its entirety, then the general direction that this approach must take is rather obvious. The first concern must be to establish the closest possible cooperation of the theoretical and the synoptic-statistical meteorologists such that all theory and hypotheses may be influenced and rigidly checked by the observational facts. The primary objectives towards which investigation should be directed include the determination of the primary energy sources and sinks in the atmosphere, as well as of the energy transformations and transportation from source to sink; the establishment of the entire dynamics and thermodynamics of the operation of the general circulation as a whole between energy source and energy sink, and an understanding of the mechanics of interaction between the zonal and cellular branches of the general circulation. Particular attention should be directed to the determination of the physical nature of the irregularly variable solar activity, to its direct effects in the higher atmosphere, and to the transmission of all such direct or indirect effects to the lower atmosphere, notably to the troposphere in the tropics, where the influence of the irregular solar activity on the weather is most directly in evidence.

All investigations of this general character, to be practically effective for the physical interpretation of the existing states and observed changes of the state of the general circulation, must be guided and verified by the synoptic and statistical analysis of reliable global observational weather data, and tested over as long a period as possible by extended series of past climatic data. Obviously, the efficient organization of a program such as that suggested above can be accomplished effectively only by a strong international meteorological organization, for its effective prosecution in all of its ramifications more or less of necessity entails the following procedures:

1. The establishment of a well-integrated world-wide system of uniformly distributed synoptically reporting stations, operating under the direct control and support of a strong international meteorological organization, as discussed above. Since the entire research program should be conducted by the same central organization, the details of the density, character, and distribution of the synoptic network of observing stations must be planned, and occasionally modified, with an eye to research requirements quite as much as to practical current forecasting requirements.

2. The development of more effective synoptic tools or techniques for the presentation and analysis of synoptic weather observations on a world-wide scale. Emphatically this development and standardization of synoptic techniques can be effectively implemented only by a strong central meteorological organization, as discussed above. However, the development contemplated as a necessary part of any program of basic research in weather forecasting goes far beyond the mere selection and standardization of synoptic techniques which already are in practice, as previously considered. This development should contemplate the evolution of basically new and improved synoptic techniques as an integral part of, and guided by the needs of, the basic research program.

3. The establishment of one central international meteorological research center, under competent direction, where leading meteorologists of all nationalities would be enabled to work under conditions of complete cooperation and exchange of ideas, with a maximum of readily available synoptic data and the necessary amount of clerical assistance. In this manner it should be made possible for the meteorologists with the most to

offer to forecasting research to make the greatest possible contribution irrespective of nationality or local resources, and to concentrate all available talent with the greatest efficiency on the primary objectives to be attained. Much of the efficiency of operation of such a research center would stem from the fact that the observational synoptic-statistical, and theoretical phases of the research program could be integrated and coordinated to the greatest advantage of the necessary basic forecasting research. A complete long-range program could be formulated and prosecuted under the combined effort of the best minds in meteorology and in the related fields. If eventually the problem of scientific weather forecasting is to be solved to some degree of satisfaction, the solution will be greatly hastened by a program such as that outlined above.

REFERENCES

1. Baur, F., *Einführung in die Grosswetterkunde.* Wiesbaden, Dieterich, 1948.
2. —— "Zurückführung des Grosswetters auf Solare Erscheinungen." *Arch. Meteor. Geophys. Biokl.*, (A) 1: 358–374 (1949).
3. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
4. Charney, J. G., "On a Physical Basis for Numerical Prediction of Large-Scale Motions in the Atmosphere." *J. Meteor.*, 6: 371–385 (1949).
5. Miller, J. E., "Cyclogenesis in the Atlantic Coastal Region of the United States." *J. Meteor.*, 3: 31–44 (1946).
6. Namias, J., *Extended Forecasting by Mean Circulation Methods*, 89 pp. U. S. Weather Bureau, Washington, D. C., 1947.
7. Petterssen, S., *Weather Analysis and Forecasting.* New York, McGraw, 1940.
8. Richardson, L. F., *Weather Prediction by Numerical Process.* Cambridge, University Press, 1922.
9. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2: 38–55 (1939).
10. Rossby, C.-G., "Kinematic and Hydrostatic Properties of Certain Long Waves in the Westerlies." *Dept. Meteor., Univ. Chicago, Misc. Rep.* No. 5, 37 pp. (1942).
11. —— "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2: 187–204 (1945).
12. —— "On the Distribution of Angular Velocity in Gaseous Envelopes under the Influence of Large-Scale Horizontal Mixing Processes." *Bull. Amer. meteor. Soc.*, 28: 53–68 (1947).
13. —— "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6: 163–180 (1949).
14. Starr, V. P., *Basic Principles of Weather Forecasting.* New York, Harper, 1942.
15. ——*A Physical Characterization of the General Circulation.* Dept. Meteor., Mass Inst. Tech., Rep. No. 1, General Circulation Project No. AF 19-122-153, Geophys. Res. Lab., Cambridge, Mass., 1949.
16. Wadsworth, G. P., *Further Analysis of Dynamics of Major Pressure Cells*, Geophysical Research Directorate Rep. No. 5; and *Position of Major Pressure Cells in Relation to Rainfall*, Geophysical Research Directorate Rep. No. 6. Contract No. W28-099-ac-398, Mass. Inst. Tech., Div. of Industrial Cooperation, Cambridge, Mass., 1949.
17. Willett, H. C., "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6: 34–50 (1949).
18. —— and others, *Final Report of the Weather Bureau—M.I.T. Extended Forecasting Project for the Fiscal Year 1948–1949*, 109 pp. Cambridge, Mass., 1949.

SHORT-RANGE WEATHER FORECASTING

By GORDON E. DUNN

U. S. Weather Bureau, Chicago, Illinois

INTRODUCTION

Literature. Literary productivity in meteorology during the past two decades has reached the highest rate in meteorological history. Much of it has been descriptive and theoretical in nature, some of it with forecasting implications, but remarkably little has been directly concerned with or has contributed significantly to the improvement of forecasting. Indeed, the few modern textbooks supposedly directed primarily toward analysis and forecasting devote comparatively little space to the practical problems of forecasting. A tendency exists for the theoretical meteorologist, and even the analyst, to remain aloof from the vicissitudes and discouragements of the practical forecaster. Research workers apparently believe their obligations have been fulfilled when their results and suggestions have been passed along to the forecaster, while the latter, as a rule, has neither the time nor the facilities to test the suggested practical applications of current research. Thus, unfortunately, the gap between the theoretical meteorologist and the forecaster remains unbridged.

The Forecasting Problem. The problem of forecasting may be divided into three separate but closely related phases: (1) analysis, (2) prognostication of pressure patterns, and (3) forecasting the weather. Airmass analysis may be defined as the study of preceding and current meteorological conditions over a prescribed area for the purpose of deriving a satisfactory and logical explanation of the physical processes and weather actually observed. The analysis should include a determination of the structure, location, direction and rate of movement of fronts, the characteristics of the various air masses with particular reference to temperature and moisture, and an explanation of any precipitation areas. Analysis of the area required for forecasting at the district level has become too great a burden for forecasters and the desirability of central analysis centers is now generally accepted.

In addition to the more normal functions, analysis centers (or extended-forecast units) should provide district forecast centers with analyses of broad or large-scale features of the general circulation as determined by the "centers of action" or "weather controls," which often lie some distance off the normal forecasting chart. These analyses would include hemispheric wave lengths and indications of blocking, since these often influence weather developments far away within very short periods of time.

Complete utilization of analysis centers must await further development of facsimile since the coding and decoding of analyses is time consuming and, more important, the consequent smoothing results in a serious loss of character in the analysis. However, transmission of prognostic charts and many types of analyses by facsimile is already satisfactory in the limited areas where this facility is in use.

The preparation of prognostic surface-pressure charts, with indicated frontal positions, and prognostic constant-pressure charts comprises the second phase of the forecast problem. These charts are derived by more or less mechanical means and are subjectively evaluated and modified at the analysis central and then transmitted to forecast centers. Duplication of the same process of analysis at individual forecast centers represents an excessive waste of time and a group of specialists, with an adequate staff for plotting the necessary additional maps and diagrams, should provide forecasters with better analyses than are obtainable by any other method. Prognostic surface charts for periods up to 30 hr now maintain a very high standard of accuracy and satisfactory progress is being made toward 54-hr surface prognostics. The preparation of the 700-mb prognostic chart has, apparently, not met, so far, with the success which might be expected from acquired experience and available techniques. The preparation of both surface and upper-air prognostic charts has not improved to the point where the district forecaster is relieved of the duty of checking and recomputing predicted frontal and pressure system positions for his own district on the basis of additional and later reports and his greater experience in his own area.

The third, and most difficult and important, phase of the forecast problem is the prediction of the weather. Because of the local nature of many forecast problems and the extreme variability and complexity of weather over any large area such as the United States, some decentralization of forecasting is required. This article will deal primarily with the preparation of prognostic charts and the prediction of weather, with particular reference to the middle latitudes, and will be concerned only casually with analysis. However, it should be emphasized that there is no distinct demarcation between these three phases of weather forecasting.

Present State of Short-Range Weather Forecasting. Douglas [19] stated in 1931 that, because of the staggering complexity of the atmosphere, forecasting was largely a matter of experience and judgment and thus almost wholly on an empirical basis. Byers [11] in his chapter on forecasting techniques remarks: "The ability to forecast weather accurately comes as much from experience as from study. The rules for prediction cannot be stated simply and it is extremely difficult to attain success in forecasting through formal study or instruction." In an earlier edition, he further said: "In general, it may be stated that . . . forecasting is the application of all the forecaster's knowledge of meteorology, augmented by thermodynamic calculation to the problem at hand." Willett [58] in a similar chapter and in like

vein declares: "The ordinary forecasting procedure makes no use of any mathematical or quantitative physical methods of diagnosis and prognosis." He had pointed out earlier that "the successful introduction of such quantitative physical methods into practical daily and longer range weather forecasting constitutes the ultimate goal of a large part of present meteorological research. But, up to the present time, the mathematical computation of local changes of weather elements on the basis of physical principles, as distinct from quantitative kinematical extrapolation, has been effectively applied to only a few cases of short-range forecasts at the expense of much time and effort. Really scientific weather forecasting is far from being an accomplishment of the present."

Although efforts to apply dynamic and thermodynamic principles to weather forecasting continue at an accelerating pace and some progress is being made, it must be admitted that forecasting is still largely an art and only in a relatively small part a science. It is difficult at times to explain past and present weather satisfactorily in terms of physical principles and almost impossible in terms of mathematical formulas. Indeed, only two such formulas are in general use in forecasting offices at the present time.

The principal change in forecasting techniques during the past three decades has been the elevation of air-mass characteristics to an importance equal to that of pressure distribution. That is, with the vast increase in upper-air observational data, the forecaster now has a greatly improved three-dimensional synoptic weather picture. However, it should be noted that in the second phase of the forecast problem—the preparation of the surface and upper-air prognostic charts—the emphasis remains on the pressure distribution, although a successful forecast of this characteristic of the weather map far from guarantees a successful weather forecast. As emphasized by Houghton [29] and others, there is no unique correspondence, but only a statistical relationship between the pressure field and the weather. While Houghton advocates further studies of this relationship, he points out the greater importance of the field of motion.

The principal progress in forecasting in the past two decades has been evidenced by a significant improvement in the accuracy of forecasts up to 8–12 hr, in the development of certain specialized types of forecasting, and in the detail currently attempted with some success. For periods from 12 to 48 hr, forecast accuracy has improved no more than 3 per cent. In regions of rapid and frequent weather changes, such as in most of the United States, the accuracy of weather forecasts diminishes 3 to 4 per cent for every additional 6 hr of the forecast period. In view of our better knowledge of atmospheric dynamics and thermodynamics and the vastly increased observational data, it must be admitted that the accuracy of short-term forecasting has not increased proportionately.

With the development of new electronic computing devices and recent work by Charney and Eliassen [14], there is currently renewed interest in the classical attempt of Richardson [44] to forecast the weather mathematically. There is, apparently, good reason for optimism that some practicable forecast aids may be forthcoming from this source. At the moment, however, the possibility of forecasting the weather wholly on an objective basis seems very remote.

The great amount of work completed and under way on the upper air encourages the belief that the development of practical applications of dynamics and thermodynamics, useful in forecasting, will continue. But for some time to come it seems likely that, in the words of A. H. R. Goldie, "forecasting will continue to be a combination of physical reasoning with the practical experience of the synoptic charts."

PROCESSING THE OBSERVATIONAL DATA

Observations. In accordance with international agreement, a basic surface observational network consists of stations spaced approximately 140 mi apart and denser national and secondary networks. Observations from the international network are taken over large sections of the world at approximately 00, 06, 12, and 18Z and distributed to weather centrals and to district and local forecast offices by means of teletype, radio, and other communication facilities. Observations from the national network are distributed to the analysis centrals and the district and local forecast centers and from the secondary networks to adjacent centers and to such other forecasting offices and centrals as may be required. In most areas, reports are transmitted, distributed, and plotted as rapidly as possible. The spacing of pilot-balloon and radiosonde observations depends largely upon the budget of the various weather services.

Charts and Graphs. The composition of the several types of weather observations as well as the station model for plotting data on the basic surface weather chart may be found in publications of the national weather services. The basic surface chart usually represents a compromise between the desire for both as large a scale and as large an area as possible. If possible, it should include all areas whence weather in one form or another can reach the forecast district within 48–60 hr. A Lambert conformal conic projection (standard parallels at 30° and 60°, and scale 1:7,500,000) is recommended for middle latitudes. Larger-scale charts may be required for local, airway, and other specialized forecasting.

Pilot-Balloon Charts. These charts contain observations of wind velocity and direction for specific levels from stations spaced about 150–175 mi apart in accessible areas. The levels normally include surface, 2000, 4000, 6000, 8000, 10,000, 12,000, 15,000, 20,000, and 25,000 ft above sea level. The charts are prepared at six-hourly intervals. Some stations attempt a streamline and trough-and-ridge-line analysis at the 0900Z and 2100Z observational periods between the regular radiosonde observations.

Thermodynamic Diagrams. Radiosonde observations are taken at 0300Z and 1500Z at stations 300–500 mi apart although this spacing is insufficient for detailed

upper-air analysis. Radiosonde observations provide pressure, temperature, and moisture data and values of these elements at significant points on the ascent curve are plotted on various thermodynamic diagrams. In the United States, the pseudoadiabatic diagram is commonly used.

Constant-Pressure Charts. To provide a horizontal picture over a given area, the same information is plotted for certain standard isobaric surfaces. At most district forecast centers, these usually include the 850-, 700-, 500-, and occasionally the 300- and/or 200-mb surfaces. Weather centrals will usually also prepare the 1000- and 100-mb charts, differential analysis charts or chart and a maximum potential-temperature chart.

Pressure Charts. Pressure and the 3-hr pressure change and characteristic are entered on the pressure chart directly from the coded observation, and the 12-hr pressure change, corrected for the diurnal variation in pressure, is computed and plotted. Three and 12-hr isallobars are drawn and maxima and minima are tracked.

Temperature Charts. Charts containing current and maximum or minimum temperatures are usually plotted at 1200Z and 2400Z and the 24-hr change from maximum to maximum, or minimum to minimum as the case may be, is computed. Departures from the normal temperature are usually entered.

Other Charts. Graphs of the zonal index and miscellaneous charts, such as snow-on-the-ground and the 24-hr precipitation, are prepared for various purposes.

USE OF SURFACE DATA

Basic Surface Chart. When entry of data on the surface chart has been completed, the analyst locates and defines the fronts, draws the isobars, and determines the air masses. Procedures in general use are described by Petterssen [42, Chaps. I and XI]. The current rate of movement, and the acceleration and deceleration tendencies of fronts and pressure systems are determined, and the previous paths of high and low centers are plotted. Local and district forecast offices use weather-central analyses but district offices make such modifications as their denser network of observations, special reports, and pilot reports may indicate. Areas of hydrometeors are indicated in accordance with uniform national custom or regulation. Conclusions on the causes of current and past weather, drawn from surface maps, are integrated and reconciled with those derived from the upper-air and auxiliary charts.

Pressure-Change and Other Auxiliary Charts. The necessity for the pressure-change chart arises from the large amount of data, particularly in bad weather, entered on the surface chart for each individual station. The severe deadlines imposed on analysts and forecasters in many weather services preclude close inspection and analysis of these data at each station. The pressure chart presents in clearly defined form 3- and 12-hr pressure trends. In summer, the true 3-hr pressure change is masked by the greater diurnal change, and the 12-hr change, corrected for the diurnal variation in pressure, is more representative.

Cook [16] has collected some hundred empirical rules for use of the pressure-change chart, many, however, of only regional value. The 12-hr pressure change represents the net change in surface pressure in that period of time resulting from all the physical processes which affect the weight of a column of air extending from the surface to the top of the atmosphere. The direction and rate of movement and changes in the intensity of the allobars can be forecast with the same accuracy as the highs and lows themselves and the relationship is obvious.

Whether the various areas of cyclogenesis and anticyclogenesis are undergoing intensification or weakening can be deduced rather quickly and accurately from the pressure-change chart by comparing the last 3- and 12-hr pressure changes with the same allobaric values 6 and 12 hr before and for other time intervals. The explosive effect as a surface pressure fall reaches an area under a cold upper-air low or when the leading edge of a new pressure fall reaches a moist stationary front is well known to all forecasters. Miller [35] has noted that cyclogenesis can be recognized on the Atlantic Coast by 12-hr pressure changes when 3-hr changes are indistinct. Cyclogenesis on the east slope of the Rockies can be handled much better on the basis of 3- and 12-hr katallobars than by actually following pressure systems as they move out of the Gulf of Alaska.

A snow-on-the-ground chart is necessary for forecasting the rate of modification of air masses as they move from snow cover to bare ground or vice versa and for the forecasting of maximum and minimum temperatures and drifting snow.

Some forecast centers maintain a graph of the zonal index, plotting daily values or using mean values provided by the extended-forecast section. The zonal indices provide a good indication of the broad-scale features of the general circulation and assist in forecasting changes in intensity or position of the centers of action. Attempts to correlate variations of the zonal index with the formation of arctic air masses have, so far, been unsuccessful and attempts such as that of Weiss [57] to determine the relationship, suggested by Rossby [47], between a large-scale flow of warm air northward in the eastern Pacific Ocean toward Alaska and marked anticyclogenesis over the Mackenzie Basin have generally failed. In this particular situation, anticyclogenesis appears to take place at more southern latitudes. Namias [36] has described a case where a general fall in index over Asia was followed by the progressive establishment of low-index conditions from west to east resulting in strong polar anticyclogenesis. It has been known for some time that an analogous condition may be initiated in Europe, developing westward into North America and resulting in the condition generally referred to as "blocking." Indeed, exceptionally strong and persistent arctic anticyclogenesis seems to form more often in the latter than in the former manner. These important trends are not often available to the district forecaster but fairly short-term forecast effects arise from them since it is not unusual for an arctic

front to accelerate from an almost stationary condition to 40–60 mph within 12–18 hr.

Temperature-change charts have relatively little forecast value but are required for weather bulletins for press and radio, for preparation of shippers' forecasts and other information for commercial and agricultural interests, and for the general public.

Analogues. The system of locating a number of previous weather charts which are most nearly analogous to the current chart, and then determining the prognostic chart by analogy, has been used successfully by a number of forecasters, particularly J. J. George. The past maps are known as "analogues." Weather charts for several decades are catalogued on the basis of their dominant characteristics, for example, Colorado lows or Alberta highs. If, for example, on some particular December day those are the two principal characteristics of the current weather chart, the December, and also the January and November, file of past situations is inspected and the five to ten charts with greatest similarity to the current chart are selected and studied. A choice of two or three is finally made, primarily on the basis of history and secondarily on the basis of detail. "Good" analogues should provide reasonably good prognostic charts since the numerous hydrodynamic and thermodynamic parameters that describe the current weather situation are conveniently integrated in the analogue. Where selection of the analogue is made mechanically (*i.e.*, by machine) the system has been a complete failure, since historical sequence is much more important than actual location and intensities of pressure systems. Manual selection, which permits consideration of both surface and upper-air patterns, does provide better results but time and patience are required. Because of the persistence of weather types, the best analogue may often be found from the charts within the three or four weeks previous to the current chart. The purpose of the analogue procedure is to provide the inexperienced forecaster with assistance, but, in practice, he does not have the experience required for the selection of the best analogues and for the recognition of the significant differences which are always present between the analogue and the current chart. The use of analogues will never prove satisfactory until the upper-air patterns and historical sequences are included in the parameters upon which their selection is based.

Types. Attempts have been made to type weather maps and situations since the beginning of forecasting. Many of these attempts have met with limited or no success because of the infinite variations of weather situations and the superficiality of some methods of typing.

Studies of storm types, frequencies, and normal storm tracks, such as those of Bowie and Weightman [7] in the United States and of Van Bebber [54] and Braak [8] in Europe, are widely used by forecasters. The San Francisco forecasters were moderately successful for many years in forecasting for the Pacific Coast by dividing their weather into three main steering types: northwesterly, westerly, and southwesterly. This typing was expanded and refined by the Meteorology Department at the California Institute of Technology under the leadership of I. P. Krick.

The C. I. T. typing method is based on the location of the "center of action," here the semipermanent eastern Pacific high. Krick, Elliott, and colleagues [12] have classified North American weather into five main types, with a considerable number of subtypes, depending upon the location of the Pacific high. It was found that a type tended to persist for six days (five to seven days) and then to repeat or change to another type. Composite charts, by seasons, were derived for each of the six days of each type. As indicated by Elliott [20], changes in index and in the various arrangements of the large upper-level waves (meridional flow patterns) and different degrees of expansion of the ring of strongest westerlies (zonal flow patterns) are regularly associated with the shift and change in intensity of the Pacific high. While typing of this kind is most helpful in extended forecasting, it has also been found of value in 24–72-hr forecasting as far east as central United States. It is understood that further investigations are under way to develop composite upper-air maps analogous to the composite surface charts and to determine the limits of variation of upper-air situations with individual surface types. The principal limitations of the type technique are (1) the frequent and extensive variations of a single type, and (2) the difficulty in recognizing some types in complex situations. Typing on a somewhat more circumscribed scale, such as that of Saucier [49] in connection with Texas-West Gulf cyclones, can provide the district forecaster with a valuable tool.

USE OF UPPER-AIR DATA

Radiosonde Diagrams. For a number of reasons, including the requirements for the preparation of certain specialized and detailed forecasts for very short-term periods as well as thorough analysis of local weather conditions, it is desirable and necessary to make a detailed analysis of individual radiosonde observations within and adjacent to the forecast district. The analysis enables the forecaster to locate cloud layers and their thicknesses as well as areas of freezing rain and snow (when present) and also to ascertain stability conditions in order to forecast the probability of cloud, cloud heights, and convectional shower activity. For convenience, the forecaster may wish to classify the lower troposphere according to the customary degrees of stability, namely, stable, conditionally unstable, and absolutely unstable. During certain periods and seasons, the forecaster will wish, as a matter of course, to determine the lifting condensation level, the convection condensation level, and the energy available. For a full discussion of stability and instability see any standard textbook, particularly Petterssen [42, Chap. II].

Two methods exist for determining the stability of the atmosphere: the *parcel* and the *slice* methods. The parcel method is less exact in that as a parcel of air rises, other air is entrained from the outside and when the air is conditionally unstable (when the exact degree

of stability is most important to the forecaster) the parcel method underestimates the resistance against lifting and overestimates the available energy. The slice method, first suggested by J. Bjerknes, does take into consideration changes in environment of the ascending air, but the method is too complicated for practical use.

Constant-Pressure Charts. At district forecast centers, constant-pressure charts are normally prepared for the 850-, 700-, 500-, and occasionally for the 200-mb surfaces. Fulks [22] has described procedures for the preparation of these charts which normally include entry of dry-bulb and dew-point temperatures, height of the pressure surface, and 12- or 24-hr height changes. Contours (isohypses) and isotherms are then drawn. At many centers, isolines of equal height changes are drawn and areas of certain moisture values, based on the dew points, are shaded. At weather centrals, other pressure surfaces may be useful, particularly the 1000-mb chart, as well as tropopause and density charts.

The 850-, 700-, and 500-mb charts are among the finest tools available to the forecaster. A large number of techniques and rules, mostly subjective, have been developed or formulated during the past few years for use in forecasting.

Constant-pressure charts, and the 850-mb chart in particular, can provide the forecaster with much of the information yielded by the isentropic chart. Since the isotherms are lines of potential temperature, they represent the intersections of isentropic surfaces with constant-pressure surfaces. Means [33] has found the warm air advection in the lower layers of the atmosphere (2000–8000 ft and above m.s.l.) useful in forecasting thunderstorms.

By far the greatest amount of work has been done on the 700-mb chart. During the past few years, however, the 500-mb chart has been finding greater favor with forecasters, since the field of motion is more conservative at this level, and for other reasons. The most important use of the 700- and 500-mb charts is the "steering" of surface highs and lows or pressure rises and falls indicated by the field of motion at these levels. "Steering" appeared prominently in the publications of German meteorologists in the 1930's particularly in the work of Baur [3], who found that the direction of motion of regions of rise and fall of pressure was controlled by the steering at 5 km. The mean duration of a broad-weather situation (*Grosswetterlage*) was 5 to 5½ days which, with the allowance of one transitional day, corresponds to the C.I.T. six-day type. Steering was broken down into four divisions:

1. Westerly steering (high index)—about 18 per cent of the cases.

2. Northwest and southwest steering (moderately low index and moderate troughs and ridges)—about 17 per cent each.

3. Northerly and southerly trough and ridge steering (very low index)—smaller percentages.

4. Easterly steering (very low index)—observed very infrequently. Presumably a number of cases could not be classified.

V. J. and M. B. Oliver [40] have summarized a large number of rules obtained from many sources for the use of the upper-air charts. Most of these rules have not been objectively tested but some of the more important of them, applicable to the 700-mb chart, are:

1. Surface cyclones move in the direction of the 700- and/or 500-mb flow.

2. Surface cyclones move in the direction of the 700–500-mb mean isotherms, inclining slightly toward the colder air.

3. If an upper isallobaric maximum (24 hr) is found in the direction in which the surface cyclone will move, the cyclone will move into the region, or just to the west of it, in 24 hr.

4. Surface cyclones will move along a line from the center of the isallobaric minimum (24 hr) in the rear to the center of the isallobaric maximum in front.

5. The smaller the angle between the upper isobars and the mean isotherms of the low troposphere, the closer the speed of the cyclone approaches the speed of the upper flow. But perhaps the most important and widely accepted rule in current use states: The surface low or surface katallobar moves with approximately half the speed of the 500-mb wind over it.

Table I. Average Deviation of the Direction of Movement of Cyclones from the Orientation of the Upper-Level Contours and Isotherms

Upper-level pattern	Comparison A		Comparison B	
	Average deviation	Number of cases	Average deviation	Number of cases
Contours, 850–700 mb	31°	212	24°	92
Contours, 700–500 mb	28°	216	21°	92
Isotherms, 850–700 mb	31°	215	23°	92
Isotherms, 700–500 mb	33°	218	23°	92
Contours, 200 mb	26°	159	21°	92

The forecasting value of rule 1, if valid, is obvious. Austin [1] has tested this rule by comparing the direction of movement of the cyclone on the surface with the orientation of contours and isotherms at various levels. This direction of movement was assumed to be given by the direction from the position of the center 12 hr before to the position 12 hr after the time for which the comparison was made. Austin's conclusions are summarized in Table I. Comparison A includes all observations, while Comparison B includes only those cases in which the speed of the cyclonic center exceeded 20 mph.

Table II. Frequency Distribution of the Deviations in Direction

Upper-level pattern	−180° to −46°	−45° to −16°	−15° to +15°	+16° to +45°	+46° to +180°
Contours, 850–700 mb	1	5	48	27	11
Contours, 700–500 mb	2	7	60	12	11
Isotherms, 850–700 mb	4	13	51	18	6
Isotherms, 700–500 mb	3	7	49	22	11
Contours, 200 mb	3	14	50	17	8

The frequency distribution (Table II) shows the spread of the deviations for the 92 cases in Comparison

B. A positive deviation means that the cyclone moved to the right of the upper-level contour or isotherm. The mean deviation was reduced to 19° by averaging the direction of contours in the layer from 700 mb to 500 mb with the direction at the 200-mb surface. Austin also found that a cyclone which was being steered tended to continue being steered, that the deviation appeared to be independent of intensity, that approximately 65 per cent of the cyclones moved to the right of the upper-level isotherms and contours, that the difference between filling and deepening cyclones was negligible, and that the average deviation was greater for the Rocky Mountain plateau than for the eastern United States. Forecasters will agree with Austin's conclusions that the forecasting principle of steering compares favorably with other prognostic procedures for determining the direction of movement of a cyclone and, indeed, is the best, although definitely not a precise tool, for this purpose. In this most useful study it did not appear that the speed of a cyclone could be determined from the geostrophic wind speed at any particular level, and it is believed that the poor correlation obtained came as a surprise to most forecasters.

In a somewhat similar study of cyclones and anticyclones, Longley [32] found the deviations given in Table III.

Table III. Mean Values for Classes Grouped According to the 700-mb Flow

Class	Number of cases	Mean deviation	Mean 24-hr movement (mi)	Mean deviation distance (mi)
Cyclones				
Closed centers aloft	87	44°	468	332
Troughs and ridges aloft	36	20°	537	310
Straight contours aloft	47	17°	628	295
Anticyclones				
Closed centers aloft	46	65°	311	295
Troughs and ridges aloft	30	37°	515	400
Straight contours aloft	33	27°	587	328

The frequency distribution of the deviation angle derived in Longley's study, which, for cyclones, is in good agreement with Austin's, appears in Table IV.

Table IV. Frequency Distribution of the Deviation Angle

	+180° to +50°	+45° to +20°	+15° to −15°	−20° to −45°	−50° to −180°
Anticyclones	14	13	38	20	24
Cyclones	22	35	82	22	9

Longley found that, of 47 anticyclones with a deviation of 45° or more, 33 were located over the Atlantic along a line approximately from 45°N and 30°W northeastward to 55°N and 20°W, and that a greater concentration of cyclones with large deviations existed in a rectangle bounded by the 40°N and 50°N parallels and the 50°W and 70°W meridians, or as the pressure systems approached the Atlantic high and the Iceland-Greenland low. Longley differed with Austin in regard to the tendency of a cyclone to continue to be steered if following the upper-air flow but agreed that the wind velocity at the 700-mb surface was not a good indication of rate of movement of the surface cyclone.

Palmer [41] obtained as good or better results on direction of movement by a simple graphical device employing the following parameters:

1. Normal 24-hr direction of movement based on data published by Bowie and Weightman [7];
2. The direction in which the cyclone moved during the past 6 hr;
3. The orientation of the major trough in the sea-level pressure pattern; and
4. The direction from the 3-hr anallobaric center behind the cyclone to the 3-hr katallobaric center ahead of the cyclone.

Tests indicated that this tool would be correct within 15° about 75 per cent of the time. Palmer considered only winter storms; summer lows might not yield as good results.

The procedures employed by both Austin and Longley perhaps do not provide a wholly fair test of the validity of the steering principle, but tests should be made of all forecasting techniques and rules which are held in high regard by all forecasters. In general, forecasters simply do not know the accuracy of many of the rules and techniques in use or the range of the possible deviation of nearly all of them.

In applying the results of Austin's study, the forecaster is faced with the determination of the probable direction of deviation from the steering indicated by the 700- and 500-mb charts.

The deeper the low, the less applicable the steering principle. Surface lows which have closed centers at 700 mb and higher tend to move with the upper center. The best indication of the direction of movement of closed lows at the 700-mb surface and higher is the 12-hr pressure change around the center; another method is the determination of the resultant of the wind circulation around the upper-level low. The calculation of the distance such a low will move in a given time is even more difficult and should be determined by the broad-scale weather processes going on.

Little has been written on the practical applications of the 200-mb chart to forecasting. Wulf and Obloy [59] have emphasized the importance of the compensating effects between the stratosphere and the lower troposphere and have suggested that stratospheric conditions may exert considerable influence upon frontogenesis and that advection and other processes in the lower stratosphere should receive consideration in prognosticating surface and tropospheric pressure patterns. Austin [1] found that the 200-mb chart is almost as effective a steering level as any other. Some forecasters have found the 200-mb chart helpful in determining displacement of shallow migratory troughs at the 700-mb level.

Under the tropopause, at approximately the 300-mb level, an extremely fast and narrow current has been noted which is called the "jet stream" by the University of Chicago group [53]. Large-scale mixing, interrupted

at a critical zone in middle latitudes, has been suggested by the Chicago group, and the "confluence" theory by Namias and Clapp [38], as possible explanations of the mechanism of this current. Which are the causes and which are the effects have not been definitely established. Riehl [45] has suggested that the ascent of tropospheric air below and slightly to the north of the jet might affect the distribution of precipitation. Starrett [52], investigating this relationship, found a significant concentration of precipitation activity under the jet stream, subject to the normal variations of local dynamical fields of convergence and divergence associated with the short baroclinic waves within the westerlies.

It seems probable that the jet stream and George's "isotherm ribbon" [26] are all manifestations of the same phenomenon, namely, the frontal zone, and that they are usually located along the same latitude at any one time and that this zone is favorable for cyclogenesis.

Pilot-Balloon and Other Upper-Air Charts. The principal use of the 6-hr pilot-balloon charts is obviously the forecasting of upper-air winds and, to a much lesser extent, surface winds. The interim charts between the regular constant-pressure charts are useful in checking the movement of troughs and ridges and the latest trends.

Isentropic analysis was introduced around 1937 by Rossby and collaborators [48]. Isentropic charts were prepared generally by forecast centers for a number of years in accordance with procedures outlined by Namias (see Petterssen [42, Chap. VIII]), but were eventually abandoned when forecasters failed to find in them the hoped-for precise tool for precipitation forecasting. The poor results have been blamed on the nonadiabatic processes common in the atmosphere, which tend to destroy the conservatism of isentropic surfaces, principally (1) radiational cooling and heating, (2) evaporation and condensation, and (3) convection. Also one thin, and often wavy, isentropic surface, arbitrarily selected, may be unrepresentative of the principal upslope area. In the Middle West, determination of the motion of air particles, some distance away from but in line of flow toward a station, relative to the contour lines of the isentropic surface over the station, was frequently very difficult. Contours might move in the same direction and at the same rate as the air particles and the expected upslope movement would not occur, with resultant failure in the precipitation forecast. There is some question whether isentropic analysis has been given a fair trial by forecasters, since (1) time was rarely available for careful construction and analysis, and (2) many forecasters never fully understood the meaning and significance of all the information on the isentropic chart.

Clouds. With the considerable increase in the amount of upper-air information now available to the forecaster, the importance of cloud data has decreased somewhat. However, as Brooks [10] states: "Clouds are indications of humidity, lapse-rate, (and temperature), direction and velocity (including shear and turbulence) in the free air. As such, they are, of course, valuable aids in air-mass analysis and forecasting." Thus, as indirect aerology, clouds provide the local forecaster particularly, and the district forecaster and analyst as well, with valuable information regarding the stability and moisture trends of the lower atmosphere. A close watch on the clouds as shown on the hourly sequences and interim charts will provide the forecaster with valuable clues on the rate of moistening of a previously dry trough.

PREPARATION OF PROGNOSTIC CHARTS

It is desirable to prepare, in a formal fashion, prognostic charts for 24-hr intervals. These charts should include isobars, frontal positions, high and low centers, and precipitation areas. The positions of fronts and pressure systems can be indicated informally on the regular six-hourly synoptic chart for six or twelve hours or for any other time interval. Upper-air prognostic charts should include contours and ridge and trough lines.

Preparation of the 700-mb Prognostic Chart. Because of their complexity and the time involved, the preparation of upper-air prognostic charts is logically a function of weather centrals and not of forecast centers. Techniques have been suggested by Starr [51] but a description of those in use by American meteorologists has not, as yet, been published for general distribution. However, the principal procedure appears to be extrapolation with a check for consistency by drawing the implied mean virtual temperature between the 1000- and 700-mb surfaces. The forecasting of mean virtual temperature and mean density patterns should provide a better indication of future height changes than pure extrapolation of changes at one level. Other techniques which might be used include:

1. Extrapolation of trough and ridge lines. The pattern at 700-mb is more conservative than at the surface and conservatism increases further with height.

2. Extrapolation of isallobaric centers, applying corrections for indicated intensification or diminution.

3. Application of the formula for the movement of long waves in the westerlies, developed by Rossby [47]. This formula is $c = U - \beta L^2/4\pi^2$, in which c is the speed of the wave, U is the zonal wind speed (best results are apparently obtained at approximately 600 mb), L is the wave length, and β is the rate of change of the Coriolis parameter northward with latitude. Recent studies by Cressman [17, 18] indicate that good to excellent results may be obtained if adequate daily 700-mb charts are available on a three-quarters or full hemispheric scale. Use of this formula is not practicable for an area as small as North America.

4. Isotherm-isobar relationship. Another principle based on the kinematics of wave motion provides some indication of the rate of movement of minor waves in the westerlies. Assuming that there is no convergence or divergence and that temperature is a conservative element, Rossby [46] derived the equation $U/(U - c) = A_T/A_p$ in which U and c have the same significance as before, A_T is the amplitude of the isotherms, and A_p is the amplitude of the isobars. It can readily be

seen that when the isobars and isotherms are in phase but the isotherms possess the larger amplitude, the wave will move with moderate velocity. In the same case but with isobars of an amplitude larger than the isotherms, the perturbations will move slowly or even retrograde. When the isobars and isotherms are 180° out of phase, the wave will move very rapidly. When cold air moves into the system, the wave will tend to deepen; when warm air moves into the system, the wave will usually weaken.

5. Minor waves tend to intensify as they move into a major trough although maximum intensity is usually found as they reach the midway point between the major trough and the major ridge downstream. Similarly, a pressure fall intensifies as it moves into a major trough and a pressure rise intensifies as it reaches a major ridge. However, care should be exercised if the pattern is changing.

6. Use of pressure tendencies at upper levels. Miller and Thompson [34] devised a method for the calculation of 3-hr pressure changes from pilot-balloon observations. It was hoped that these pressure tendencies might permit the use of Petterssen's extrapolation formula, but this has not become standard practice largely because of the work involved and because only advection is considered, while convergence and vertical motion (and any nongradient winds) are omitted from consideration.

7. Rossby's trajectory method. Rossby has developed a trajectory technique which is essentially the determination of the future path of a system of particles based on the change in vorticity with change of latitude while the absolute vorticity is conserved. Fultz [24] has described the technique and the attempts which have been made to apply it to forecasting the 10,000-ft chart. Tests have shown that the 10,000-ft trajectories can be computed by this method for winds in major flow patterns with a significant degree of accuracy up to 72 hr. Convergence and divergence are neglected in the equation and subjectivity is required in the selection or rejection of the computed winds. In certain situations wave computations are difficult and the constant absolute vorticity trajectory methods give better results. Use of this technique in extended-range forecasting is described by Namias [37].

8. Supergradient wind velocities. It has been noted that supergradient wind velocities in the eastern Pacific have often been attended or followed by an increase in pressure over the far western United States. The supergradient condition results in a net unbalanced force on each unbalanced particle directed toward the right and thus in a piling up of air on that side. This has been observed in other regions and has become known as "anticyclogenesis upstream."

9. "Cutoffs." Occasionally a rise of pressure will appear behind a trough, move eastward in the maximum westerlies and cut off the southern portion of the trough resulting in a closed circulation in that area. This tends to happen in preferred locations, for example, in the far southwestern United States. The closed low will either move southwestward off the coast of southern California and usually dissipate or remain stationary for around 48 hr awaiting the arrival of a new fall in pressure from the north. Somewhat similarly, warm highs may be cut off. Closed lows may develop in almost any section, usually building upward from the surface, and their prognostication is very important because of the extensive bad weather associated with them. This development may be detected 12–24 hr in advance when a slowly moving isallobaric minimum is located south of a weak westerly gradient.

For the most part, the 700-mb prognostic chart should be prepared independently of the surface prognostic chart since one of its principal uses is as a check on the latter. However, there may be a few elements of the surface prognostic chart which will be subject to little or no question and the 700-mb chart should be consistent with respect to these elements.

The usual order of operations followed by the analyst making the 700-mb prognosis in the Weather Bureau-Air Force-Navy Analysis Center[1] is as follows:

> The prognosticator makes a 700-mb analysis, and then draws a thickness chart (1000–700 mb) which also includes the 850-mb flow. While the thickness chart is being made, an assistant makes the 24-hr height change chart. A tentative prognosis is then made by application of the various techniques discussed in the following paragraphs. Then, in consultation with the surface prognostic analyst, a series of checks and adjustments of the two prognoses is made to make them mutually consistent, with special attention given to surface patterns and to trends determined from closer examination of old data or from a check of new three-hourly data. This informal discussion of the general situation is followed by a discussion of the points of difference or of inconsistency between the 700-mb and the surface charts and is participated in by the two prognostic analysts and the supervising analyst. This leads to the final version of the prognosis.

No one procedure or set of procedures is used regularly in making the 700-mb prognosis. Of the wide variety available, the choice depends upon the analyst and the occasion. Good judgment is of paramount importance in making a prognosis. The analyst relies on it in determining rather quickly to what extent pure extrapolation and persistence can be used with better results than a time-consuming physical forecast. Attempts to explain the physical basis for the processes going on in the atmosphere lead to involved discussions and arguments so that such attempts are limited to those features which are worth the valuable time required for them. Those features which, from the analysis, provide indications that extrapolation is a good prognosis, plus those features which show definite indications of persistence, frequently make up a large portion of the chart.

The semipermanent and characteristic points and lines serve as a starting point for the prognosis. Persistence is usually the best forecast in regard to the semipermanent systems: the Aleutian low, the oceanic highs, etc. However there is always the question of whether to move them or to fill or deepen them. A deep cold low, dominating a large area of the map, and the accompanying peripheral flow determine not only the features of the area of the low but also the course of any perturbations coming into the belt dominated by the

1. The author is greatly indebted to Mr. Charles M. Lennahan and Mr. J. R. Fulks of the Analysis Center for the outline and description of the procedures used there for the preparation of the 700-mb prognostic charts.

peripheral flow. The speed of these perturbations may also be inferred so that the pattern of the region in the vicinity of their prognosticated locations may be deduced. In this way persistence not only controls its own area but extends its influence and prognostic value to adjoining regions and, with diminishing influence, to more distant regions. The dominant warm highs have a similar role in the prognosis since they persist and tend to be blocking highs. Any cold trough which is strong enough to move a semipermanent warm high is usually so obviously strong that at least a slight movement of the high will be included in the prognosis.

As in all forecasting, the first step in pure extrapolation is to find some element that is conservative enough on the particular occasion to be used as a starting point. This element may very well be different from the one used on the previous day and may even be different *each day* for a week or more. Thus the speed of the front, the speed of the low, the speed of the isallobaric trough, etc., may be different from the corresponding points and lines of a high or ridge, and one may be conservative when the others are not.

With a sequence of good analyses as a basis, the prognosticator extrapolates the conservative characteristic points and lines, at uniform speed if they have been moving uniformly—with positive or negative acceleration if they have shown signs of accelerating. Trends have to be checked closely so that the beginning of transition periods between high- and low-index conditions can be detected and the prognosis adjusted to conform to the increase or decrease in displacement. A blocking high, whether one of the warm oceanic highs or a cold high over the continent, has to be considered in its effect on the flow.

The complex interrelations between sea-level isobars and 700-mb contours make any feature on the sea-level chart important insofar as it affects the 700-mb contours. Conversely, since one of the forecaster's main interests in the 700-mb prognosis is its value as an aid in forecasting surface wind, weather, and temperature changes, any important feature on either the surface or the 700-mb chart is usually an important feature on the other. It should not, however, be implied that the 700-mb prognosis is strictly dependent on the surface prognosis; often the reverse is true because some features aloft can be prognosticated with greater certainty, and in such cases the upper-air prognosis will influence the surface prognosis.

Because of the close coordination required between the surface and the 700-mb prognoses, it is difficult to separate completely the factors used in making the one from those used in making the other. One of the best aids that the person responsible for the 700-mb prognosis can have is a competent surface prognosticator working with him. A good surface prognosis is the best way to point up the inconsistencies in the 700-mb prognosis.

The 700-mb prognosticator must be familiar with the sea-level analysis and this familiarity influences him not only in the movement of systems at 700 mb but also in the deepening and filling of the systems. The 700-mb and sea-level analyses and prognoses require parallel and coordinated treatment. It is inherent in the whole prognostic procedure that the surface chart tends to be the dominant one, because the surface network of stations is much denser and provides more frequent reports than the upper-air network. Therefore, there is a tendency to start from the surface when making a 700-mb prognosis, and final revisions are usually made on the basis of the latest three-hourly map. Any system which is important enough to be included in the 30-hr surface prognosis is of sufficient magnitude to affect, even though slightly, the 700-mb contours.

The close connection between the sea-level pressure pattern and the 700-mb surface makes many studies of sea-level systems applicable indirectly in the prognostication of 700-mb contours. Thus Bowie and Weightman normals, C.I.T. types, Palmer's method, analogues, etc., all contribute to the 700-mb prognosis. However, here again good judgment is the prime factor; there is always the question of how much such objective measures should be modified or ignored. They cannot be used blindly. Considerable qualitative use is made of the characteristic behavior of storms in a typical sequence. The path and speed of the storms, the isobar and contour patterns, all contribute to the final version of the prognosis.

Any deepening of the surface systems will usually cause a corresponding deepening of the 700-mb systems. Therefore, any kinematic or dynamic effects which are known from experience or from theory to favor deepening of surface systems will have to be considered in making a 700-mb prognosis. Consideration must be given to such kinematic processes as the converging of two katallobaric systems or of the movement of a low in such a direction as to take it under progressively lower contours at 200 mb, and to such dynamic processes as the converging of fresh cA and mT air masses.

Sometimes an inactive front (the most pronounced cases are on the Texas and Louisiana coasts) will become very active and rapid cyclogenesis will take place on it because of the approach of a cold trough at 700 mb or higher and the subsequent coupling of the cold trough with the sea-level cold front. This coupling has the effect of steepening the slope of the cold front and of increasing its vertical extent. As a result, the total mass of air involved and the total temperature difference are much greater than those involved in the original front. The deepening storm quickly affects the 700-mb contours and a very important feature of the chart is created. These storms usually move fast since they are generated in well-developed upper troughs.

A frequently recurring feature is the movement of sea-level lows and highs into the region of the mean trough. This process usually causes a mutual deepening of the sea-level and of the 700-mb troughs; similarly it causes weakening of the sea-level high and a corresponding weakening of the 700-mb ridge.

Frequent use is made of the 200-mb contours in forecasting the deepening of surface lows. Often the high-level contours will be only slightly influenced by the surface system, and if the future position of a surface low is known, it is found from experience that when the low moves into a region of lower 200-mb height, it will deepen by approximately 60 per cent of the amount which would result from projection onto the ground of the change in height of the 200-mb level above the surface low. This effect causes a corresponding change in the 700-mb surface.

Contrary to the general rule that a fast-moving low will not deepen, lows in well-established troughs with a strong gradient on the warm side will move rapidly and if the movement has a good component northward, the storm will deepen. This feature of movement or steering with the upper winds is generally known but it has many modifying factors.

Cold air moving into the region to the rear of an upper-level trough will tend to cause the trough to retrograde. The timing of the action of the mechanism is of the utmost importance. If the initial trough is in a position where it will move (*i.e.*, where topographic and kinematic reasons favor its continued movement), the cold air will tend only to make the flow more westerly and cause the trough to move faster. Ordinarily, however, the cold air will cause the trough to move more slowly or even to retrograde.

Frequently supergradient winds are noticed, or strong winds

in a field where the contour pattern is expected to move slowly with respect to the winds, so that the strong winds will soon reach that part of the field where the gradient is much weaker than would be necessary to balance the velocity of the wind. In these cases the winds will tend to turn to the right because the Coriolis force exceeds the pressure-gradient force, and air will cross the contours toward higher values with a consequent decrease in speed. Usually the speed *decreases* more than enough to balance the gradient because the momentum of the parcel carries it beyond the point of equilibrium so that the parcel is turned to flow toward lower heights with an *increase* in speed. This cyclonic trajectory has the effect of forming a new trough or of deepening an already existing trough and intensifying the adjoining ridge; it also causes stationary or even retrograde conditions to prevail in that area of the map.

The use of the jet stream in the prognosis is very qualitative. Although the jet is at a much higher level, its effects are apparent in the stronger flow induced at 700 mb. Thus from a consideration of the confluence theory of jet-stream formation, it is apparent that any system moving into the vicinity of the strong flow will accelerate and usually will not deepen. Thus the effects of these storms on the 700-mb contours are limited to their contributions to the confluence and the speed of the jet. Also any storm moving in the vicinity of the jet will not cross it but will tend to move with it.

The 24-hr height changes at 700 mb are used in much the same way as sea-level pressure changes. Here again, however, the interrelation between the 700-mb and the sea-level charts must be studied in order to determine some criterion for using the height changes rationally. Any conservative feature common to both charts, whether of movement or of configuration, is helpful in deciding how much and where to apply the changes.

Along with the dominant role played by the surface analysis and the surface prognosis in the making of the 700-mb analysis and prognosis, it is important to note that the thickness chart constitutes an important check on the consistency of the prognoses. (The 1000-mb prognosis is quickly made by direct extrapolation from the sea-level prognosis.) Thus in the absence of strong vertical motions over the prognosticated area, the temperature field should change in the direction and with the speed of the flow in the "thickness" layer. The thickness lines are usually more conservative than the contours or than the sea-level isobars.

Lennahan concludes that it would certainly be pleasing to all if some good theoretical or objective method could be developed for preparation of the prognoses. However, for the present at least, it seems that we must be satisfied with our "cut and try" methods for short-term forecasts. In a review of Scherhag's new book,[2] E. Hovmöller remarks: "A theoretical meteorologist would probably, after reading this section [G of first part], feel discouraged once more by the contrast between the magnificent building of theory itself and the modest, almost crippled part of it which is thought to be applicable in the weather service." It is unfortunate that this is also equally true of so much of the recently developed meteorological theory.

Although, quantitatively, the 700-mb techniques appear to be equal to those available for the surface prognostic charts, and the higher-level chart is somewhat more conservative, most forecasters believe that the 700-mb prognostic charts are not as satisfactory; however, this belief is difficult to substantiate. The principal weaknesses are the forecasting of the development of closed lows and their future movement, the movement and changes in intensity of minor troughs, and the emergence of major troughs from mean troughs. The reasons for these inadequacies may be (1) nonutilization of all techniques, (2) lack of experience, (3) preoccupation with surface prognostic charts and with mean troughs and ridges, and (4) lack of satisfactory techniques for dealing with closed circulations. There is also considerable evidence of confusion in defining major, mean, and minor troughs. When does a minor trough become a major trough? The consideration of major troughs and ridges as synonymous with mean troughs and ridges by many forecasters and prognosticators is believed to be erroneous. There appear to be certain large-scale controls, not very well understood, which during the persistence of any given regime tend to result in trough (ridge) formation or intensification (weakening) repeatedly at some one geographical location. So long as the regime persists—and it may change suddenly—pressure falls (rises) intensify (weaken) as they approach the region of the mean trough (ridge). The result is a mean trough or ridge in the favorable location, but troughs or ridges may move out of the mean positions and, apparently depending upon the wave length, become migratory major troughs or ridges. The crux of the forecasting problem in such cases is whether troughs and ridges emerging from the mean positions will attain major intensity or remain minor waves.

Recent work by Charney [13, 14] and collaborators gives some promise that the new electronic computing devices will eventually provide assistance in prognosticating constant-pressure surfaces.

Preparation of the Surface Prognostic Chart. In the preparation of the surface prognostic chart, the pressure field is given primary consideration since an accurate pressure prognosis is the synthesis of all computations by, and experience of, the forecaster. The customary procedures used in the preparation of this chart include:

1. Determination of the movement of fronts and pressure systems. The six-hourly positions of significant fronts and highs and lows are plotted for the past 12 to 24 hr or more, and changes in direction and rate of movement are carefully noted. As a starting point, highs and lows might be typed in accordance with, and the 24-hr average direction and rate of movement ascertained from, some previous statistical study such as that of Bowie and Weightman [7] for the United States. From the previous history of the system, an immediate deduction can be made whether it is operating under the average steering indicated by the study. A second position may be obtained by modified extrapolation, sometimes called the "path" method. The six-hourly projected positions should be further modified in accordance with changes indicated by the general synoptic situation. Another predicted frontal position

2. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose*. Berlin, Springer, 1948. Reviewed in *Tellus*, Vol. 1, No. 4, pp. 70–74 (1949).

can be obtained by the geostrophic wind method, which usually yields good results with cold fronts but must be used with caution with warm fronts. Petterssen states that warm fronts will move with 60–80 per cent of the geostrophic wind, while Byers in the North Pacific observed warm fronts moving with 50 per cent or less of the normal component of the gradient wind.

Further determination of the movement of fronts and pressure centers should be obtained by the Petterssen extrapolation formulas [42, Chap. IX]. These formulas give excellent results over oceanic areas, good results over homogeneous land areas, but rather poor results over mountainous sections. Since the movement of fronts and pressure systems is almost always changing with time, the formulas rarely hold good longer than 24 hr. After that time, the forecaster must rely on the normal evolution of pressure systems and the large-scale factors indicative of acceleration or deceleration. The direction and rate of movement of pressure systems as indicated by the steering shown by the upper-air charts—current and prognostic—should be determined. In computing the displacement of pressure systems, the forecaster should be careful that the estimated displacement of each individual system agrees logically with the simultaneous displacement of adjacent systems. In the United States, errors occur most frequently when (1) rapidly moving occlusions, in a period of high index, cross the Pacific coastline with at least moderate intensity but become damped out by the time they should have reached central North America, and (2) a storm with a marked allobaric minimum reaches the North Pacific Coast during a period of moderate or high index simultaneously with the development of a rather strong but poorly organized depression in the Great Plains and the Mississippi Valley. In the latter case, marked anticyclogenesis sets in immediately over the eastern Rockies and the Great Plains and the low frequently moves off more rapidly than expected.

A forecast rule in general use states that wave cyclones will move parallel to the warm-sector isobars when the latter are parallel to the isotherms and this movement is not inconsistent with the pressure tendencies. As a rule, in a family of lows, each low will develop a course farther south than its predecessor but, in strong northwest steering, successive lows usually trend toward a more northerly course.

Thus, a number of approximations of prognostic positions of fronts and pressure systems can be obtained. The results should be compared and a final approximation derived which, largely on the basis of experience, appears most logical from the physical processes evidenced by the most recent observations.

2. Forecast of changes in intensity of pressure systems. The forecaster will, of course, keep in mind the normal deepening and subsequent filling of the typical cyclone as it passes through its life cycle. Cyclones developing in middle latitudes will be attended by maximum deepening when the northward component in direction of movement is greatest. The relationship between direction of movement, intensity of meridional flow, and availability of moisture is obvious. According to Petterssen, the rate of deepening remains constant as long as a warm sector remains on the ground and for 6–12 hr after occlusion sets in. Changes in the three- and twelve-hourly pressure tendencies should be watched carefully since they provide a basis for estimating the rate of deepening and filling.

Intensification or weakening of pressure systems can be detected by inspection of central pressures in the systems on the last several regular and interim synoptic charts, and by the position of the maximum isallobar relative to the center, correcting for the diurnal variation in pressure.

Some significant rules for determining filling or deepening have been summarized by V. J. and M. B. Oliver [40] as follows:

a. A wave will deepen or a front become more pronounced if the 10,000-ft wind field possesses cyclonic vorticity and the wave has a temperature contrast through it.

b. A wave will weaken and a front will undergo frontolysis if the 10,000-ft wind field possesses anticyclonic vorticity.

c. If there are several waves along a front, the one with the most intense cyclonic vorticity aloft will develop at the expense of the others. This is usually the one nearest the axis of the trough aloft (at 10,000 ft).

d. Waves at the surface will deepen if the 700-mb contours diverge ahead of them

e. Waves at the surface will weaken if the 700-mb contours converge ahead of the wave.

Austin [1] found no definite relationship between the lapse rate of temperature above the center of cyclones and their future change in intensity. In the same study he tested changes in cyclone intensity with the spacing of isotherms at 10,000 ft and for the layer between 700 and 500 mb. That cyclones are observed in regions of strong temperature contrast was confirmed; otherwise no definite correlation was established. In this study, cyclones apparently were not classified according to the stage of development.

3. Determination of cyclogenesis, anticyclogenesis, frontogenesis, and frontolysis. Cyclogenesis, or the formation of wave cyclones on stationary or slow-moving cold fronts, is one of the more difficult problems facing the forecaster. There are certain preferred areas for wave development, such as the southern portion of mountain ranges where deformation of the cold front is induced. Under certain conditions, flat waves emerge from this region every 24 hr or so and move rapidly in a general easterly or northeasterly direction. These cannot be forecast in detail more than 12 hr in advance, and for longer periods precipitation should be forecast without attempting to define times of beginning and ending.

Waves form most frequently on stationary fronts or slowly moving cold fronts. The formation of a wave may be indicated by a new surge of pressure rises in the cold air, from the general pressure field, deformation of the front as it passes over mountain ranges, and cyclonic circulation such as may frequently be observed

on the Texas coast. Further deductions can be made from the distribution of clouds and precipitation.

The solenoidal field on the east coast of continents, in the Northern Hemisphere, is favorable for wave development in connection with stationary fronts. Techniques for handling cyclogenesis of this type have been described by Miller [35]. Byers postulates that no really primary cyclone forms entirely independently of other nearby disturbances and this appears to be correct for extratropical regions. The arrival of even a weak fall in pressure over a stationary front will almost always result in strong cyclogenesis.

Cyclogenesis tends to occur where there is a concentration of isotherms aloft (5000–15,000 ft) which provides considerable potential energy, and the wind field is favorable (*i.e.*, the wind blows across the isobars). Cyclogenesis (anticyclogenesis) tends to occur on the warm (cold) side of an area where isotherms are packed.

Baum [2] has described the Scherhag divergence theorem and states, on the basis of some informal experimentation with it, that it merits a trial by forecasters in the United States.

Anticyclogenesis can be detected by the pressure tendencies. For the forecasting of anticyclogenesis upstream, the transition from a cold to a warm high with resultant blocking should be carefully watched.

Frontogenesis and frontolysis are not, as a rule, particularly troublesome although there is, perhaps, a tendency to expect frontolysis too quickly. Over land areas, the forecaster must frequently deal with warm frontogenesis dynamically induced in the lee of mountain ranges and, less frequently, developing between two highs. The latter usually has some previous history. Important cold frontogenesis will occur in polar regions with the development and initial southward surge of arctic air masses.

The prognosticator finally completes the approximations of the future positions of fronts and the positions and intensities of the pressure centers. He then tests the consistency of these approximations with the upper-air synoptic and prognostic charts and the indicated extrapolation of mean isotherms between the 1000- and 700-mb surfaces. If inconsistencies appear, one or more of the factors used in the preparation of the surface charts has been incorrectly calculated or interpreted. Inconsistencies must be resolved by further checking or by according the greatest weight to the factors of greatest certainty.

In an effort to derive an objective method of forecasting the surface pressure pattern, Haurwitz and collaborators [28] investigated a number of procedures based on advection. Isopycnic lines were drawn for five levels between sea level and 14 km and forecasts of the density changes at these levels were prepared on the hypothesis that air moves with the geostrophic wind velocity, preserving its density. After allowance had been made for the thickness of each layer, the density changes were added together to obtain the sea-level pressure change. Vertical motions were not considered in the study. Correlation coefficients of significance were not obtained. This does not imply, however, that the subjective evaluation of advection in forecasting pressure change is not a helpful procedure. More recent research by Houghton and Austin [30], while furthering our knowledge of the mechanism of pressure changes, has failed, so far, to provide new tools for forecasting pressure changes.

No purely objective techniques are available for the preparation of prognostic charts and thus procedures remain generally subjective.

Vederman [55] has listed the techniques used in the preparation of prognostic charts and includes (1) vertical extent of highs and lows, and (2) temperature and height changes at various constant-pressure surfaces in addition to those discussed in this section. Some other procedures developed by American meteorologists have been described by Fulks and collaborators [23].

STEPS IN FORECASTING THE WEATHER

Preforecast Study. The organization of forecasting at a district forecast center should provide for at least two days' study of past maps each month. There are important month-to-month and seasonal variations in weather sequences, frequency of certain weather types, depth of surface heating, effect of nearby water areas, and many other similar influences which the forecaster must keep in mind. The forecaster can best integrate them in his mind by, for example, at the end of the month, running through several past years' charts for the following month and preparing practice forecasts for one or two areas or cities in his district.

Preliminary Steps in Forecasting. A certain amount of preparatory work is required for a forecaster when coming on duty. This may include:

1. Inspection of weather charts prepared since he was last on duty, or if he has been absent for an extensive period, for the last four to seven days, in order that he may bring himself up to date on the prevailing type of weather.

2. Briefing by the forecaster going off duty, who describes and explains the weather now in progress, the forecasts in effect and the basis for them.

3. Analysis of radiosonde observations from stations in the forecast district and in adjacent areas from which weather is approaching, for the purpose of ascertaining stability, cloud decks, freezing level, temperature and moisture changes, structure of moist layers, and indicated maximum temperature.

4. Inspection of the 850-mb chart for moist tongues, the general moisture pattern, advection of warm and cold air, and upslope areas.

5. Analysis of the 700-mb chart for depth of moisture, trends toward increased zonal or meridional flow, intensification of troughs and ridges, possible development of closed circulations, advection of colder and warmer air, pools of warm and cold air, and steering; checking the 700-mb prognostic chart from the weather central and modifying it as indicated.

6. Checking of the 500-mb and 200-mb charts, particularly the 500-mb, for the same features as given above for the 700-mb chart, except for moisture, which is not especially significant at the higher elevations. A

study of wave lengths with reference to the speed of movement of contour systems is especially important at the 500-mb level since this level is nearest the level of nondivergence. Also, when vorticity-trajectory computations are not made, some subjective study of the 500-mb chart with respect to constant vorticity trajectories is possible and helpful in forecasting movements and other developments.

7. Inspection of the latest pilot-balloon charts for trends and changes during the preceding six hours and for computation of the 6-hr trough and ridge movements. The forecaster should be familiar with the diurnal variation of winds under 8000 ft.

8. Checking the analysis of the last six- and three-hourly surface charts with the latest hourly sequences, noting recent frontal passages and movement of precipitation areas.

9. Reading the latest district forecasts from all available forecast centers, noting areas where precipitation and any unusual weather are forecast as well as any weather that does not seem warranted by the prognostic charts, and attempting to rationalize the basis for the differing forecasts.

10. Reading the synopses and forecasts from the various airway forecast centers, noting rate of movement given for fronts and pressure centers in the area where the forecast is made and also noting forecasts of unusual weather. Forecast centers in mountainous and distant areas are usually in a better position to determine these facts accurately than are remote centrals or forecast centers.

11. Inspecting the surface prognostic chart from the weather central and noting deviations from it by the preceding district forecaster and probable deviations from it by district forecasters in other forecast districts as indicated by the district forecasts. This check should be repeated when the new six-hourly synoptic chart is completed and analyzed.

Intermediate Steps in the Preparation of the Forecast. These may include:

1. Analysis of the new six-hourly synoptic chart, using weather-central analysis, modifying it in accordance with more detailed observational data which the forecast center may have in and near its own district.

2. Analysis of the pressure-change chart, forecasting future positions of each isallobaric center.

3. Using the 24-hr prognostic chart from the weather central as a basis, checking the positions and intensities of fronts and pressure centers, making such revisions as are indicated by the latest surface and upper-air charts, using techniques described earlier (pp. 750–758).

4. Final check based on subjective evaluation of broad-scale influences of centers of action, zonal indices and blocking highs.

5. Final reconciliation of all conflicting indications, objective, subjective, dynamical, and empirical. Weather processes are logical, and conflicting indications are evidence of an error in computation, in interpretation of the data and charts, or in reasoning. The original error may have been made by the observer, the computer, the analyst, or the forecaster but, in most calculations and chartwork, the error should be fairly obvious from the large amount of reasonably accurate information available. The forecaster should make the strongest possible effort to harmonize the prognostic information derived from a fairly large number of sources rather than to reject entirely a portion of it. Recomputation and re-evaluation may be necessary but, usually, the time required is well worth while. Most often a forecast error arises in the subjective weighting of the several factors from which the forecast is derived.

Steps Recommended by Petterssen. The following steps have been listed by Petterssen [42, Chap. XI]:

1. Displacement of pressure systems, fronts, etc.
2. Forecasting deepening and filling.
3. Determination of whether new systems will form.
4. Readjustment of displacements.
5. Determination of the position and properties of air masses—detailed analyses of clouds, hydrometeors, adiabatic charts, moisture patterns aloft. Constant-pressure charts will reveal physical and kinematic conditions of the air masses in question.
6. Determination of changes in air masses as they reach the forecast district. Possibilities of cooling, heating, depletion or addition of moisture, subsidence, etc.
7. Modification of local influences, mountain ranges, valleys, lakes, land and sea breezes, and other coastal effects.
8. Last re-examination: What can upset the forecast?
9. Wording of the forecast—clear and unambiguous —degree of certainty.

Influence of Main Centers of Action. By following the steps outlined above, the forecaster can, with the primitive tools at his disposal, complete the pressure prognosis for the first 24 hr of the forecast period. Because of acceleration and deceleration factors which are constantly operating, extrapolation becomes a progressively poorer tool with time. For periods in excess of 24 hr, the forecaster must expand his horizon in space and time. Data describing the instantaneous state of the atmosphere are inadequate to define, in any detail, the evolution of developments which may significantly affect the weather beyond 24 hr. Large-scale circulation changes cannot be ignored and important developments in one portion of the hemisphere may be compensated for in another far away. The forecaster must not become preoccupied with the latest hourly sequences or even the last six-hourly weather map.

The influence of centers of action was pointed out as early as 1881 by Teisserenc de Bort. The forecaster must keep close watch on deviations from the normal positions of the semipermanent areas of high and low pressure and the resultant position of the mean troughs and ridges and carefully consider their probable effects on migratory depressions and the preferred areas of cyclogenesis and anticyclogenesis under the prevailing type.

The centers of action are a conservative factor in the prevailing weather, that is, there is a tendency for a prevailing weather type to continue. An example of the

persistency of wet and dry weather has been given by Newnham [39] for Kew Observatory in England and Aberdeen in Scotland, as shown in Table V.

The forecaster should prepare a skeleton prognostic chart for each 24-hr period after the first 24-hr period,

Table V. Probability of a Rain Day

Station	Number of preceding successive fine days								
	1	2	3	4	5	6	7	8	9
Aberdeen	0.50	0.41	0.37	0.39	0.37	0.36	0.37	—	—
Kew	0.45	0.34	0.36	0.32	0.26	0.27	0.27	0.22	—
Station	**Number of preceding successive rain days**								
	1	2	3	4	5	6	7	8	9
Aberdeen	0.67	0.70	0.76	0.70	0.66	0.67	0.72	0.75	0.78
Kew	0.56	0.61	0.61	0.64	0.64	0.67	0.73	0.69	0.70

if one is not available from the weather central. The actual timing of frontal passages, wave development, precipitation surges and similar phenomena becomes increasingly difficult with time and, consequently, the forecast should be couched in more general terms.

Usefulness of Various Charts and Techniques. An analysis by Elliott, Olson, and Strauss [21] of questionnaires on the usefulness of certain charts and forecasting techniques returned by some 1460 forecasters in the military, civilian, and private weather services indicated the percentages of forecasters finding the several charts and techniques "often useful" or "occasionally useful" as given in Table VI. One may safely say that charts and techniques only "occasionally useful" are not regularly used in the preparation of forecasts. The question may be asked, Are forecasters using all the most valuable available techniques? If not, is it because of lack of clerical assistance, lack of time or lack of knowledge of how to use them, or do the few simple extrapolation and stability estimation techniques provide the same accuracy as additional checks using more complicated techniques?

Table VI. Usefulness of Charts and Techniques

	Often useful	Occasionally useful
Charts		
700-mb chart	95.3	3.0
500-mb chart	57.5	26.2
850-mb chart	37*	—
300-mb chart	22*	—
200-mb chart	16*	—
Cross sections	22.8	21.0
Thickness charts	10.4	9.6
Mean virtual temperature charts	5.0	4.2
Isentropic chart	2*	4.9
Adiabatic diagrams	92.0	4.0
T-ϕ diagrams	4.2	2*
Rossby diagrams	8.6	13.7
Other charts or diagrams	28.8	—
Techniques		
Petterssen's kinematic technique	10.3	26.4
Steering of surface system along upper-air flow	76.2	19.1
Extrapolation based upon history	81.0	15.1
Frontal movement from gradient winds	61.6	31.3
Upper-level isobar-isotherm relationship for movement of waves	37.6	36.1
Upper-level isobar-isotherm relationship for deepening or filling	53.5	29.4
Average cyclone and anticyclone tracks and movements	32.3	42.0
Analogues	4.7	17.5
Weather types	23.6	31.1
Rossby's constant-vorticity trajectory	2*	7.9
Other techniques	21.5	—

* Approximate per cent.

Prognosis of the Weather. The forecasting of the various weather elements is the last step in the preparation of the weather forecast but it is the most important and the most difficult. An understanding of the current weather, together with the surface and upper-air prognostic pressure patterns, forms in a general way the basis of weather forecasting. Tests have shown that forecasters demonstrate but little skill in forecasting the various weather elements when given a "perfect" prognostic pressure pattern alone.[3] One must again emphasize the *individuality* of each day's synoptic chart with the infinite variations of lapse rates, temperature and moisture, convergence, radiation, and all the other factors involved in future weather.

Cloudiness and Precipitation. Following the development of the concept of air-mass analysis by the Norwegian meteorologists, J. Bjerknes [4] in 1918 first described the arrangement of clouds and precipitation around a cyclone. The classical arrangement of hydrometeors around an idealized cyclone was described by Bjerknes and Solberg [5] in 1921 and was extended in 1922 [6] to include all stages of cyclone development. Since the ideal cyclone is rarely present on the daily weather chart, the marked variations in stability and moisture in the air masses normally found around the cyclone are of paramount interest to the forecaster. These elements frequently possess little relationship to the isobaric pattern. In many areas, the forecaster is faced with the complicated problem of determining where, as well as when, precipitation will develop in various sectors of a cyclone. No very satisfactory techniques are available for forecasting the transition of a dry to a wet trough other than by a rather subjective evaluation of moisture and stability factors.

In summer, particularly, the precipitation pattern is even more imperfectly associated with frontal and cyclonic systems. Pure air-mass thunderstorms may be forecast with fair success from the thermodynamic diagrams, but the general shower and thunderstorm activity, often nocturnal in nature, over large areas in middle latitudes, is very difficult to handle. Means [33] has suggested techniques for forecasting nocturnal thundershowers in the Midwest which show a slight increase in skill over more subjective methods, but the mechanics of the nocturnal thunderstorm in the Great Plains and upper Mississippi Valley are still imperfectly

3. It is not known whether, in this experiment, forecasters had access to the preceding weather charts. In any case it is not intended to imply that the prognostic pressure chart is not useful. Indeed, a good prognostic pressure chart is one of the most useful tools available to the forecaster.

understood and the forecasting of this phenomenon is still unsatisfactory. Verification of thunderstorms even in the relatively short-range airway forecasts (12 hr) is occasionally less than 20 per cent.

This summertime type of precipitation is of major economic consequence over many of the most important agricultural areas in the entire world and yet there is no comprehensive and satisfactory description of any reliable technique for forecasting it. Quantitative precipitation forecasts are needed by river engineers, hydroelectric power companies, and many other interests in commerce and agriculture. A rather primitive procedure for a forecast of this type has been devised by Showalter [50] based on the amount of moisture which will enter a given region and the unprecipitated amount likely to leave it. Verification is unsatisfactory since no precise methods exist for a rapid quantitative calculation of effective convergence over specific areas. The convergence of flow over mountain ranges is more susceptible to computation.

Klein [31] has correlated precipitation anomalies and the five-day mean 700-mb chart with considerable success but verification dropped from about 66 per cent on past weather charts to 25 per cent when his techniques were used on prognostic 700-mb charts. Although the technique was derived from five-day mean charts, it appears equally applicable to daily forecasts.

Precipitation results from the lifting of saturated air and is therefore associated with horizontal mass convergence in lower levels and horizontal mass divergence at higher levels, of which some of the associated factors are:

1. Frontal lifting, upslope or upglide.
2. Surface friction.
3. Topography.
4. Thermal instability.

V. J. and M. B. Oliver [40] have collected a number of principles of practical use in forecasting, which include:

1. Cloudiness and precipitation are prevalent under cyclonically curved isobars aloft (about 10,000 ft), regardless of the presence or absence of fronts on the surface chart.

2. In a cold air mass, the instability showers, cumulus, and stratocumulus clouds will be found only in that portion of the air moving in a cyclonically curved path.

3. In a warm air mass moving with a component from the south, cloudiness and precipitation will be very abundant under a current turning cyclonically or even moving in a straight line. (The importance of trajectories with increasing cyclonic curvature in producing convergence and heavy precipitation is stressed by all writers on this subject.)

4. Elongated V-shaped troughs will have cloudiness and precipitation in the southerly current in advance of the trough with clearing at the trough line and behind it.

5. Cold fronts will produce no weather when the component of the wind perpendicular to the front increases with height through the front.

6. If the 10,000-ft flow is parallel to the cold front, the front will be active and cloudiness and precipitation will extend as far behind the front as this condition obtains.

The applicability of all these rules is subject to the normal variations of moisture and stability.

Fog. The physics of fog formation have been discussed by Petterssen [43]. Fog is the subject of a separate article by J. J. George in this Compendium.[4] Techniques developed during the past two decades by George [25], based largely on analogues and statistics, provide the most practical methods of forecasting fog in middle latitudes.

Temperature. The main factors governing local temperature change in the free atmosphere are horizontal advection and vertical motion. The minimum temperature may be predicted by determining the location of the air expected over the station the following night and forecasting any indicated modification. The dew point of the expected air mass is a helpful indicator. Consideration must be given to modifications resulting from emergence of air from snow-covered to bare ground, trajectories over water, and such factors as wind and clouds. With clear skies and calms, the dew point of an air mass falls somewhat during the night as moisture is extracted from the air through the formation of dew.

The maximum temperature may be predicted from the sounding in the air mass expected to be over the station during the day. Depending upon the season, the lowest 800–1400 m will be warmed by heat transferred to the air from surface insolation. Although more refined methods are available, a satisfactory approximation of the maximum temperature may be obtained by following the dry adiabatic from a point on the sounding at the top of the area expected to be warmed down to the surface and picking off the temperature. Corrections must be applied for any indicated cloudiness and precipitation. In winter, diurnal insolation may have difficulty in overcoming even very shallow cold air inversions.

Wind. The gradient wind may be determined from the prognostic surface pressure pattern. The proportion of the gradient wind actually realized on the surface varies from 40 to 80 per cent or more over water with the variation dependent upon a number of factors, of which the stability of the air is most important. Over land areas the frictional effect is greater.

Synoptic situations, mostly in spring, favorable to tornado formation may be recognized by the forecaster, but any precise definition of the areas where tornadoes may occur has, so far, been impracticable. Recently E. J. Fawbush and R. C. Miller, in an unpublished paper,[5] have claimed some progress in defining the geographical areas in Oklahoma where tornadoes may occur by locating the intersection of areas of greatest instability, increasing dew point, narrow band of high

4. Consult "Fog" by J. J. George, pp. 1179-1189.

5. (*Added in press.*) These results have now been published in an article by Fawbush, E. J., Miller, R. C., and Starrett, L. G., "An Empirical Method of Forecasting Tornado Development." *Bull. Amer. meteor. Soc.*, 32: 1–9 (1951).

winds aloft (40 knots or more), and the trigger action associated with an approaching cold front. Tornadoes appear to move with the speed and in the same direction as the thunderstorms with which they are normally associated, that is, with the winds between 6000 and 12,000 ft. These winds will normally be from the southwest.

Verification. The primary purpose of any forecast verification system is to determine whether satisfactory standards of forecasting are being maintained, and, secondarily, to ascertain the relative skill of forecasters and forecasting candidates. The system should be rational, fair to the forecaster, and should not require more than a few minutes of his time in paper work. If too many elements are verified, too much smoothing will result and final values will have little significance. Except for very short-term forecasts, spot-check verification (*e.g.*, verification of specific elements at exactly 0200, 0400, 0600, etc.) should be avoided, since, after eight hours, values based on verification of this type will show no significant differences between good and mediocre forecasters, because of the variability of weather. The longer short-term forecasts (12–48 hr) should not be verified by methods involving a refinement of detail beyond the current ability of the forecasting profession. Verification of trivia should be avoided or at least handled with intelligence. One system of verification currently in use awards the forecaster 100 per cent for a measurable amount of precipitation (0.01 in.) and 0 per cent for a trace when precipitation has been forecast. In many countries a measurable amount is 0.1 mm, a unit only four-tenths of the one used in the United States. Over large areas of the United States, precipitation amounts of either a trace or 0.01 in. comprise 50 per cent or more of the total periods of precipitation, and, in a large proportion of these cases, it is fortuitous whether the observer records a trace or 0.01 in. Yet in this system of verification, the range in skill score is almost entirely dependent upon the forecaster's luck with these very trivial precipitation amounts.

Indeed, no substitute has been developed in the United States which is an improvement over the old U. S. Weather Bureau verification system. If a sufficient number of forecasts are included in the verification, chance coincidence and normal expectancies need not be considered since contributions from these sources will become equal for everyone and the differences in the final score will represent a satisfactory approximation of the differences in skill.

STATISTICAL AIDS IN FORECASTING

Objective Forecast Aids. Gringorten [27] defines an objective forecast as one that is made without recourse to the personal judgment of the forecaster. He further states meteorology is much too complex to allow one to believe that objectivity in forecasting will, eventually, completely replace subjectivity. To date, wholly objective forecast techniques are very rare and have shown improvement over accepted subjective methods only when the latter are exceptionally inadequate and verification is unusually poor, such as for certain types of precipitation forecasting, or when the forecaster is relatively inexperienced. Objective methods for forecasting minimum temperatures, in which subjective selection of certain parameters if often required, have at times equaled subjective methods but have rarely surpassed them. Vernon's study [56] of winter precipitation at San Francisco is an excellent example of a useful and reasonably objective forecasting technique. The study of wintertime precipitation for Washington, D. C., by Brier [9], while a commendable attempt to attack a difficult forecasting problem, cannot be regarded as particularly successful in providing the forecaster with an improved forecasting tool. Forecast methods of this type tend to meet with passive resistance from forecasters, who feel that the derived formulas and diagrams oversimplify the problem and that other contributory variables are frequently neglected.

Climatological-Statistical Forecast Guides. From time to time, and particularly in recent years, forecasters have been promised climatic guides which statistical climatologists claim will assist and improve forecasting. So far, few have been forthcoming, possibly because it is now realized that their value may have been overemphasized. Local forecasters soon know the earliest and latest dates of killing frost, after answering a dozen inquiries on that question, and district forecasters learn climatic normals and extremes of their forecast district as part of their training. It does not appear that any considerable amount of time and expense for this purpose is justified on the basis of results to date.

Climatological summaries were of considerable value during the recent war in forecasting for foreign areas where few or no weather observations were obtainable and, of course, are of great value for other purposes.

Tests of Empirical Forecast Rules. A large number of empirical rules have been developed and formulated since the beginning of forecasting. Some, such as those of Chromov [15], have been given wide distribution, but most forecast centers have their own collection in one form or another. Too many such rules are only confusing to the forecaster who, in the limited time available, may be unable to recall those applicable to the synoptic situation at hand. Rules may be classified according to seasonal, local, or broad-scale applications, types, etc. The value of many empirical rules might be enhanced by subjecting them to objective tests with further stratification.

ORGANIZATION OF FORECASTING

It seems doubtful that the accuracy of weather forecasting can be improved materially within the near future with present or anticipated new techniques. Some radically different approach to short-range forecasting will be required if any marked improvement is to be effected. Some significant, although slight, increase in forecast accuracy may be possible with feasible changes in forecast organization. Because of economic necessity, forecast organization in many weather services leaves much to be desired.

It is beyond the purview of this article to describe in any detail how forecasting should be organized, but all forecast organizations should provide for the following:

1. An adequate method of selecting forecasters. Demonstrated interest in forecasting, ability shown in a practice forecast program, temperament and general intelligence, among other factors, should form the basis for selection.

2. Verification of forecasts made by regular forecasters. The weather service should have some method of ascertaining whether proper standards are being maintained, and whether the accuracy curve of individual forecasters is rising or falling or has leveled off, as well as some method of determining the comparative skill of individual forecasters.

3. Satisfactory environment for forecasting. Adequate forecast performance requires a high level of concentration not possible in congested and noisy offices. Provision should be made for rapid preparation or receipt by facsimile of needed charts and maps, their efficient display, and the arrangement of other references and aids where they may be located and utilized quickly.

4. Sufficient time for the proper analysis of data, application of techniques, reconciliation of conflicting inferences from the several charts and diagrams, and for the preparation and wording of forecasts. An hour and a half is estimated as the irreducible minimum required on the average for the various forecast steps after completion of the six-hourly synoptic chart, and an additional sixty to ninety minutes is needed for the preparation of forecasts and warnings; communication schedules should be arranged accordingly.

5. A forecast staff large enough to permit a forecast schedule with time for research, study, and relief from forecasting. Forecasting requires energy, enthusiasm, and imagination and the work should never be permitted to become routine or the forecaster to become stale.

6. Freedom from interruption and from administrative and service work when on forecast duty. Although the forecaster is involved in problems of extreme complexity, he frequently must operate under conditions which permit constant interruption from phone calls and pilot briefing. Work performed under these conditions must inescapably be of inferior quality. Local forecasting can, of course, be combined with briefing and public service work.

7. Proper scheduling of forecasters. There are still differences of opinion between and within forecast services with respect to the type of schedule which will result in forecasts of the best quality and continuity consistent with proper health safeguards for the forecaster. Tests should be feasible.

8. Testing of suggestions, new ideas, and techniques for the improvement of forecasting. Each weather service should have one forecast center adequately equipped and staffed for testing suggested new techniques in actual forecasting.

NEEDED RESEARCH IN SHORT-RANGE FORECASTING

Some sixteen forecast units in the United States responded to an invitation from the author to indicate the areas of research of greatest urgency in short-range weather forecasting. The replies covered every phase of the whole problem of forecasting and reflected a general admission of the inadequacy of current techniques. There was general agreement that research should proceed along two main lines: (1) toward the discovery of fundamental truths about the atmosphere, with emphasis on the general circulation, and (2) toward the development of forecast techniques which would lead to improved forecasts in the immediate future.

There is little doubt that the first is the more important, and that the second should be given a secondary role. Certainly the yield from the large amount of research on practical forecasting during the past decade or two has been exceedingly meager.

Suggested topics for further research included:

1. The general circulation of the atmosphere. Although many forecasters feel that the large proportion of research at the meteorological departments of universities which is focused on the broad problems of the general circulation of the atmosphere has little immediate and practical application to the problems of forecasting, it seems likely that any marked improvement in forecasting must come from this quarter. One may hope that discoveries will eventually be made that will revolutionize current forecast methods, but, even earlier, it seems probable that new knowledge of the general circulation would permit the preparation of better prognostic upper-air charts and other forecast aids. Continued intense research on the general circulation should have the highest priority.

2. Aperiodic departures from the normal positions of centers of action and mean troughs and ridges. Of paramount importance to short-, medium-, and long-range forecasting are the causes of the departures of the centers of action from their normal location and intensity—anomalies which may persist for a matter of months. Of equal moment are the causes of the variations of the zonal indices, which in turn permit, at times, excessive zonal or meridional flow; it is likewise desirable to learn how these variations can be forecast.

3. The qualitative and quantitative forecasting of precipitation. All forecast units listed this forecast problem, and the majority described it either as the most difficult or as the most urgent. This problem includes:

a. Determining when and where precipitation will break out in a previously dry trough.

b. Accurately timing the beginning of precipitation 12–24 hr in advance.

c. Forecasting the nocturnal type of thunderstorm.

d. The degree of concentration of the convective afternoon thundershower.

e. The development of "bursts" of rainfall, instability lines (squall lines), and other zones of convergence.

f. Correlation of precipitation with upper-air patterns and the direction and movement of lows, and development of methods of forecasting the great variation in warm-sector precipitation.

g. Investigation of the whole problem of summertime precipitation, the causes of most of which are obscure. It is suggested that further relationships between showers and convective instability be developed to reduce the forecasting of showers to some practical numerical basis.

4. The causes of deepening and filling of pressure systems. This problem was second in order of priority on the forecasters' list. With it is the associated problem of wave development.

5. The speed and direction of movement of highs, lows, and fronts. Tolerable errors (15°) in forecasting the direction of movement will result in very serious errors in the weather forecast for 24 hr hence. Less complex methods of employing acceleration and deceleration factors are desired.

6. Treatment of cold upper-air lows; their formation, movement, and dissipation.

7. Bringing the historical weather map series up to date, including 700- and 500-mb charts.

8. Development of instruments which could measure vertical motions, and changes in moisture and stability aloft.

There are numerous other problems, local in nature, which are properly the subject of investigations at the district forecast center, and, of course, only a beginning has been made in the attack on forecasting problems in the tropics.

REFERENCES

1. Austin, J. M., "An Empirical Study of Certain Rules for Forecasting the Movement and Intensity of Cyclones." *J. Meteor.*, 4:16–20 (1947).
2. Baum, W. A., "Scherhag's Divergence Theorem." *Bull. Amer. meteor. Soc.*, 25:319–326 (1944).
3. Baur, F., "Die Bedeutung der Stratosphäre für die Grosswetterlage." *Meteor. Z.*, 53:237–247 (1936).
4. Bjerknes, J., "On the Structure of Moving Cyclones." *Geofys. Publ.*, Vol. 1, No. 2 (1918).
5. —— and Solberg, H., "Meteorological Conditions for the Formation of Rain." *Geofys. Publ.*, Vol. 2, No. 3 (1921).
6. —— "Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation." *Geofys. Publ.*, Vol. 3, No. 1 (1922).
7. Bowie, E. H., and Weightman, R. H., "Types of Storms of the U. S. and Their Average Movements." *Mon. Wea. Rev. Wash.*, Supp. No. 1 (1914).
8. Braak, C., "Het Klimaat van Nederland." *Meded. ned. meteor. Inst.*, (A) Nr. 32 (1929).
9. Brier, G. W., *Predicting the Occurrence of Winter Time Precipitation for Washington, D. C.* W. B. Publ., U. S. Weather Bureau, Washington, D. C., 1945.
10. Brooks, C. F., "Clouds in Aerology and Forecasting, I." *Bull. Amer. meteor. Soc.*, 22:335–345 (1941).
11. Byers, H. R., *General Meteorology*, 2nd ed. New York, McGraw, 1944. (See pp. 449–488)
12. California Institute of Technology, Dept. Meteor., *Synoptic Weather Types of North America*. Pasadena, Calif., 1943.
13. Charney, J. G., "On a Physical Basis for Numerical Prediction of Large-Scale Motions in the Atmosphere." *J. Meteor.*, 6:371–385 (1949).
14. —— and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
15. Chromov, S. P., "Rules for Forecasting Synoptic Situations." (Translated by I. I. Schell.) *Bull. Amer. meteor. Soc.*, 16:21–22, 71–73, 108–110 (1935).
16. Cook, A. W., "The 12-Hour Pressure Change Charts" (unpublished).
17. Cressman, G. P., "On the Forecasting of Long Waves in the Upper Westerlies." *J. Meteor.*, 5:44–57 (1948).
18. —— "Some Effects of Wave-Length Variations of the Long Waves in the Upper Westerlies." *J. Meteor.*, 6:56–60 (1949).
19. Douglas, C. K. M., "Some Problems of Modern Meteorology. No. 4—The Present Position of Weather Forecasting." *Quart. J. R. meteor. Soc.*, 57:245–253 (1931).
20. Elliott, R. D., *Extended Weather Forecasting by Weather Type Methods*, 55 pp. U. S. Navy Dept., Long Range Weather Forecasting Unit, Washington, D. C., 1944.
21. —— Olson, C. A., and Strauss, M. A., "Analysis of Replies to a Questionnaire on Forecast Aids." *Bull. Amer. meteor. Soc.*, 30:314–318 (1949).
22. Fulks, J. R., "Constant Pressure Maps—Methods of Preparation and Advantages in Their Use." *Bull. Amer. meteor. Soc.*, 26:133–146 (1945).
23. —— Wobus, H. B., and Teweles, S., "A Collection of Reports on the Preparation of Prognostic Weather Charts." *U. S. Wea. Bur. Res. Pap.* No. 21. Washington, D. C. (1946).
24. Fultz, D., "Upper-Air Trajectories and Weather Forecasting." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 19 (1945).
25. George, J. J., "Fog, Its Causes and Forecasting with Special Reference to Eastern and Southern United States." *Bull. Amer. meteor. Soc.*, 21:135–148, 261–269, 285–291 (1940).
26. —— *On the Relationship between the 700-mb Surface and the Behavior of Pressure Patterns at the Ground.* Dept. Meteor., Eastern Air Lines, Atlanta, 1949.
27. Gringorten, I. I., "A Study in Objective Forecasting." *Bull. Amer. meteor. Soc.*, 30:10–15 (1949).
28. Haurwitz, B., and Collaborators, "Advection of Air and the Forecasting of Pressure Changes." *J. Meteor.*, 2:83–93 (1945).
29. Houghton, H. G., "Some Personal Opinions of the Present Status of the Science and Practice of Meteorology." *Bull. Amer. meteor. Soc.*, 29:101–105 (1948).
30. —— and Austin, J. M., "A Study of Non-geostrophic Flow with Applications to the Mechanism of Pressure Changes." *J. Meteor.*, 3:57–77 (1946).
31. Klein, W. H., "Winter Precipitation as Related to the 700-mb Circulation." *Bull. Amer. meteor. Soc.*, 29:439–453 (1948).
32. Longley, R. W., "A Study of the Relationship between the 700-mb Flow and the Movement of Surface Pressure Centers." *J. Meteor.*, 4:202–204 (1947).
33. Means, L. L., "Thunderstorms in the Central United States" (1946) (unpublished).
34. Miller, E. H., and Thompson, W. L., "A Proposed Method for the Computation of the 10,000-foot Tendency Field." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 1 (1942).
35. Miller, J. E., "Cyclogenesis in the Atlantic Coastal Region of the United States." *J. Meteor.*, 3:31–44 (1946).
36. Namias, J., "Investigation of Polar Anticyclogenesis and

Associated Variations of the Zonal Index." *U. S. Wea. Bur. Res. Pap.* No. 24, Washington, D. C. (1945).

37. —— *Extended Forecasting by Mean Circulation Methods.* U. S. Weather Bureau, Washington, D. C., 1947.
38. —— and Clapp, P. F., "Confluence Theory of the High Tropospheric Jet Stream." *J. Meteor.*, 6:330–336 (1949).
39. Newnham, E. V., "The Persistence of Wet and Dry Weather." *Quart. J. R. meteor. Soc.*, 42:153–162 (1916).
40. Oliver, V. J., and Oliver, M. B., "Forecasting the Weather with the Aid of Upper-Air Data" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, N. R. Beers, ed., pp. 813–857. New York, McGraw, 1945.
41. Palmer, W. C., "On Forecasting the Direction of Movement of Winter Cyclones." *Mon. Wea. Rev. Wash.*, 76:181–201 (1948).
42. Petterssen, S., *Weather Analysis and Forecasting.* New York, McGraw, 1940.
43. —— "Recent Fog Investigations." *J. aero. Sci.*, 8:91–104 (1941).
44. Richardson, L. F., *Weather Prediction by Numerical Process.* Cambridge, University Press, 1922.
45. Riehl, H., "Jet Streams in Upper Troposphere and Cyclone Formation." *Trans. Amer. geophys. Un.*, 29:175–186 (1948).
46. Rossby, C.-G., "Kinematic and Hydrostatic Properties of Certain Long Waves in the Westerlies." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 5 (1942).
47. —— and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
48. —— "Isentropic Analysis." *Bull. Amer. meteor. Soc.*, 18: 201–209 (1937).
49. Saucier, W. J., "Texas-West Gulf Cyclones." *Mon. Wea. Rev. Wash.*, 77:219–231 (1949).
50. Showalter, A. K., "An Approach to Quantitative Forecasting of Precipitation." *Bull. Amer. meteor. Soc.*, 25: 137–142 (1944).
51. Starr, V. P., *Basic Principles of Weather Forecasting.* New York, Harper, 1942.
52. Starrett, L. G., "The Relation of Precipitation Patterns in North America to Certain Types of Jet Streams at the 300-Millibar Level." *J. Meteor.*, 6:347–352 (1949).
53. University of Chicago, Dept. Meteor., "On the General Circulation of the Atmosphere in Middle Latitudes." *Bull. Amer. meteor. Soc.*, 28:255–280 (1947).
54. Van Bebber, W. J., "Die Zugsstrassen der barometrischen Minima." *Meteor. Z.*, 8:361–366 (1891).
55. Vederman, J., "The Weather Bureau–Air Force–Navy Analysis Center." *Bull. Amer. meteor. Soc.*, 30:335–341 (1949).
56. Vernon, E. M., "An Objective Method of Forecasting Precipitation 24–48 Hours in Advance at San Francisco, California." *Mon. Wea. Rev. Wash.*, 75:211–219 (1947).
57. Weiss, L. L., "Relation of Polar Continental Outbreaks to Meridional Indices" (1945) (unpublished).
58. Willett, H. C., *Descriptive Meteorology.* New York, Academic Press, 1944. (See pp. 282–305)
59. Wulf, O. R., and Obloy, S. J., "The Utilization of the Entire Course of Radiosonde Flights in Weather Diagnosis." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 10 (1944).

A PROCEDURE OF SHORT-RANGE WEATHER FORECASTING

By ROBERT C. BUNDGAARD

Air Weather Service, Washington, D. C.

INTRODUCTION

The practical procedure of short-range weather forecasting is simply the extrapolation, in a partly geometrical way, of sequential weather situations. On the other hand, the ideal mathematical prognosis is based completely upon physical considerations and uses only a single set of initial data. Since practical weather forecasting nevertheless aims at accomplishing the same effect as integrating the equations of atmospheric dynamics and thermodynamics, it too must take into account the physical relationships existing among the various meteorological variables, such as wind, pressure, and temperature. The extrapolation process therefore must also take physical considerations into account.

In this article we shall first consider certain general principles of practical weather forecasting. In order to make 24-hr to 36-hr forecasts, the complex analytic-prognostic process (leading from the observational data to the prognostic maps) must be centralized so as to permit the beneficial use of an operational procedure known as the synoptic cycle. By systematic use of the synoptic cycle the analysis as well as the construction of the prognostic maps is facilitated, since the prognostic maps form the first approximation to the analysis of the next set of weather maps. On the other hand, the prognostic-forecasting process (leading from the prognostic maps to the explicit forecasts) is decentralized so that the predictions of the different weather elements can be formulated locally. Such predictions can then be synthesized into explicit weather forecasts, such as general forecasts giving summary predictions for a large region, area forecasts giving more detailed local information, and special forecasts for various fields of human endeavor.

Although the involved procedure of weather forecasting should be undertaken systematically, it must necessarily vary with the prognostic region and weather situation under consideration. Still we have, partly for expository reasons, introduced a certain order of seven prognostic operations which can be recommended, in many cases at least, for use in temperate latitudes. This order serves to remind the forecaster of the many different rules, of a theoretical or practical character, which he should use. We shall see how the prognosis of the surface map can be based on geometrical extrapolation together with certain physical considerations. The prognosis of the upper-air maps will then be presented (geometrical extrapolation and physical considerations). Finally, we shall recapitulate the more important considerations to be borne in mind by the forecaster in predicting the various weather constituents: temperature; general hydrometeors; fog, stratiform clouds, and drizzle; cumuliform clouds, showers, thunderstorms, and hail; wind, turbulence, and tornadoes; and aircraft icing.

This article is essentially an epitome based primarily upon certain chapters of *Dynamic Meteorology and Weather Forecasting*, soon to appear as a publication of the Carnegie Institution. Most sincere thanks go to Professor C. L. Godske, who, as the principal author of the book, has so very generously and warmly approved the use of the book in this article. Some of the subject matter presented in this article has arisen through an intimate collaboration among C. L. Godske, T. Bergeron, and R. C. Bundgaard. Some new material has been introduced by the author, partly for the purpose of taking into account the scope and objectives of this Compendium. The author also wishes to acknowledge his deep indebtedness to Professors T. Bergeron, J. Bjerknes, and C. L. Godske for the great inspiration he has received through being associated with them.

THE GENERAL PRINCIPLES OF PRACTICAL WEATHER FORECASTING

The Synoptic Cycle. From just *a single initial state* of the atmosphere, its subsequent development can be mathematico-physically predicted by means of a certain system of differential equations and boundary conditions. Practical weather forecasting, on the other hand, is based upon prognostic maps, first sketched by geometrically extrapolating (in a purely formal way) a *sequence* of completely analyzed weather maps, and then improved and amplified by considerations based on experience and/or general physical principles. The forecaster should, therefore, first survey carefully the analyzed map sequences, to amend all recognized errors and to add the finishing touch to the maps if time limitations had forced him to leave them in an unsatisfactory condition. Through the complete, detailed, and correct analysis of the maps, dynamical and thermodynamical considerations can *indirectly* be taken into account in the preparation of the prognostic maps preliminarily based only upon pure extrapolation. Owing to the importance of map analysis as the foundation for the prognosis, we have introduced *a synoptic cycle involving the three basic operations of "epignosis," diagnosis, and prognosis.*

Let us first consider the synoptic cycle as it presents itself in connection with the upper-air maps. The observations at time $t = t_0$ are plotted on the prognostic upper-air maps referring to the time t_0 and prepared at $t_0 - 24^h$; we shall refer to these as maps I. These maps represent a first approximation to the analysis, which, like that of the surface map, has the form of

an "epignosis" of prognostic maps. When the diagnostic-analytic task is finished, the analyst centers his attention upon the prognostic upper-air maps, maps II, prepared at the time $t_0 - 12^h$ and referring to $t_0 + 12^h$; these maps are readjusted, both by performing the first stage of prognosis upon maps I and by using the curves already drawn and only checking and correcting them in the light of the new observations now available. Finally, using maps I and II, the prognostic maps, maps III, for the time $t_0 + 24^h$ are prepared. They form the basis for the forecast. At the next synoptic hour the analyst then starts with the maps II (prepared at time $t_0 - 12^h$ but readjusted at time t_0), on which the $t_0 + 12^h$ observations have been plotted.

The synoptic cycle described above not only improves the analysis and prognosis of the weather maps, but also shows the analyst to what extent the prognostic extrapolation has been successful in any particular weather situation. Thus, it reveals the influences that have been totally or partly neglected during the prognosis. In particular, we may hope that the application of the synoptic cycle may yield valuable information as to the nonadvective effects in the upper air.

For practical purposes it is necessary to employ several meteorologists for the simultaneous prognosis of the system of surface and upper-air maps; a certain subdivision of work thus becomes necessary. However, a strict partitioning should not be maintained in practice between the different parts of weather prognosis, for instance by dividing (as is often done) a weather analysis center into a "surface" section and an "upper-air" section. Instead, the upper-air prognosis should be regarded as part of a single procedure involving the surface-map prognosis. The forecaster continually improves the surface-map prognosis from the results of the upper-air analysis and prognosis. On the other hand, he builds his system of prognostic upper-air maps on the prognostic surface map and must continually adjust his prognostic upper-air maps so as to profit from the recent improvements in the surface-map prognosis.

The General Forecast. Over a very large *prognostic region*, the extent of which varies from one weather situation to another, a smoothed prognostic map is derived from the analyzed weather maps based on representative observations of a more inclusive *analytic region* (see Fig. 1). From this map system is issued the *general forecast*, which—for a *forecast period* more inclusive than the *prognostic period* of the maps—states in a semi-Lagrangian way the probable gross weather development, in the absence of marked local influences, for an entire administratively fixed *forecast region* situated within the central part of the prognostic region. In particular, the general forecast must tell specifically the *what* and *where* of the following weather phenomena, including their present state and position as well as their motion and development during the forecast period: (1) the main air currents, (2) the main cyclones and anticyclones, (3) the main air masses and fronts, (4) the main areas of cloud and precipitation that may occur within an air mass and at a front, (5) the regions of especially strong wind, and (6) the general temperature distribution. Valid mainly for the synoptically representative stations, the general forecast is primarily of interest only to the weather service and for the strategic planning of region-wide activities such as communication, transportation, and aviation.

The Area Forecast. The general forecast, which is quite insufficient for a specified smaller *forecast area* within the forecast region, must be amplified, detailed, and even supplemented by an *area forecast* which—although based on the maps and the general forecast for the region—takes into consideration important local effects that were purposefully neglected during the regional analysis and prognosis. At the area forecasting centers, the general analysis, prognosis, and forecast are received at 12-hr intervals from the regional analysis center by teletype and/or facsimile machines. In order to feel the pulse of the locally influenced present and

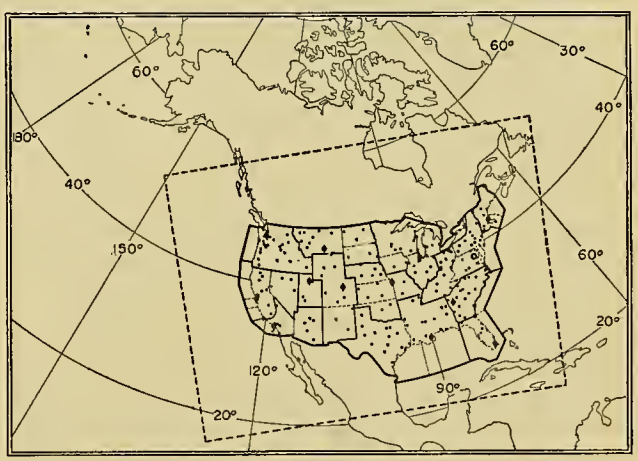


FIG. 1.—Analytic, prognostic, and forecast regions for the United States with its areas, districts, and localities.

past weather of his area, the area forecaster may often be obliged to prepare, mostly on a scale greater than that of the regional map system, his own sequences of more frequent and detailed maps (special maps, or even surface maps limited to the extent of his analytic and prognostic area). Being based upon the synoptically *unrepresentative* observations of a dense area network of primitive auxiliary stations which do not ordinarily belong to the international network, the area analysis forms a supplement to, not a duplication of, the regional map analysis. Moreover, the area forecaster is able, by the use of a network denser in time as well as in space and from more recent observational data, to fix the position of fronts, isobars, etc., in his area with greater accuracy than is possible at the regional analysis center. To further both these purposes he also, if residing within his area, makes use of such supplementary detailed "unwritten" observations as those he collects verbally from pilots and other contacts using the so-called local

method (see [1]). Fixing his attention upon the atmospheric fields within his area, the area forecaster predicts their form at any instant during his forecast period and thereby issues a more detailed and accurate area forecast, which describes in a semi-Eulerian way the local peculiarities of the weather and the time of arrival of the weather changes—the *what* and *when* of the weather.

The preparation of the general forecast is primarily the employment of the more theoretical aspects of meteorology for sketching in broad terms the probable future meteorological development and a rather idealized weather state associated with the development. The area forecast, on the other hand, applies the general forecast together with supplementary information so as to satisfy the requirements of its areas as to detailed predictions. The general forecast is a synthesis and a generalization, whereas the area forecast is a division and a specialization; they complement each other.

The area forecast bulletin should be perspicuously expressed, especially so as take into account the extreme local variability in time and space of the weather phenomena. Although brevity is essential, the forecast should not be resolved into such simple but ambiguous terms as "clear," "rain," or "snow." For example, a light cirrus overcast might be accepted as "clear" by the average person but would likely be regarded as "cloudy" by, say, a photographer. The ordinary area forecasts are regularly issued several times each day. Transmitted by radio, they are announced preferably by the forecaster himself (otherwise read *verbatim* by the radio announcer). Under the personal direction of the area forecaster, televised discussions of the map and other visual weather aids make possible a clearer presentation of the forecast than the verbal radio emission, and a more rapid dissemination than through the newspapers.

Further, the area forecast is particularized for each of several *forecast districts* into which the forecast area is subdivided. For delimiting the forecast districts on the basis only of weather similarity, it would be tempting to apply a rational method such as Schaffer's criterion [70], namely, the mean of the probable number of coincidences of rain days and of dry days for any two district stations. But at the present time the subdivision into districts is determined mostly by political reasons, so that the area forecaster must be at liberty to vary their extent whenever necessary. Although the district forecast bulletins are generally issued together in a prescribed order, it is sometimes advantageous to collect the districts into moving "zones" having practically the same weather (*e.g.*, a zone parallel to the general direction of an approaching front) and to give the same forecast for the whole zone.

The optimum extent of the forecast area and districts depends on its latitude and topography as well as the needs and activities of its inhabitants. To apply successfully the supplementary data furnished by the special area network of unrepresentative observations and by the local method, the area forecaster must be familiar with the degree of unrepresentativeness of the different meteorological stations within his districts and must possess an intimate knowledge of the local topographic factors affecting the weather in the vicinity of these stations. Then too, the forecasting work should not be unnecessarily complicated by having a forecast area so large as to be simultaneously influenced by different weather systems.

The Local Forecast. Although the area forecast is still partly a general forecast, it tends considerably toward specialization in the operational interests of such area-wide activities as communication, transportation, aviation, hydrology, and fire control, and especially of such district-wide activities as, say, fishing and planting. But if the area forecast were to be sufficiently detailed to meet the manifold needs of the populace of the area, it would be, especially in the United States, so voluminous and involved that the individual layman (*e.g.*, editor, shipper, farmer) could not easily extract from it for his particular needs. To localize, detail, and amplify further the area forecast for a particular *forecast locality*, and especially to specialize it for the benefit of various local endeavors, is instead the task of a locally residing forecaster, who, so to speak, acts as the local agent or representative of the weather service, supplying the local weather consumers with what they really need.

Every six hours the local forecast office receives from its forecast center the area forecast—but not the area map analysis or prognosis—which is then further detailed by the local forecaster as to arrival time of local weather phenomena (*e.g.*, showers, etc.) and is also amplified and specialized for the needs of the various and particular endeavors of his locality. The Eulerian considerations are thus, to a lesser extent, applied also by the local forecaster. Using mainly the district forecast of the area forecast center as guidance material, the local forecaster formulates his many local and special forecasts through an intimate knowledge of the weather peculiarities and activities of his locality. To some extent he also utilizes various supplementary, detailed, and reliable local observations that he himself can make up to the very moment of issuing the local forecasts when the area forecast may be up to three hours old. As a matter of fact, the local forecast office generally takes part in an observing program contributing to the area and/or international network, so that at least a part of these observations can be applied toward his local forecast problems. Although he does not prepare maps, he may often apply specialized techniques of analyzing and prognosticating locally the various weather elements, such as his locally and empirically developed diagrams for predicting minimum and maximum temperatures. Finally, the local forecaster must have, in addition to his intimate knowledge of the local topography as well as his in-station study of local forecasting problems [1], a thorough acquaintance with the various weather needs of his community and must be ready to supply special forecasts regularly and whenever called upon.

But the local forecaster himself cannot provide all the information wanted for special purposes. Therefore, in addition to the area forecast centers, *special fore-*

casting centers are needed to assist the local forecaster in issuing certain special forecasts. For example, we refer to such formal U. S. Weather Bureau organizations of the local forecast offices under special forecasting centers as the Corn, Wheat, Cotton, and Livestock Weather Service, the Hurricane-Warning Service, the Airways Forecasting Service, the Flood-Warning Service, the Fruit-Frost Service, and the Fire Weather Service. In addition we may mention the application of special forecasts to such enterprises as the shipping of perishable goods, the distribution of seasonal merchandise (winter clothing, ice cream, power, light, and fuel), the cinematic industry, the conservation of water, and irrigation.

The geographical extent of the forecast localities, of roughly a 40-km radius, generally should be somewhat irregular and should be subject to occasional changes because of the fluctuating administrative and professional functions of the forecast localities. Wherever there is no local forecast office to service new requirements, a nearby local forecast office is appointed to fulfill the service. Such an appointment is based on several factors in addition to the geographic one, for instance on the size and ability of the administrative and professional staff to handle the new task.

Summing up, we may state that map plotting, analysis, and prognosis—for reasons of economy—call for a certain *centralization of analytic and prognostic work*, especially of upper-air and routine work. On the other hand, the necessity of detailed forecasts (of different types) valid for restricted areas and localities leads to a demand for a *decentralization of prognostic and forecasting work*. The most practical compromise between these two tendencies will depend on the topography of the region, on the geographical latitude, and on the types of forecasts desired [41].

The Accuracy of Forecasts. Fronts, precipitation areas, etc., linearly extrapolated 24 hr to 36 hr ahead, may well have space errors from two to four times as large as do their analyzed positions, which can be fixed with an accuracy of from about one-half to one-fourth of the distance between adjacent synoptic stations, that is, with an analyzed accuracy of about 50 km. Depending on their velocity of propagation as well as on the errors in their analyzed positions, the inaccuracies of the arrival time of a fast-running disturbance moving, say, with a constant speed of 100 km hr^{-1} may therefore be of the order of magnitude of 2 hr. Further inaccuracies—the so-called "weather surprises"—are involved in the extrapolation of fast-running disturbances coming from a region where, owing to a sparse network, their positions have been analyzed inaccurately.

The value of a forecast depends upon its minutiae as well as upon its accuracy; a forecast proved correct according to purely formal methods is not necessarily the most useful forecast (for more details see [10, 18, 38, 52, 78]). Although it is not easy to verify a short-range weather forecast and to judge fairly the efficiency of a forecaster, statistics of success may be of some interest in weather services where the methods of analysis and prognosis are to a certain extent standardized. (For a survey of literature on the verification of short-range weather forecasts, see [54] and the discussion elsewhere in this Compendium.[1])

THE PROGNOSIS OF THE SURFACE MAP

The Geometrical Extrapolation of the Surface Map. *The first stage* in the order of operations for obtaining the $t_0 + 24^h$ prognostic surface map *is the geometrical extrapolation of the analyzed surface maps for finding the $t_0 + 3^h$ prognostic surface map.* In a sequence of surface maps carefully analyzed 3 hr apart,[2] the forecaster can determine directly the motion and development of their generally nonphysical features, such as isobars, pressure centers, ridges, troughs, and fronts with their cloud and precipitation areas: By superimposing on the tracing table the maps for $t_0 - 3^h$ and t_0, average values are found for the 3-hr period ending at t_0. Though less reliable than such direct determination of velocity, the kinematical formulas independently introduced by Dedebant during 1927–32 [19, 20], Angervo during 1928–30 [2, 3, 4], Wagemann in 1932 [81], and Petterssen in 1933 [58] are nevertheless of some use for determining, from the reported 3-hr pressure tendencies, the motion and development of features and systems which have just entered the analytic region. Valid only wherever there are no discontinuities in the space and time derivatives of pressure, these speed formulas should preferably be applied to sea-level systems without marked frontal structure. Thus, for determining the tangential speed of a frontal cyclone, Petterssen [60, pp. 398, 411] has recommended that the nonfrontal trough formula be based upon the pressure and tendency profiles poleward of its center where there are no fronts. The deepening and filling of sea-level pressure centers and cols, as determined either from the application of numerical formulas to the last surface map or from the last two available surface maps, must be corrected for an appreciable diurnal variation before being used for extrapolation.

By means of the average values thus found for the 3-hr period ending at the time t_0 of the last map, the forecaster extrapolates linearly and in a merely formal way the position, 3 hr ahead, of the map lineaments and the deepening and filling of sea-level pressure systems. A correction, which is next performed as part of stage one, consists in using the sequence of 3-hr maps from $t_0 - 24^h$ to t_0 in order to take into consideration higher-order terms in the extrapolation. Moreover, if the sea-level allobaric centers (preferably for 12 hr) move with approximately the same velocity as the pressure center, their past movement may be used to determine more surely the past movement of the pressure system. If this is not the case, the future velocity changes of the sea-level pressure systems are sometimes indicated by the recent movement of the corresponding sea-level allobaric centers.

As stage two, a rough sketch of the sea-level prognostic surface for $t_0 + 24^h$ is made by applying these geometrical

1. Consult "Verification of Weather Forecasts" by R. A. Allen and G. W. Brier, pp. 841–848.

2. Prior to 1930–35 the maps were prepared at 6-hr intervals.

extrapolation methods to the sea-level pressure maps for t_0 *and* $t_0 + 3^h$. After the completed prognostic surface map for $t_0 + 3^h$ has been used as an intermediate step for finding the sea-level prognostic surface map for $t_0 + 24^h$ (by reapplying the foregoing methods), it is laid aside to be used later as the next surface-map analysis.

The Physical Extrapolation of the Surface Map. As an analyst, the meteorologist is concerned with selecting models which can be fitted to the observations. As a forecaster, he considers how well the development of the chosen model fits the geometrical extrapolation described previously, and to what extent adjustments are necessary, either to this extrapolation or to the idealized development of the models during the forecast period. In both their ideal and deformed states, these tropospheric models are formulated from experience (in daily weather analysis) combined with special scientific investigations. Synthesizing, roughly, our increasing knowledge about atmospheric dynamics and thermodynamics, these models should not only be continually readjusted and improved but should also be supplemented by new models. Since no rigid set of tropospheric models exists, weather-map prognosis should not stiffen into a series of routine mechanical procedures.

The prognosis of the surface map can be facilitated by the following general considerations concerning the physical consistency of the tropospheric models. Air masses, fronts, cyclones, and anticyclones extend over large areas. In other words, the models must be *geometrically consistent*, a condition which leads to simplifications for the prognosis. Further, the displacement and deformation of the models during the prognostic period must be kinematically probable. Equipollently, the extrapolated position of a model must be *kinematically consistent* with the model at the present time and with the average field of motion during the prognostic period, both in the region covered or passed by the model and within the model itself. For example, a shallow, lenticular-shaped depression must be connected with an initial wave at a frontal zone, a deeper and rather circular one with an occlusion. Moreover, the structure and the stage of development of the chosen model must also be in agreement with the field of forces acting upon it during the prognostic period; that is, the model must be *dynamically consistent*. For example, a shallow and lenticular depression, implying small vorticity and a comparatively small supply of kinetic energy, will correspond to the initial stage of a frontal disturbance with a maximum of potential energy. On the other hand, the deep and circular depression, implying maximum vorticity and kinetic energy, could as a rule result only from the occluding process of a frontal disturbance. Finally, the chosen model must harmonize with the observed processes of radiation, heat conduction, condensation, and precipitation, that is, the model must be *thermodynamically consistent*.

The tropospheric models possess certain physical characteristics, found by theory and synoptic experience, which must be borne in mind by the meteorologist in making the map prognosis and the weather forecast. We shall now recapitulate these characteristics in the same order as the irreversible, cyclic development of the models, namely, (1) frontogenesis, frontal cyclogenesis, and anticyclogenesis, (2) the movement of sea-level pressure systems, (3) frontal occlusion, (4) degeneration and regeneration of cyclones and anticyclones, (5) local modifications of air currents, air masses, and fronts, and (6) frontolysis, cyclolysis, and anticyclolysis.

Frontogenesis, Frontal Cyclogenesis, and Anticyclogenesis. In an extensive quasi-stationary pressure col on the surface map, the forecaster should probably expect kinematic frontogenesis if the isotherms form a smaller angle with the axis of stretching than with the axis of shrinking (the part of the front marked "active" in Fig. 2). At least in the lowest layer, the general poleward decrease of temperature may sometimes be exceeded, especially along coastal regions, by the longitudinal temperature differences connected with the distribution of land and sea. Then, in these coastal regions, the cols with a north-south (west-east) axis of stretching give rise to frontogenesis (frontolysis).

Aloft, an accelerating jet stream in the direction of the frontogenetically effective axis of stretching aids in the upward extension of the frontogenetic field of deformation. Experience has indicated that a weak frontal surface will change in intensity (either by frontolyzing or by frontogenizing) wherever on a constant-pressure map one or more of the following conditions are observed: The *relative isohypses*[3] within the *frontal strip*[4] (1) have essentially the same spacing as they have outside the frontal strip, (2) are not parallel to the frontal strip, or (3) are not strongly curved in crossing the fronts.

For the quasi-stationary front thus formed, friction tends to increase the slope up to a certain height if the colder air is situated on the low-pressure side of the current—as is mostly the case with fronts in the westerlies and within the outskirts of an old, cold cyclone. A corresponding increase in sharpness and shear must be expected. This is one probable reason why such quasi-stationary fronts show a maximum (kinematic) frontogenetic tendency.

A front kinematically created along some thermal discontinuity at the surface of the earth may be thermodynamically conserved or intensified by prolonged temperature differences between, say, land and sea, snow-free and snow-covered land, or adjacent sea currents. In fact, it is quite probable that the atmospheric thermal discontinuity at the ice limit may in itself create a hyperbolic flow with a slanting surface of stretching. A cold front of limited extent may form within the lower layer of an air mass crossing a mountain range, and be intensified through cooling by its forced ascent and by the evaporation of the falling, orographically released precipitation. The forecaster should also expect the intensification of a quasi-stationary front when the isallobaric gradient is directed

3. For a definition of the aerological terms used in this text, see Table I.

4. The frontal strip is defined as the area of the map bounded by the intersections of a frontal surface with the boundary surfaces of a mandatory layer, as shown for instance in Fig. 3.

against it, and of a moving front as it approaches a deep pressure trough.

It has been known for some time that the formation of frontal waves is somehow partly initiated by a lessening of *dynamic* stability in the adjacent air masses, a decrease that has taken place through the preceding frontogenesis. For example, as observed already in 1909 by Guilbert [36] and subsequently affirmed by Hesselberg [40] in 1915, a cyclonic, subgeostrophic, and divergent flow with indraft is followed locally by a stronger pressure gradient and so by cyclonic deepening. A more complete and up-to-date theory of frontal wave formation has been presented in "Extratropical Cyclones" by J. Bjerknes, pp. 577–598 in this Compendium. The main points of his theory are as follows: The first formation of the frontal wave depends on the intensity that can be reached by organized upgliding and downgliding on the quasi-stationary frontal slope which, as a result of the preceding frontogenesis, has just acquired a tilt almost the same as the saturation isentropes. Organized upgliding of saturated air is dynamically restricted to those parts of the frontal slope where the geostrophic wind increases along the warm air streamline. It will be stronger the more nearly the isentropic anticyclonic shear of the geostrophic wind approaches $2\Omega_z$, where Ω_z is the vertical component of the earth's rotation. The latter condition is identical with the *lessening of dynamic stability*. That condition is satisfied along the eastward half of the axis of stretching of a col (the front portion marked "active" in Fig. 2), where the first active upgliding of the beginning cyclogenesis is most frequently found. The first organized downgliding will occur where the progressive frontogenesis has produced a large local decrease in the geostrophic wind on the cold side of the front. Such downgliding, following a dry isentrope of much flatter slope than that of the frontal zone, will form the first cold bulge on the quasi-stationary front. The warm sector of the nascent cyclone forms immediately east of that cold bulge. Wave formation thus is the combined product of warm-air upgliding and cold-air downgliding along neighboring front sectors, which, during the process, become the warm front and the cold front of the new wave.

The first cold push must be a shallow one. The first formation of the frontal wave therefore depends on trigger action from low levels and should be studied carefully on the surface map. With the colder air situated on the low-pressure side of the current, the frictionally steepened and sharpened quasi-stationary fronts in the westerlies and in the rear of large polar-front lows have attained the slope of a saturation isentropic surface and therefore present initial conditions favorable for wave formation (see Fig. 2). With the opposite temperature distribution (such as in high-latitude easterlies, in westerlies with poleward increase of temperature, and within the outskirts of anticyclones) the quasi-stationary fronts have a diminished slope and shear and therefore no cyclogenetic tendency. Waves are therefore to be expected along a quasi-stationary *polar* front between *principal* air masses, especially if the front has a west-east orientation with cold air poleward. Although a medium- or high-speed warm front is an *anafront* (for definition, see Fig. 2), the friction will tend markedly to diminish its slope, at least in the lowest kilometer of the atmosphere. Within strongly cyclonically curving easterly flow near a cold low, the low-layer convergence and lifting of polar air contributes to a steepening of the cold-front surface; here, however, cyclogenesis is generally prevented because near the center of the low the cold front becomes a *katafront* (defined in Fig. 2).

Since wave formation depends on the degree of lessening of dynamic stability that has taken place through the preceding frontogenesis, the parameter $2\Omega_z - \partial V_G/\partial y$ should preferably be checked on low-level aerological profiles constructed across the front part that is expected to produce a frontal wave. (On an isentropic plane the horizontal direction of its steepest slope is the y-direction, which is directed northward.) The increase of geostrophic wind in the warm air from the surface front to the front on the 500-mb map can also be used in a rapid estimate of $2\Omega_z - \partial V_G/\partial y$.

The determination of exactly where and when a frontal wave will appear is very difficult, because it is difficult to assess the surface maps for probable future lessening of dynamic stability in the vicinity of the front. So, the forecaster must usually be satisfied with just being able to detect, as soon as possible, an already existing frontal wave on his surface map. In this connection, he therefore should carefully assay the analyzed maps for *indirect indications* of even the slightest wavelike formation, which sometimes cannot be found directly, as the inaccuracies in the analyzed position of the front far exceed its initial wave amplitude. For example, he should (1) examine the analyzed limits of the altostratus-nimbostratus cloud systems for upglide clouds on the cold side of the front, (2) look for protuberance into the cold air of the isallobar which hitherto most nearly coincided with a slowly moving and almost rectilinear cold front, and (3) observe especially the almost rectilinear warm fronts slowly approaching mountainous areas where a forced wave flow may form as a result of orographic drag. On the other hand, he should (4) be mindful that slightly unstable short waves (200-km wave length), already formed, are susceptible to turbulent and orographic influences tending to break them up into smaller and dynamically stable, transient waves. Even waves less than 1000 km in length are generally stable. Only the longer, more unstable waves (1500–2500-km wave length), immune to small-scale local effects, will develop into cyclones.

When a frontal wave has been identified on the surface map, the forecaster must then decide whether or not it will develop into a cyclone. The meteorologist should first bear in mind that, according to synoptic-climatological experience, certain geographical regions are more favored than others for the occurrence of cyclones. For the 1899–1939 series of daily surface

maps, Petterssen has reported how the percentage frequency of cyclones and cyclogenesis is distributed over the whole Northern Hemisphere [62, Figs. 14–21]. He has also given the rate of alternation between cyclones and anticyclones, indicating the distribution of traveling disturbances [62, Figs. 22–23].

The frontogenesis builds up cyclonic shear vorticity which is transferred into vorticity of cyclonically curved flow at the apex of the frontal wave. Large initial frontal shear is therefore a sign of "stored" kinetic energy which, in being transformed to curvature vorticity, favors the evolution of the wave into a cyclone. The intensity of the frontal cyclogenesis may partly be conjectured by the forecaster from the observed horizontal velocity differences between the air masses. In particular, under the assumption of a normal air-mass stratification, sufficiently long waves are usually unstable provided that—according to an old rule—the wind shear along the front in knots is greater than four times the temperature discontinuity in centigrade degrees.

Qualitative examples of such reasoning concerning cyclogenesis due to the individual change of vorticity of traveling particles are demonstrated in Fig. 2. The

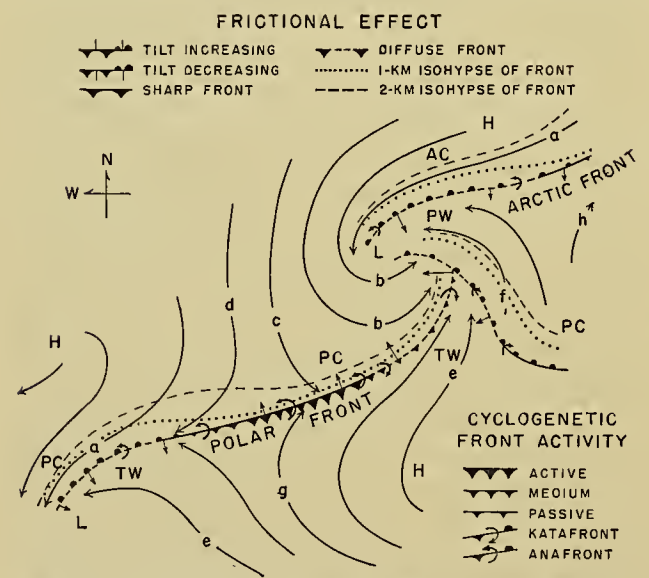


Fig. 2.—Frontal activity and passivity according to Bergeron, and vorticity dynamics of cyclones according to Bjerknes [33].

cold air sample *a* moves equatorward and partakes in the general convergence typical for the forward half of a cyclone. For both reasons cyclonic relative vorticity is acquired, which at first shows up on the surface map mainly as a horizontal cyclonic shear inside the cold air along the warm front. In passing the wave apex, this air maintains its cyclonic relative vorticity, which then is manifested as curvature vorticity. Within the strongly cyclonically curving easterly branch *b* of the cold current, a low-layer frictional convergence and lifting of the polar air contributes toward increasing the slope of the advancing cold front surface as shown by the frontal topography in the figure. Farther behind the apex the cold air samples *c* and *d* are in a region of horizontal divergence strong enough to make their cyclonic vorticity decrease or change into anticyclonic vorticity despite the southward displacement. During the growth of the cyclone, more and more of the cold air is able to maintain its cyclonic vorticity after passing the wave apex, such as represented by the life history *a–b*. Within the anticyclonically curving westerly branch *d* the low-layer subsidence combines with the effect of ground friction to decrease the slope of the front, as shown through frontal isohypses in Fig. 2. The samples *e* and *f*, although moving poleward into their respective cyclones, experience sufficient horizontal convergence to acquire some cyclonic vorticity. The warm air samples *g* and *h*, passing at greater distances from their respective centers, do not have enough horizontal convergence to prevent their acquiring anticyclonic vorticity through poleward displacement. Warm air also ascends to the base of the westerly jet stream aloft and at the same time arrives at the poleward edge of the region of maximum frontal upgliding. Hence, this warm air also gains anticyclonic vorticity from the effect of the vertical velocity terms, which under these conditions is appreciable even in comparison with the effects of both poleward motion and divergence.

Already in 1934 Scherhag [71] had pointed out that a cyclone with strong upper winds deepens. The stronger the upper westerlies have been built up during the frontogenetical period, the more they are apt to form unstable upper waves. A check on the possible occurrence of unstable anticyclonic curvature upwind from the cyclogenetic area will often give a clue to the deepening of the upper cold trough over the rear part of the cyclone. This shows up in the surface map as a deepening of the nonfrontal cold trough and the central part of the cyclone. (For more details, see pp. 587–597 in "Extratropical Cyclones" by J. Bjerknes in this Compendium.)

The instability of the wave varies also with the difference in the *static* stability of the adjacent air masses. In this connection we may mention that, in general, waves on a polar front are more often unstable than those on an arctic front. Since increasing instability of an air mass favors cyclonic deepening, cyclones can deepen in winter by motion from land out over the sea. Moreover, where (in the analyzed upper-air maps) both the *absolute isohypses*[5] and the *total relative isohypses*[5] are close together and the two sets of lines intersect approximately at right angles, cyclogenesis is often occurring in the region of higher *relative height*[5] and lower *absolute height*,[5] according to synoptic experience. In particular, if the spacing of the relative isohypses on the rear side of the low is less than that on the forward side—and this is the normal case—then, according to Pogade [64], the cyclone will deepen. (In the next subsection, we mention Petterssen's rule that the cyclones also tend to migrate toward this anterior concentration of horizontal,

5. For definition see Table I.

isosteric-isobaric solenoids.) The foregoing, rather well-known observation, which dates from 1938, has probably been considered of questionable value for some time, but it serves historically to introduce Sutcliffe's much more recent (1947) and reliable considerations in also applying the *relative hypsography*[5] for estimating sea-level cyclogenesis.

The development of sea-level systems can also be inferred from the relative hypsography. Using certain assumptions based on the vorticity equation, Sutcliffe [76, p. 204] has arrived at the following expression for the divergence of the thermal wind:

$$\text{div } \mathbf{V}' = \underbrace{-\frac{V'}{f}\frac{\partial f}{\partial s}}_{(i)} \underbrace{-\frac{V'}{f}\frac{\partial \zeta'}{\partial s}}_{(ii)} \underbrace{-\frac{2V'}{f}\frac{\partial \zeta_0}{\partial s}}_{(iii)}, \qquad (1)$$

where $\mathbf{V}'$ is the thermal wind vector defined by $\partial \mathbf{V}_\theta/\partial p$, f is the Coriolis parameter $2\Omega_z$, ζ' and ζ_0 are the vertical components of the vorticities of the thermal wind and the geostrophic wind at 1000 mb, respectively, and $\partial/\partial s$ represents differentiation along the relative isohypse. Term (*i*) may also be thought of as, for instance, the geostrophic divergence at 500 mb relative to that at 1000 mb. In other words, it is the divergence at 500 mb in excess of that at 1000 mb. It is thus directly connected with vertical motion.

Term (*iii*) is the so-called thermal vorticity or development term. It is on the basis of this term that we apply the relative hypsography for estimating sea-level cyclogenesis and anticyclogenesis. During the initial period of a quasi-stationary and developing sea-level system, term (*iii*) makes its maximum contribution, as shown empirically by Sawyer [69]. Positive centers of the computed field of term (*iii*) lie to the right, or east of the cold trough in the relative hypsography; negative centers, to the left. The pretrough areas on the surface map are cyclogenetic, the post-trough areas anticyclogenetic. A low will readily pass through a trough and intensify. On the other hand, a "self-developing" system is one with a low in the right half of the thermal trough and a high in the left half. Analogous considerations apply to highs and thermal ridges.

Next, let us apply term (*iii*) to a pattern of concentrated relative isohypses,[5] such as is found in a frontal strip, the so-called "thermal jet." The thermal jet has an "entrance" region and an "exit" region. An example of such a difluent-confluent jet is found in Fig. 3. According to term (*iii*), *there is sea-level cyclogenesis in the colder half of the difluent thermal jet and also in the warmer half of the confluent thermal jet.* (Similar considerations can be formulated for anticyclogenesis.) The genetic regions of the thermal jet thus favor the existence of lows in the confluent warmer half of the thermal jet and in the difluent

TABLE I. CONSTANT-PRESSURE TERMINOLOGY*

(1)	(2)		(3)	(4)	(5)
Terms in columns (2) to (5) refer to:	Adjectives modifying entries in columns (3), (4), and (5)		Quantity	Isopleth	Configuration
	General	Specific			
mandatory surface	*absolute*	*500-mb*	*height*	*isohypse* contour	*hypsography* *baric topography* contour map
	partial (*relative*)	*500/700-mb*†	*height*	*isohypse* contour	*hypsography* *baric topography* contour map
	total (*relative*)	*500/1000-mb*	*height, change in the*§	*isallohypse* contour change-line	*allohypsography*
mandatory layer	*partial*	*700–500-mb*‡	*layer-temperature* thickness	*layer-isotherm* *layer-isentrope* thickness	"thick-thin" map mean-temperature map
	total	*1000–500-mb*	*layer temperature, change in the* thickness, change in the	*layer-isallotherm* *layer-isallentrope* thickness change-line	*layer-isallotherms* *layer-isallentropes* thickness change-line map

* Sets of consistent expressions are given in succeeding columns. Italicized expressions are those used in text.

† 500/700-mb—read as "the ——— of the 500-mb surface above (relative to) the 700-mb surface"; *cf*, *500/700-mb hypsography* is read as "the hypsography, 500 mb above 700 mb."

‡ 700–500-mb—read as "the ——— from 700 mb to 500 mb"; *cf*, *700–500-mb layer* is read as "the layer from 700 mb to 500 mb."

§ Written also as *500/700-mb height change*, etc.

colder half, and favor the existence of highs in the confluent colder half and the difluent warmer half.

In the rear of a marked cold front of a more or less convex shape as seen from the warm-air side, a large cold-air outbreak, moving equatorward and invading regions of warmer air, spreads at the surface of the earth, a process which is connected with the production of *anticyclonic* vorticity in lower strata and the sinking of the center of gravity of the whole system of cold and adjoining air. It is obvious that such anticyclogenesis will be more efficient, all other factors remaining the same, the greater the temperature difference between the cold and the warm air and the more high-reaching and vast the cold outbreak. Although the *real* arctic air does not reach so high, synoptic experience seems to indicate that the outbreaks of arctic air often form more intense and stable anticyclones than polar air outbreaks. This may be explained by the fact that the whole process attending a pronounced outbreak of the arctic front equatorward leads to a much greater drop of temperature in temperate latitudes than a corresponding advance of the polar front. The arctic air and the overlying polar air in the former case have their origin in much higher latitudes.

Anticyclogenesis depends upon the supply of kinetic energy made available to it by the juxtaposition of tropical air over the polar air in which the anticyclone has first been formed. The forecaster should be on the lookout for regions on the constant-pressure maps where a zonal upper westerly with very high velocity is bounded (to the north and to the south) by two regions with relatively low zonal velocity. Elliott [23] has pointed out that an intensifying belt of abnormally strong upper westerlies exists for several days prior to the anticyclogenesis, and reaches a maximum just before it. Associating this development with an increased north-south thermal gradient across the region, he suggests that such anticyclogenesis is associated with excessive heat exchanges across the normal westerlies between tropical and polar air. There are sometimes other clues which are helpful in anticipating anticyclogenesis. On the surface maps, there is often the development of low pressure, a few days prior to anticyclogenesis, at half the distance of a stationary long wave (*i.e.*, about thirty degrees of latitude) upstream or downstream.

Using the quasi-static assumption, we see that the pressure change at the base of the atmosphere is associated with the depletion or accumulation of air aloft. In particular, the pressure tendency at the ground is generally decided as to sign by the horizontal (mass) divergence or convergence above the level of nondivergence (see the following paragraph for substantiating evidence). This divergence (or convergence), acting in the same sense through the stratosphere and the upper half of the troposphere, may thus be an important effect to consider in the development and movement of sea-level systems. The patterns of velocity divergence are of course determined by the distribution of the instantaneous ageostrophic upper flow. As will be shown in a following subsection on the prediction of winds, a relatively large indraft of the streamlines across the absolute isohypses occurs in the right half of the *entrance region* and large outdraft in the right half of the *exit region* (see Fig. 3*a*). On the other hand a relatively small angle of geostrophic deviation (indraft and outdraft, respectively) is found in the left half of the entrance and exit regions. Hence there is velocity divergence in the exit (or delta) region, convergence in the entrance (or confluence) region. Provided that the distribution of the horizontal mass divergence can also be represented by this "delta-divergence" and "confluence-convergence," then the pressure tendency at the ground should be positive

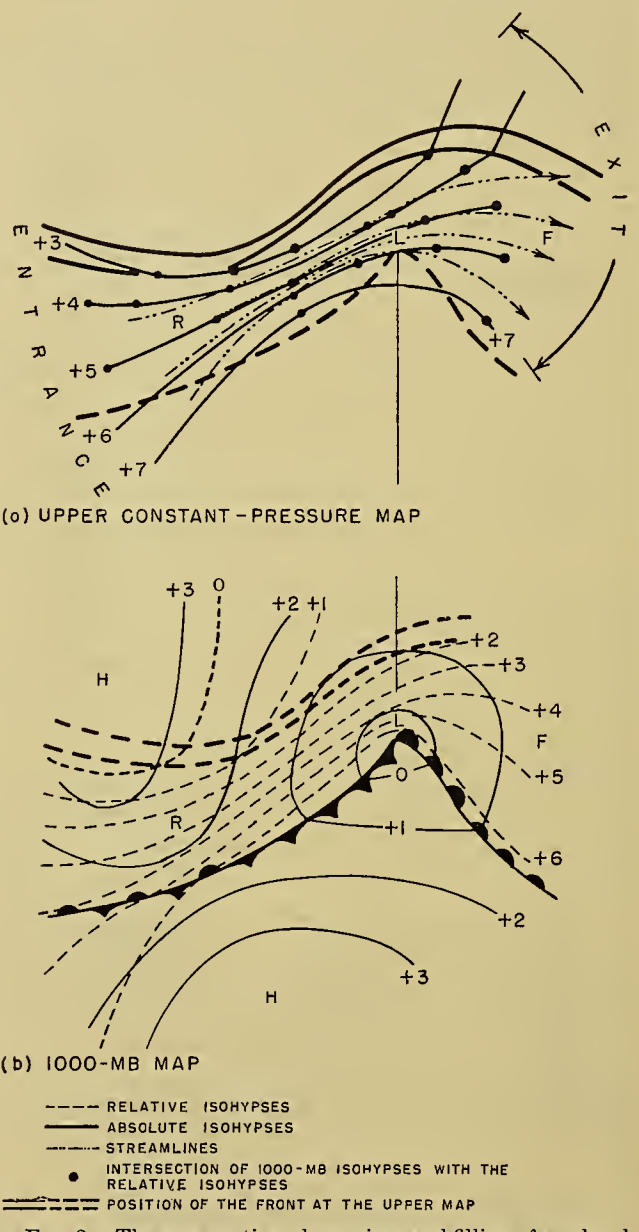


FIG. 3.—The propagation, deepening, and filling of sea-level pressure disturbances (*b*) as determined from the absolute hypsography (*a*) of an upper constant-pressure map, according to Ryd-Scherhag [71].

in the confluence region, negative in the delta region, as shown in Fig. 3*b*.

In a study of the relation between sea-level pressure change and the upper-air density change, Hess [39] has found that intensifying cyclones are associated with upper-air local density decreases, primarily in the stratosphere. Corroborating Hess's observations, Vederman [80] particularizes that the partial relative height of the upper one-third of the atmosphere (by weight) increases markedly over rapidly deepening lows. Moreover, Austin [5] observed that there is a tendency for the stability to increase over deepening cyclones in the lower stratospheric layers from 400 mb to 100 mb. These two relationships reported by Hess and Austin both corroborate the previously mentioned fact that the exit (divergence) region above a fairly low level of nondivergence determines the sign of the pressure tendency at the ground, namely falling pressure on the surface map. The entrance region similarly should be associated with rising pressure at the ground.

Upon the relationship between the ageostrophic upper flow and the local changes in sea-level pressure, Scherhag [72, pp. 23–26] already in 1943 had introduced some interesting rules for predicting the deepening and filling of sea-level pressure systems. These rules were formulated from his synoptic experience which had been guided somewhat by Ryd's earlier (1923) theoretical contributions on the subject. Although some of his rules still seem plausible in the light of Sutcliffe's more recent investigations [76], a good deal of skepticism regarding them now exists. Still, for historical and illustrative purposes, we have summarized below some of his more popular rules.

1. In the region where, on the constant-pressure maps, the absolute isohypses diverge markedly, such as at *F* in Fig. 3*a*, there is falling pressure on the surface map (*cf.* at *F* in Fig. 3*b*). On the other hand, the convergence of the upper absolute isohypses (*cf.* *R* in Fig. 3*a*) is associated with rising pressure on the surface map (*cf.* *R* in Fig. 3*b*).

2. If the divergence of the upper absolute isohypses is vertically above a low center (*i.e.*, if the low in Fig. 3*b* were displaced rightward to the position of *F*), then this center has a tendency to deepen. A high center beneath converging absolute isohypses likewise increases.

3. Cyclogenesis on the surface map usually occurs in those regions where, on an upper constant-pressure map, the absolute isohypses diverge. Anticyclogenesis is similarly associated with converging absolute isohypses. On the surface map, frontal waves which move toward the exit (entrance) region will intensify (weaken).

4. If (as in Fig. 3) the frontal strip, the jet stream above the frontal strip, and the entrance and exit regions of the jet stream all have the same orientation, then the sea-level cyclones will not deepen. Many stable waves on such a front are likely, but no waves are likely when the upper flow is mainly perpendicular to the front.

5. If, as indicated by the increased local crowding of the relative isohypses into the frontal strip (*e.g.*, the sharpening of the frontal zone by the approach of a new front), a front becomes more sharply defined, the jet stream is intensified, and so are the convergence and divergence in its exit and entrance regions. (As will be shown in the later subsection on wind (pp. 791), an increased local crowding of the absolute isohypses tends to produce large positive or negative angles of geostrophic deviation in the exit and entrance regions, respectively.) Then the *fall*[6] in the exit region and the rise in the entrance region will both intensify.

6. If, as in Fig. 3*a*, the warm-air advection in the exit region is less than that in the entrance region, the absolute isohypses in the upper jet stream then, according to Scherhag, diverge more in the exit region of the jet stream than they converge in the entrance region. As a result the wave cyclone *L* in Fig. 3 intensifies and deepens rapidly. Similarly, if the warm-air advection appears to be greater in the exit region than in the entrance region (strong warm front, weak cold front), the absolute isohypses in the upper stream then converge more in the entrance region than they diverge in the exit region. As a result the cyclone will weaken and fill.

With warm-air advection in the entrance region, the spacing of the relative isohypses decreases and their cyclonic curvature increases. With cold-air advection in the exit region, the spacing and anticyclonic curvature of the relative isohypses increases. Where there is a sharp reversal of the upper flow, the fall is usually weakened, the rises strengthened.

7. Deepening occurs in a low where there is cyclonic shear in the cyclonic flow aloft (Fig. 4*a*). A surface frontal wave does not develop where there is anticyclonic shear in the cyclonic flow aloft (Fig. 4*b*). Surface anticyclogenesis tends to occur where there is anticyclonic shear in the anticyclonic flow aloft (Fig. 4*c*). Highs do not develop where there is cyclonic shear in the anticyclonic flow aloft (Fig. 4*d*). If such a type of upper flow moves over a high, or if it forms there, the high will dissipate.

The foregoing rules may be modified by several compensating processes.

1. As will be shown in the later subsection on winds (see p. 791), the conditions for a relatively small ageostrophic flow (leading to small amounts of divergence) are (1) that the horizontal shear of the geostrophic wind be cyclonic or only slightly anticyclonic, and (2) that, if the horizontal shear is anticyclonic, the curvature of the path be cyclonic or only slightly anticyclonic. A pronounced cyclonic curvature of the absolute isohypses at a *stationary* trough must therefore be associated with their marked divergence. Long and narrow V-shaped troughs in the upper absolute hypsographies, such as in Fig. 5, should be quasi-stationary.

6. The expression *fall* refers to the center of maximum decrease in sea-level pressure. If, as in this case, the length of the period is not indicated, it is assumed that the change under consideration is valid for any period from three to twenty-four hours.

2. From the foregoing rule, as well as from synoptic experience, Rodewald [66] had observed (1937) that a rise (fall) is weakened as it approaches the position of the trough (ridge) in the upper absolute hypsography.

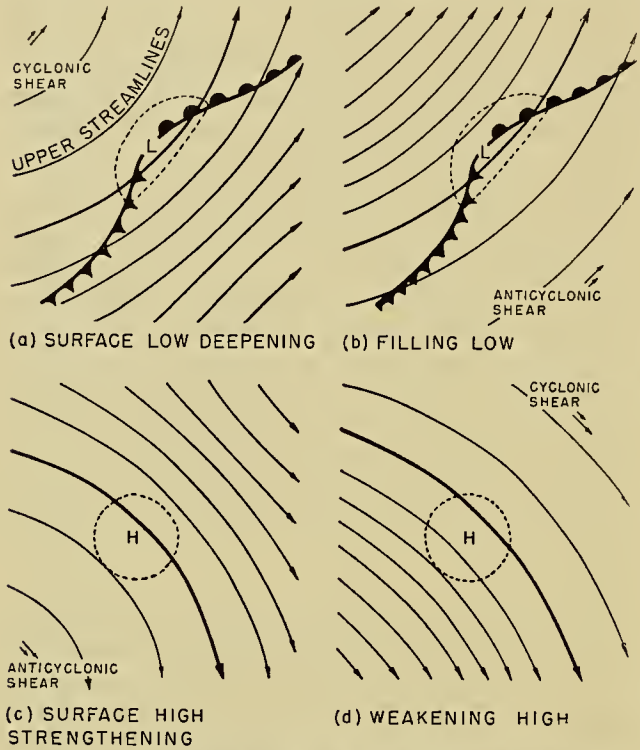


Fig. 4.—The deepening and filling of disturbances at sea level and the shear in the flow aloft.

3. A sea-level pressure trough or ridge will weaken if the surface and upper-air maps show that its axis is vertical or nearly so. This hypothesis, which follows

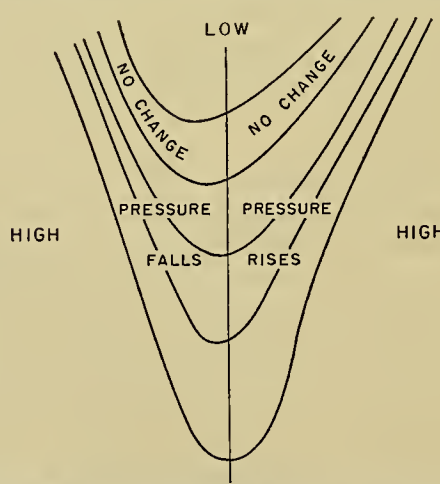


Fig. 5.—The compensating effects of the curvature and spreading of the absolute isohypses of a quasi-stationary upper trough. (*After Scherhag* [72].)

from the two foregoing rules, is correct if the frictional effect of the earth upon the surface flow is not compensated by other diabatic effects or by an appreciable transfer of angular momentum.

Finally, some physical reasoning should be involved in estimating accelerations in the deepening (filling) of the pressure system. If, for instance, the surface map shows a wave in its early stage of development, the forecaster would advisably use a *greater* average deepening per 3 hr for the coming 12 hr or 24 hr than the rate found either by formulas or from the map sequence at the past 3 hr. In other words, a "deepening acceleration" should be estimated according to the particular stage in the life history of the chosen tropospheric model and applied qualitatively as a correction to the simple first-order extrapolation of the actual pressure systems.

The Movement of Sea-Level Pressure Systems. A young cyclone generally propagates in the direction of its warm-sector isobars. An older low (high) with strongly diverging, noncircular isobars tends to have a straight path in a direction between its longer axis of symmetry and the isallobaric gradient (ascendent); the more the isobars diverge, the closer is the path in the direction of this long axis, according to Philipps [63]. A low (high) with approximately circular isobars progresses in the direction of the isallobaric gradient (ascendent), that is, normal to the isallobar through its center or, according to synoptic experience, sometimes a little to the left; it may also be extrapolated toward the domain of subnormal (supernormal) winds. Temperature, as well as pressure and pressure tendency, should be considered. For example, a stable wave progresses along a front in the direction of the warm-air flow, while a cyclone with sharply defined zones of temperature maximum and minimum moves normal to the temperature gradient. The center of an anticyclone shifts in the direction of most rapid decrease of temperature. In the usual case of cold air poleward of a front along which warm air is flowing eastward, the eastward paths of the successive members of a cyclone family—the duration of which is roughly 5–6 days—have a progressive translation equatorward. Anticyclones and ridges of high pressure which lie between cyclones move generally in the same direction and with the same speed as the frontal cyclones. But anticyclones which are cut off from a series have an equatorward component of motion. Moreover, a cyclone seeks to revolve in a clockwise sense around a neighboring stationary anticyclone, while two cyclonic disturbances of the same intensity tend to revolve about each other in a counterclockwise sense.

Centers with a linear pressure profile or a strong pressure gradient may be extrapolated with a constant propagation speed which, for a circular center, is directly proportional to the isallobaric gradient and inversely proportional to the curvature of the pressure profile. The speed of stable waves and of cyclones with weak pressure and isallobaric gradients is greater than that of a deepening cyclone, which in turn moves more rapidly than a filling cyclone. Centers with a strong pressure gradient move even more slowly; a center is quasi-stationary if it is concentric with a symmetrical isallobaric center. The cyclone should be extrapolated with an increasing (decreasing) speed of propagation

if the curvature of its tendency profile is anticyclonic (cyclonic). An anticyclonic center experiences an acceleration (deceleration) if the curvature of the tendency profile is cyclonic (anticyclonic). Often a depression with very strong winds on its forward side tends to become stationary and rapidly disappears.

The foregoing extrapolation of the motion of the sea-level pressure systems is based largely upon a recognition of certain events which had been observed to occur within the system in question (at the surface of the earth). However, the well-known *steering* "principle" or hypothesis had early been advanced that the ageostrophic upper flow, and the accompanying centers of change in the sea-level pressure, propagate with the real upper flow at a certain pressure surface, called the *steering surface*. Strictly speaking, this principle applies only to *rises* and *falls*, respectively, and not to the *highs* and *lows*. As a preliminary remark we caution that pressure changes represent processes and therefore are liable to undergo modifications. They are not systems passively drifting with the flow. Even so, the movement of the sea-level systems is assumed to be closely associated with the pressure field at the steering surface, that is, with certain events observed *outside* the systems on the surface map. The following historical and critical account of steering—introduced in 1910—considers the Scherhag complex and culminates in a discussion of Sutcliffe's practicable rules of thermal steering.

In 1931—three years before daily upper-air maps had been introduced in Germany—Mügge [53] had already formulated the loose rule that *the rises and falls follow downstream the 500-mb isohypses, the speed of their propagation being roughly half the 500-mb gradient wind.* The 3-hr rises and falls seem to move in the direction of the 700-mb flow, while the 24-hr rises and falls move with the 500-mb flow. The movement of rises and falls, which are associated with the young and shallow pressure disturbances (highs, lows, troughs, and ridges), is very roughly determined by Mügge's empirical relation so long as the development of the highs and lows in the hypsography of the steering surface is negligible. The falls have a tendency to follow the curved 500-mb isohypses better than the cuspal 500-mb isohypses, which the falls tend to cut across. On the other hand, the rises are inclined to follow the cyclonically curved 500-mb isohypses and to depart from those curved anticyclonically. By their thermal influence in transforming the upper pressure field, the falls swerve to the left, the rises to the right of the established upper flow. The rises and falls of the older and more developed disturbances are observed to move in the mean or predominant direction of the wave-shaped flow at about 400 mb to 300 mb. In this connection, constant-pressure maps higher than 500 mb should be consulted to select the steering surface with relatively straight absolute isohypses in the regions of the older and deeper disturbances. Pressure systems with closed isohypses up to the stratosphere, such as an occluded cyclone or a blocking high, are themselves considered steering centers, so that there is no well-defined relation between the 500-mb flow and the movement of their rises and falls. If a closed upper low is formed through a strong pressure fall at the surface of the earth, the fall tends to split, with the rapidly weakening part swerving to the left, the stronger to the right, of the old path (occlusion stage of development). The cold lows in the upper relative hypsographies (*Kaltlufttropfen*) are not steered by the upper flow, but their movement can be predicted from other considerations (see Fig. 6).

For preparing the prognostic map of 24-hr change in sea-level pressure, the following empirical relation was first suggested by Guilbert and Grossman and later formulated by Rodewald [66]. *The 24-hr fall (rise) for t_0 moves downstream in the direction of the 500-mb isohypses for t_0 so that its position at $t_0 + 24^h$ is on the axis of past 24-hr increase (decrease) in the 500/1000-mb hypsography*[7] *for t_0.*[8] The prognostic sea-level pressure map for $t_0 + 24^h$ can be immediately obtained by simply adding graphically the prognostic map of 24-hr change in sea-level pressure (obtained approximately by the Guilbert-Grossman rule) to the sea-level pressure map for t_0.

In 1939 Ertel introduced his theory of singular advection of the first order [26], according to which atmospheric pressure variations originate at discontinuities where changes of momentum occur. The best-defined and most persistent temperature discontinuity surface of the first order is the tropopause. By disregarding all other discontinuity surfaces (*viz.*, fronts), Lucht [50] has attempted to adapt and apply Ertel's theory at the tropopause in order to find how the upper atmosphere is associated with variations in the sea-level pressure. Essentially, the adapted equation of Ertel gives the individual sea-level pressure changes as a quantity depending on (1) the solenoidal field of the mean temperature of the troposphere, (2) the hypsography of the tropopause, and (3) the stability discontinuity of the tropopause. The tropopause discontinuity of the temperature gradient and inclination of the tropopause are apparently the most influential factors in deciding the speed of movement. The conclusions from Lucht's investigation indicate, for the rises and falls, that the direction of motion can be predicted from the hypsography of the tropopause and that Ertel's equation [26, eq. 50, p. 406] gives satisfactory values for the speed of motion. But Lucht's efforts to apply Ertel's equation are not persuasive, because of the sparsity of the data in 1939 for determining the slope of the tropopause and the permanent difficulties in determining its temperature discontinuity. Lucht has thus attempted to substantiate and refine the observations made already in 1926 by Stüve and Palmén [56] and by von Ficker [29] that the rises and falls move in the direction of the isohypses of the tropopause.

In an effort to extend the steering hypothesis to lows, Austin [5] has observed that the steered lows (1)

7. Defined in Table I.

8. Variously stated in [9, 66].

tend to continue to be steered (but nonsteered lows may eventually move parallel to the steering flow), and (2) usually veer relative to the steering flow. In this study, he has also observed that (3), especially for filling lows, the angular deviation of their movement from the steering flow is generally independent of their intensity (number of closed isobars) and less for cyclones than for anticyclones, and that (4) this angular deviation increases as the upper flow departs more and more from straight flow. Longley [49] has found evidence to state that strengthening highs move toward the left of the steering flow. It is also possible, as shown by experience, to predict the deepening and filling of the highs and lows.

The motion of sea-level systems can also be estimated from the relative hypsography. Term (*iv*) in equation (1) is the so-called thermal steering term. This term indicates *the tendency for the sea-level systems to be steered by the relative isohypses, that is, to move in the direction of the thermal wind with a propagation speed proportional to the thermal wind.* Sawyer [69, p. 113] has shown empirically that term (*iv*) becomes increasingly important as the sea-level system develops. In practice, the speed of displacement by thermal steering can at best be assessed by the forecaster only from experience and statistics. The failure in practice of term (*iv*) to give the correct propagation speed is due to the neglected (1) stabilizing control of the vertical motion, (2) diabatic processes, and (3) variation of the vorticity of the thermal wind (term *iii*). As we have already noted in the previous subsection, the genetic regions of the thermal jet thus favor the existence of lows in the confluent warmer half of the thermal jet and in the difluent colder half and favor the existence of highs in the confluent colder half and the difluent warmer half. Consequently cyclones tend to swing out of the thermal jet toward lower relative height, anticyclones toward higher relative height—and slow down.

The veering of lows relative to the steering flow, as cited by Austin [5], may not be in disagreement with the above-mentioned backing of lows relative to the thermal jet. In the baroclinic region just ahead of a low, the thermal jet veers relative to the steering flow (*i.e.*, the sense of the horizontal, isosteric-isobaric solenoids is negative). So, by combining Austin's observations with our interpretation of term (*iii*), we should expect the lows to move in a direction which is somewhere between that of the steering flow and the thermal flow. In other words, the lows move toward the concentration of horizontal, isosteric-isobaric solenoids. This tendency for lows to migrate toward the solenoid concentration has been physically confirmed by Petterssen [62, p. 136, eq. 12]. By applying the horizontal del-operator to the vorticity equation and considering only horizontal motion, he found an expression for the horizontal ascendent of the vorticity tendency. Lows, being centers of maximum relative vorticity, must move in the direction of this ascendent. According to Petterssen's expression for this ascendent, lows have a component of motion toward the concentration of horizontal, isosteric-isobaric solenoids; opposite considerations apply to the highs.

Frontal Occlusion. The frontal wave, even in the case of a long wave length, is a fairly long-lived system, whereas the so-called ideal cyclone with a narrow but unoccluded warm sector exists only for about 6–18 hr. Thus, a wave of a quite innocent appearance may well occlude into an intensive cyclone during the next prognostic period (t_0 + 24–36^h).

At the very first stage of occlusion, particularly in the lower layers and near the cyclonic core, the converging and ascending precyclonic air is colder than the diverging and subsiding postcyclonic air. These dynamic effects should favor the formation of a warm-front occlusion. Later, especially at higher levels and in the outer region of cyclones which are propagating eastward at temperate latitudes, the equatorward current of postcyclonic polar air is usually colder than the poleward, precyclonic current, either because it originates in higher latitudes or because its recent life history is colder. In Europe during the summer (winter), the continental precyclonic air of the ordinary west-east cyclone is warmer (colder) than the maritime postcyclonic air, so that the forecaster may expect the occlusion to be of the cold-front (warm-front) type. In the eastern United States the foregoing conditions should, of course, be reversed. For cyclones moving in other directions than towards the east, the forecaster can still predict the type of occlusion by systematically using the general temperature distribution as shown by the surface and upper-air maps. At the last stage of occlusion, the rapidly sinking tropopause and the marked anticyclonic motion within the upper air result in a general tendency towards subsidence and dynamic warming of the precyclonic air, while in the postcyclonic air a tendency exists towards ascending motion and dynamic cooling.

Whereas a cyclone neither occludes nor deepens so long as the warm-sector tendency is zero, according to Petterssen [60, p. 433] the occluding rate of a young cyclone is proportional to its deepening and to the negative tendency in its warm sector. Moreover, if a cold front flows slowly over a mountain range normal to its path, an orographic occlusion may occur. Upon occlusion the path of the cyclone deviates to the left, moving in the direction of the isobars of its warmest part (the "false" warm sector). Once the occlusion process starts, the propagation speed of the cyclone diminishes. If an occluded front approaches a stationary continental anticyclone from the west, its movement is retarded. Finally, the forecaster should watch especially for secondary disturbances often arising at the point of occlusion.

Degeneration and Regeneration of Cyclones and Anticyclones. The potential energy of adjacent air masses is gradually consumed during the occlusion process, so that the cyclone normally degenerates during this stage. But the occluded cyclone and its fronts may be regenerated either by the same general frontogenetical effects that, according to synoptic experience, create the secondary cold front at the end of the bent-

back occlusion or by the influence of the convergence at the front which causes ascent and the release of new potential lability. In Europe during summer and in the eastern United States during winter, these factors are especially effective at the cold-front occlusion. As a result of a "false" warm sector forming in its bent-back trough, the warm-front occluded cyclone acquires not only additional potential energy of horizontally adjacent air masses but also a mechanism (*viz.*, the bent-back occluding process) which can transform this potential energy into kinetic energy in the form of increasing cyclonic vorticity. The consequent conditional instability of the air mass within the false warm sector and the existence there of labile energy favors the regeneration of the cyclone. In the eastern United States this kind of regeneration is presumably more frequent in summer than in winter. A dying cyclone also deepens anew when it comes under the influence of a fresh cold air mass or when there is a general sharpening of its fronts. An apparent regeneration of the occluded cyclone often takes place when a new wave of the main front overtakes the occlusion. If an occluded cyclone regenerates, its speed of propagation is increased.

As the anticyclone develops, both its movement and its anticyclogenesis decrease, so that the anticyclone and cyclone cease to resemble each other in the manner of their development. In particular, the duration of the anticyclone is highly variable, from a week or so to a month or even longer. The anticyclone, now stable both physically and dynamically, assumes a semifixed geographical location. In the stationary anticyclone, the subsidence inversion intensifies and lowers, hindering convection. The transformation in properties of its air mass has reached a quasi-equilibrium state as the air mass, now thoroughly homogenized, is adapted to the various steady-state processes operative in its region. The anticyclone imposes decisive restrictions on the movement, development, and position of adjacent cyclone systems. Certain types of anticyclones exhibit a remarkable permanence, in time as well as in space. In mature anticyclones, the importance of this stability in regulating the remote weather is paramount. Persistent abnormal weather is usually an indirect consequence of the permanent nature of the mature anticyclone.

The essential difference between the cyclone and the anticyclone is the fact that the latter is usually conterminous with a single air mass. One important characteristic of the anticyclone, found in classifying the air masses according to their source regions, is the homogenizing influence of the quasi-permanent subtropical anticyclones upon the migratory cold anticyclones, which are absorbed into these warm subtropical highs. The success of weather prognosis depends, to a large extent, upon the somewhat orderly behavior of the cyclone. The maximum extent of time in which detailed weather prognosis is possible is determined partly by the time of evolution of the cyclone. The weather immediately associated with the anticyclone is not so hazardous as the cyclonic weather; and, partly for this reason, anticyclones are not known so well as cyclones. Anticyclones with divergent motion are nevertheless primarily instrumental in the transformation of the solar radiative energy received by the atmosphere into the kinetic energy of its motion relative to the earth. They are sources, then, for the conversion of geopotential energy or internal heat energy into horizontal kinetic energy [75]. From them the processes of advection and work by pressure forces are continually and rapidly transferring this kinetic energy to cyclonic areas, that is, to the centers of major weather activity.

Local Modifications of Air Currents, Air Masses, and Fronts. Whenever possible, the forecaster should apply the foregoing models while working out the prognostic maps. Sometimes, however, their application must be introduced between the preparation of the prognostic map and the formulation of the forecast. This introduction may be particularly necessary because of important (thermal and orographic) local modifications to the idealized, normal evolution of the lower-tropospheric models. These disturbances are particularly important to the area forecaster; but they often operate on a scale so large that, if ignored by the regional forecaster, they would introduce grave errors into the general forecast.

When an air mass moves over a colder (warmer) surface, its stability (instability) is increased. An air mass becomes stable in winter (unstable in summer) by motion from sea to land; for motion from land to sea the reverse is true. The stability is increased (decreased) for an air mass moving poleward (equatorward) over an isothermal portion of the surface of the earth. The air-mass stability (instability) of a stationary continental anticyclone increases from day to day in fall and winter (spring and summer). Especially near the ground, the properties of warm and cold air masses show a marked diurnal variation, because they depend mainly on the vertical lapse rate of temperature. At night (in the afternoon), a warm (cold) air mass will have all its properties accentuated and a cold (warm) air mass will, in the lowest layer, temporarily acquire the properties of a warm (cold) air mass: uniform wind, poor visibility, stratiform clouds, even drizzle occasionally in the case of a warm air mass; gusty wind, improved visibility, cumuliform clouds and showers in the case of a cold air mass. Over sea the effects will be the opposite, although much less marked.

Moreover, periodic wind systems are often created owing to differences in the diurnal temperature variation over land and sea, and over mountain and plain these wind systems may become both intensive and extensive enough to influence even the general forecast. The more stable a moving air mass, the more it tends to converge horizontally about an obstacle such as an elevated part of a continent bordering a plain or a lowland bordering the sea. To the low-pressure side of a stable warm air-mass current, the forecaster should thus predict an extraordinary intensification of the wind at the "corner" of the obstacle—the *corner effect*—and of the "coastal" convergence and precipitation at

the windward coast. Orographic showers or rain may thus be superimposed upon the idealized precipitation areas of the models. Leeward of the mountain range, on the other hand, the foehn effect may lead to a dissolution of the idealized precipitation and cloud areas of the models. Mountain chains in general obstruct the rapid motion of all fronts; the hindrance is greater the higher the obstruction and the less the vertical extent of the front itself. But a swiftly moving cold front will flow over a low obstruction without especial deformation.

Frontolysis, Cyclolysis, and Anticyclolysis. When the cold air lies to the right of the general flow, frontolysis occurs in the lower portion of the front. A front which is traveling normal to the general flow aloft is very subject to frontolysis. A front dissipates as it departs from a deep pressure trough.

As a consequence of the convective mixing of the two air masses below and above a frontal surface, the thermodynamic effect generally will tend to smooth and delete any weak fronts rather than to create new ones. A *medium-* or *high-speed cold front* (mostly found in the interior of an already existing cyclone) will generally be a katafront in the upper layers (see Fig. 2). It will then be subject to frontolysis, tending to make the frontal surface diffuse and thus dynamically ineffective (even if the effect of friction upon the motion tends to increase its slope and thereby presumably the horizontal shear). A *medium-* or *high-speed warm front* is generally an anafront, but the friction will tend to diminish its slope markedly, at least in the lowest kilometer of the atmosphere and thus, according to synoptic experience, also reduce the horizontal shear.

After occlusion, the deepening of the cyclone ceases and its filling commences, provided that a false warm sector does not occur. The existing pressure gradient weakens and the cyclone fills more or less rapidly when it possesses supergeostrophic, divergent flow (Hesselberg [40] and Guilbert [36]). In particular a cyclone with marked supergeostrophic wind on its forward side fills during the following 12–24 hr. With weak upper flow around its periphery, a cyclone will fill. If an upper low reaches to the surface without especial tilting of the axis, the surface low sometimes fills.

A young anticyclone weakens when approached by a cold front, particularly with weak upper flow around its periphery. The degeneration of the stable anticyclone, unfortunately, can seldom be anticipated in advance. This fact is especially regrettable, for seldom is a reliable forecast more insistently demanded than during the prolonged periods of abnormal weather due to the blocking action of the mature anticyclone. The anticyclolysis and the movement of such an anticyclone is generally desultory. Sometimes the weather regime of the blocking anticyclone appears to be at last giving way, as the anticyclone starts to move and dissolve, only to be renewed again as it restrengthens and meanders back to its original position. In fact, in closing this section on surface map prognosis, we mention that another weather forecasting deficiency on which much future study is needed concerns our lack of knowledge about anticyclone development.

THE PROGNOSIS OF THE UPPER-AIR MAPS

The upper-air prognosis of the pressure field proceeds in the same way as the analysis. By the layer method, we first prognosticate the partial relative hypsography of a mandatory surface, then graphically add it to the prognostic absolute hypsography of the adjacent and lower mandatory surface. By the layer method the forecaster builds upward one prognosticated mandatory layer on top of another, beginning with the prognostic surface map as the base. This method not only requires that the prognostic absolute hypsographies of all mandatory surfaces for a given time be mutually consistent and hydrostatically interrelated, but also provides a means of developing the baric prognosis upward from the more reliable prognosis of the surface map to the less predictable future state aloft. The main attention is directed toward prognosticating the pressure field at sea level and the relative hypsographies of each mandatory surface. The upper-air prognosis of the pressure field is then based solely upon *predicting locally* the rate of change in the partial relative height. *Upper-air prognosis is thus concerned with extrapolating the relative hypsography on the basis of an estimate of its future over-all rate of change.*

The Geometrical Extrapolation of the Relative Hypsography. Closed centers of both low and high values in the relative hypsography, so-called *thermal lows* and *thermal highs*, respectively, and formerly termed "cold-air drops" (*Kaltlufttropfen*) and "warm-air drops," are often observed to occur in the partial and total relative hypsographies of higher level constant-pressure maps. These thermal lows are important *conservative* features of the constant-pressure maps. The same holds for the thermal highs. In prognosticating the relative hypsographies of the constant-pressure maps, therefore, attention should be given to preserving the real continuity in time of these models. *Stage three*, then, in the order of our prognostic operations is the *geometrical extrapolation of these thermal lows and highs.*

A series of actual surface and upper-air maps showing the typical features of the thermal centers are shown in Fig. 6, which is taken from a paper by Schwerdtfeger [73]. Following the cold front, the thermal low *C* in the 500/1000-mb hypsography moves (southeast) for 2½ days with remarkable regularity and little change in size and value—536 gpDm to 540 gpDm (geopotential decameters). As a second example we may refer to the 2½-day movement of the 500/1000-mb low in Fig. 8*b*, the innermost closed 516 gpDm isohypse of which remained about the same size during that period. It has been generally observed that this low moves not with the velocity of the air in which it exists but roughly with the horizontal velocity of the cold-front surface underneath it. Although the island of cold air in a moving upper low-pressure center

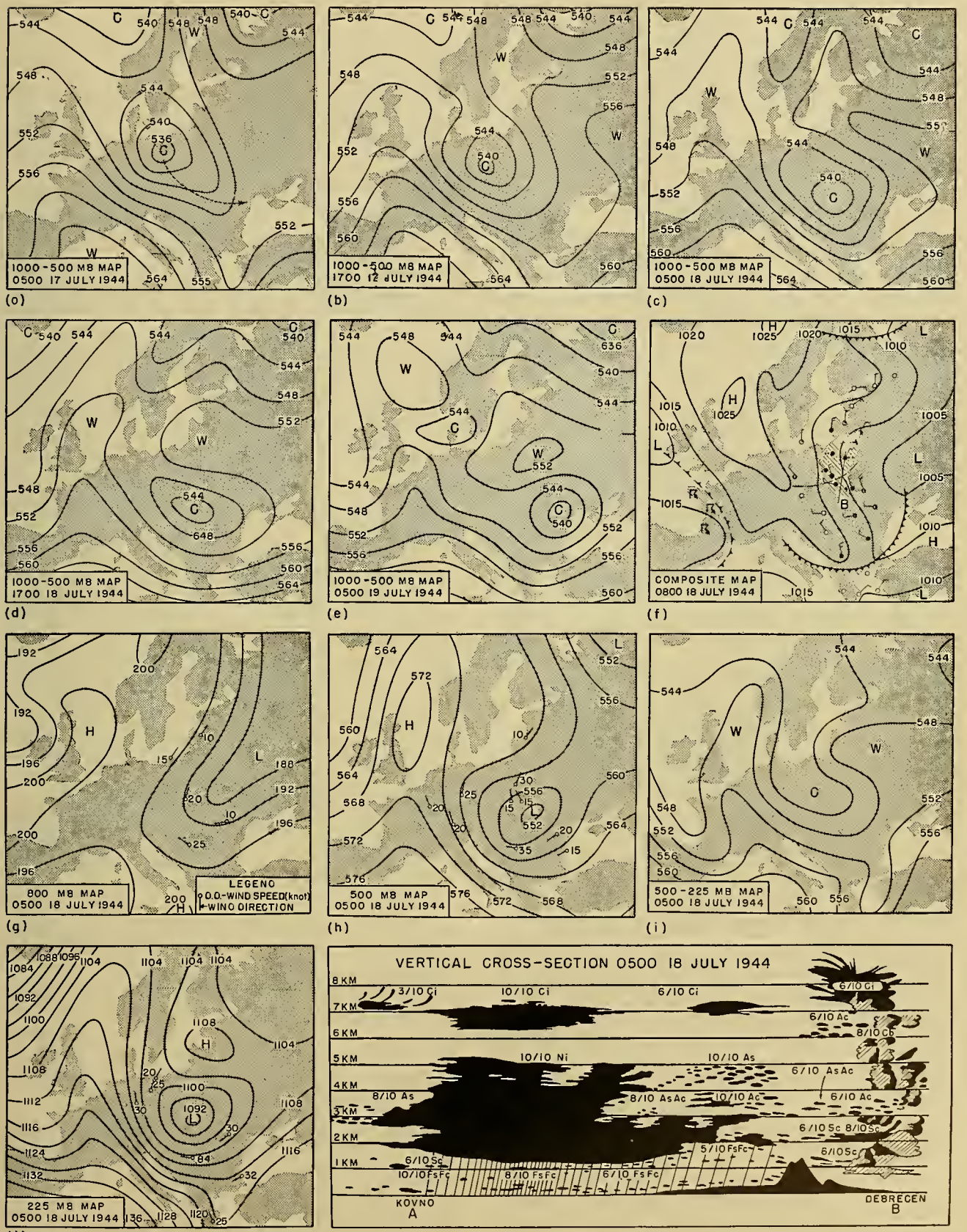


FIG. 6.—The thermal low. Units are geopotential decameters on the upper-level charts and millibars on the composite map.

would be advectively advanced to the right of the upper low (see Fig. 7), the anterior ascent of air in the low and posterior descent alone could maintain a concentric system of relative and absolute isohypses,

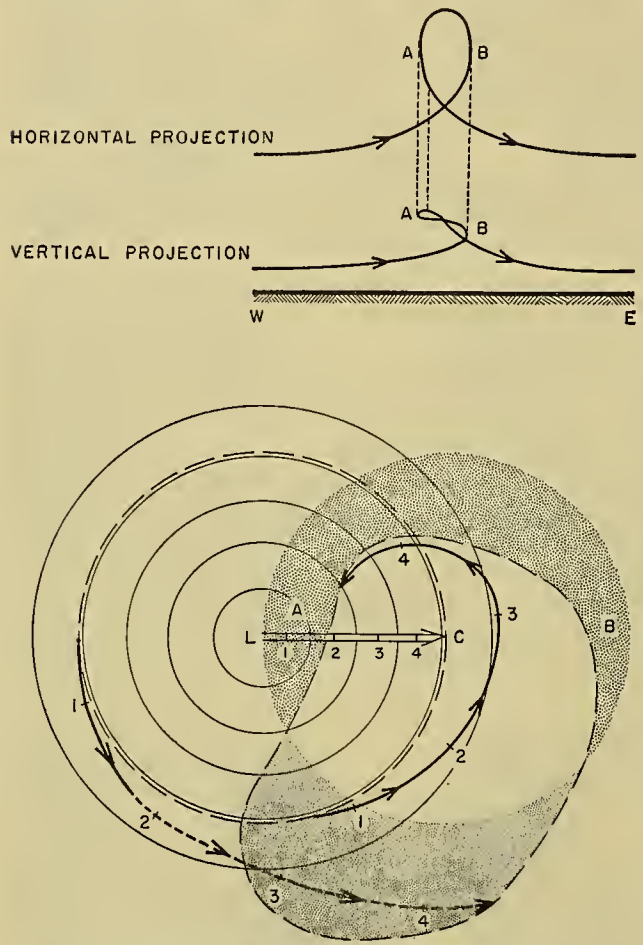


FIG. 7.—Three-dimensional motion in a nondeveloping upper low concentric with its island of cold air. In the lower part, the continuous concentric circles are the absolute isohypses of an upper low at, say, t_0. The dashed circle congruent with an absolute isohypse is a relative isohypse, also for t_0, of an upper island of cold air. The double-stroked arrow is the subsequent displacement of the nondeveloping cold low from L to C, while the two curved single-stroked arrows give the resulting trajectories of two selected points on the relative isohypse, according to the field of gradient motion of the propagating cold low. When continuous, the single-stroked arrows also indicate upward motion; when dashed, downward motion (*cf.* upper part of figure). The displacements and trajectories during four successive 6-hourly periods are indicated by the numbers (1, 2, 3, and 4) along all three arrows. With C as its center, the circle indicated by the shaded areas is the position of the relative isohypse at $t_0 + 24^h$. The eccentric, dashed closed curve is the advected position of the relative isohypse also at $t_0 + 24^h$. The shaded areas therefore indicate the nonadvective displacement of the relative isohypse—light-shaded for nonadvective increase in relative height, dark-shaded for nonadvective decrease. The letters A and B are references to the upper part of the figure.

(even disregarding the contributing diabatic effects). The thermal low thus appears as a sort of cold-air source region, mainly in the middle tropopause just above the cold-front surface. Most important in upper-air map prognosis, this conservatism of the cold-air islands allows for their geometrical extrapolation. In fact, the forecaster, at least in the beginning, devotes considerable attention to the thermal lows, treating the prediction of them as a mainstay around which the map prognosis can be built.

The Prognosis of the Relative Hypsography by Advective-Adiabatic Extrapolation. Local changes in the relative height may be considered as being associated mainly with the isobaric transport—into the locality—of an air column with a height different from that of the initial column. As a semi-Lagrangian method of predicting this advective change in the relative height, *the relative isohypses for t_0 are considered as conservative mean-isentropes of their mandatory layer embedded in its isobaric flow; and, as the fourth prognostic stage, they are displaced passively with the estimated future flow of the same layer, for the period, say, from t_0 to $t_0 + 24^h$.* The transport of the relative isohypses should include not only that part of the advection which is due to the gradient wind at either boundary surface of their mandatory layer, but also the advection by the gradient wind shear between these boundary surfaces. (The advection effect of the gradient wind shear is zero only if, along the vertical, the horizontal gradient of the virtual temperature maintains the same direction throughout the layer.)

The most accurate procedure, therefore, is to transport the relative isohypses according to the vector mean of the gradient flow at the boundary surfaces of their mandatory layer. This vector mean is determined approximately by the mean absolute hypsography of the bounding surfaces, obtained by the graphical addition of the two absolute hypsographies with only every second possible line being drawn. Since the mean absolute isohypses form a greater angle of intersection with the relative isohypses than do the absolute isohypses of either bounding surface, the mean absolute hypsography is more useful than either of the absolute hypsographies for estimating the transport of the relative isohypses due to gradient advection.

On a hitherto unused map (printed on vellum tracing paper), superimposed on the constant-pressure map for t_0, a rough draft of the prognostic relative hypsography for $t_0 + 24^h$ is, in this way, completed for each mandatory layer under the preliminary assumptions that advection is perfectly isobaric and that the field of horizontal motion at t_0 is constant during the prognostic period. But, after having first found the prognostic absolute hypsographies for $t_0 + 24^h$, we may then apply—as a correction to stage four—an average field of motion during the 24-hr prognostic period and arrive at better-advected positions of the prognostic relative isohypses. A method of successive approximations can thus be introduced for finding the varying field of horizontal motion during the 24-hr prognostic period, thereby improving the advective-adiabatic extrapolation of the relative hypsography.

We shall now proceed to consider the nonadvective and diabatic extrapolations of the relative hypsography. This prognostic operation takes into account the changes occurring in the quasi-conservative aspects of relative height and neglected in stage four. These

nonadvective changes in the relative height are due to (1) vertical motion, and (2) condensation or evaporation and radiative heating or cooling of the air.

The Local Change in the Relative Height Associated with Transisobaric Motion. Also associated with the local changes in the relative height is another atmospheric process known as *the transisobaric displacement of the air*. This is a somewhat less important process than the isobaric advection and involves the conservation of entropy. Although in reality both effects occur simultaneously with various degrees of relative intensity, for the sake of convenience in the operations of prognosis they are considered separately.

The greatest convective temperature changes are produced by the vertical advection of pressure, whereas only generally small convective changes in the temperature are caused by the relative local pressure change and by the horizontal advective pressure change. For example, a modest and frequently observed value of the vertical motion is 1 km per 12 hr, corresponding to an *individual* pressure change of about 23 mb per 3 hr for $p = 700$ mb and $T = 0$C. This rate of pressure change in turn corresponds to a convective isobaric change in temperature of 1.3C per 3 hr. In extreme instances the isobaric temperature change resulting from transisobaric displacement may be as large as 4C. On the other hand, at 3 km the pressure tendency seldom exceeds 3 mb per 3 hr, corresponding to a temperature change of only 0.18C per 3 hr at 700 mb and 0C. Extreme limits of cross-isobaric velocity component and pressure gradient are 5 m sec^{-1} (about 10 knots) and 6×10^{-5} cb m^{-1} (about 1 mb in 10 miles), respectively; such ageostrophic flow leads to a pressure change of 3 mb per 3 hr, corresponding to a temperature change of 0.18C per 3 hr at 700 mb and 0C.

The "convective" variation in the thickness of a mandatory layer is found by differentiating the statical equation with respect to time and ascribing the total convective changes in the temperature to the vertical advective pressure changes alone. In this differential equation, the pressure variable is eliminated by means of the statical equation, and the lapse rate of potential temperature is then introduced. Integrating this equation by using appropriate mean values for the mandatory layer, we obtain $-\Phi(\theta_2 - \theta_1)/\bar{\theta}$, where θ and θ_2 are values of the potential temperature at the lower and upper boundary surfaces, respectively, of the mandatory layer, and $\bar{\theta}$ is the mean potential temperature of the layer.

Since θ always increases upward in a deep layer, *unsaturated adiabatic descent* relative to a mandatory layer *increases* the partial relative height of the upper boundary surface. In a total layer of the lower troposphere, this motion is generally connected with subsidence which, after some time, makes the upward increase of θ even larger and thereby tends to produce an especially large rate of local increase in the total relative height. (In Fig. 8*a*, for example, it is seen that there is a pronounced center of nonadvective increase in the 500/1000-mb height over the south-central United States, where the northwesterly flow is descending from the Plateau Region.) However, the initial descent is soon hindered more and more by the increasing upward ascendent of θ, which, as a second-order contribution, dampens and somewhat reduces this otherwise large rate of local increase in the total relative height. Over inactive fronts the subsidence produces an increase in the relative height. (In Fig. 8*a*, for example, the cold front and occlusion in eastern United States are *inactive*, as evinced by the precipitation and altostratus-nimbostratus cloud system. It is seen that there is a ridge of nonadvective increase in the 500/1000-mb height to the rear of this front, the region traversed by this front during the past 24 hr.) At a cold front this effect partly reduces the generally larger relative-height decrease due to the advection of colder air, while at a warm front, it augments the relative-height increase due to the advection of warmer air. Consequently the spacing of the relative isohypses would be smaller for warm fronts than for inactive cold fronts.

On the other hand, *unsaturated adiabatic ascent* relative to a mandatory layer is accompanied by a *decrease* in its relative height. In a total mandatory layer of the lower troposphere, this motion is generally connected with vertical stretching, whereby the upward increase of θ through the mandatory layer is appreciably reduced, so that the associated rate of local decrease in the relative height will be smaller than the increase in the case of descent. (In Fig. 8*a*, for example, it is seen that there is a tongue of nonadvective decrease in the 500/1000-mb hypsography along the western seaboard of British Columbia and Washington where there is forced ascent of the westerly flow in crossing the Coast Mountains and the Cascade Range.) However, the initial ascent of air through a mandatory layer is abetted by the consequent reduction in the upward increase of θ between the boundary surfaces of the layer, so that, as a second-order contribution, this decreasing vertical ascendent of θ partly increases the otherwise small rate of local decrease in the relative height.

The *lifting* of *saturated* and *conditionally stable air* (or, more generally, of all saturated air *below its level of free convection*), such as in most cases of warm air-mass or warm-front precipitation, also produces a decrease in the relative height. (In Fig. 8*a*, for example, it is seen that there is a pattern of nonadvective decrease in the 500/1000-mb hypsography along the eastern seaboard of the United States and in southern Quebec, where a saturated and conditionally stable, warm air mass is being lifted in its poleward ascent over quasi-stationary warm-front surfaces. Over the corresponding (active) warm-front surface this relative-height decrease partly compensates for the larger relative-height increase due to the advection of warmer air. On the other hand, above the level of free convection, the relative-height increase due to the ascent of the conditionally unstable, *saturated* air over a cold-front surface would reduce the large relative-height decrease due to advection; above some warm-front surfaces the same effect would add to the height in-

crease due to advection. In most instances of saturated ascent, the initial lapse rate departs only slightly from the saturated indifferent one. In such cases even large upward vertical velocities have negligible effects upon the relative-height change.

The Local Changes in the Relative Height Associated with the Diabatic Processes. The local change in the relative height can be regarded as due in part also to a variation in the entropy of the air particles within the isobaric layer. Mainly because of evaporation, the latent heat being supplied by the air, and also by direct cooling, falling rain produces an appreciable diabatic decrease of the relative height in the lower layers. For example, the 850/1000-mb height is reduced about 20 gpm (geopotential meters) during the first four hours of a typical warm-front rainfall; after that (as the layer approaches a saturated state) no appreciable reduction takes place. Of course, in the adjacent lower layer of a temperature inversion the direct warming from falling precipitation partly reduces the larger relative-height decrease due to the evaporation. (In Fig. 8*a*, for example, it can be seen that the nonadvective decrease in the 500/1000-mb height corresponds roughly to the areas of precipitation and of the altostratus-nimbostratus cloud system.)

Diabatic changes in the relative height may also be brought about by the contact of the air mass of the total mandatory layers with warmer or colder ground or sea surface. Owing to increased turbulence by heating from below, a warm earth surface affects the relative height of a deeper layer than does a cold surface. The increase in the relative height is also particularly large for the 1000–700-mb layer in the cold air mass of polar outbreaks. (In Fig. 8*a*, for example, where the northwesterly flow of the cold air passes from the snow-covered Great Plains to the snow-free south-central states, a crescent-shaped center of nonadvective increase is found in the 500/1000-mb hypsography; the opposite considerations apply to the southerly flow of warm air as it passes from snow-free southeastern United States to snow-covered New England and Quebec.) Yet, even the ordinary turbulence in winds around 15 knots produces a decrease in the relative height for as much as the lowest hundred meters of a warm air mass. Even so, this decrease is still relatively large compared with the decrease by radiative cooling of that layer, the last atmospheric process of importance to be considered for relative-height change.

The relative-height decrease due to the net radiative cooling of the cloudless atmosphere is greater in summer, at low latitudes with warm humid air, and for low altitudes than in winter, at high latitudes with cold dry air, and for high levels. For example, the summertime radiative decrease in the 700/1000-mb height of a cloudless atmosphere is very roughly 5–10 gpm per day in a typical tropical air mass. Although according to estimates of Elsasser [24] the net rate of radiative cooling decreases steadily from 2 km to 5 km, later investigations by Penner [57] in an arctic region show that this rate is a maximum in the 600-mb to 500-mb layer.

The Prognosis of the Relative Hypsography by Nonadvective and Diabatic Extrapolation. These changes occurring in the quasi-conservative aspects of relative height are described by the nonadvective relative allohypsography, the allo-entropic centers of which are mainly due to (1) transisobaric lifting or sinking and to (2) condensation or evaporation (the latter especially in connection with precipitation) if the period is 24 hr. For periods shorter than 24 hr they will often be due mainly to (3) radiative heating or cooling of the air, which appreciably modifies this pattern, especially over high plateau areas such as western North America.

Further subdivision of the change in the relative height according to its component processes is not feasibly incorporated into the daily prognosis. In the practical method for calculating the transisobaric displacement of the air, the necessary assumption that the entire nonadvective change in relative height is adiabatic prevents us from separating the nonadvective relative allohypsography into, say, its "convective," "precipitative," and "radiative" relative allohypsographies. (Godske *et al* [33] have shown how the comparative intensities in these diabatic processes can be determined.) The results of Godske's method can then be introduced into stage five (see below).

The foregoing considerations of the nonadvective changes in the relative height, substantiated by synoptic experience, can be generalized into a *synoptic upper-air model*, which relates the nonadvective relative allohypsography to the corresponding relative hypsography. For example, the center R of nonadvective increase in Fig. 8*a* is just behind the equatorward portions of the troughs in the 500/1000-mb hypsography in Fig. 8*b*. The ridges in the relative hypsography (see Fig. 8*b*) and its centers of nonadvective decrease, F in Fig. 8*a*, are similarly related. *Generally speaking, these centers of nonadvective increase and decrease in the relative hypsography for the period* $t_0 - (t_0 - 24^h)$ *appear respectively at, or just behind, the equatorward portion of the trough and the poleward part of the crest in the relative hypsography for* t_0.

Important, then, in preparing the way for the prognosis is the analysis of the nonadvective relative allohypsography for the past 24-hr period, the zero isopleth of which, when traced onto the relative hypsography for t_0, can be applied as the starting place of prognosis by advective-adiabatic extrapolation. At the position of this isopleth, the relative isohypses have a past 24-hr displacement due just to their isobaric advective transport according to the past 24-hr flow of their layer. As the starting point for the fourth prognostic stage, already introduced, we may proceed, then, to advect adiabatically, during the 24-hr prognostic period, those portions of the relative isohypses for t_0 which are in the vicinity of this isopleth. The prognostic positions so obtained constitute a tentative, fixed framework for the prognosis by nonadvective and diabatic extrapolation.

Thus, as *stage five* in the order of prognostic operations, *we now complete the layer prognosis by supple-*

menting the advective-adiabatic extrapolation of the layer isentropes with physical considerations regarding the nonadvective and diabatic contributions to their future displacement. We choose an appropriately idealized prognostic pattern of the nonadvective relative allohypsography for $(t_0 + 24^h) - t_0$. Whereas the advective-adiabatic extrapolation in stage four is mainly mechanical in its application, this prognosis, on the other hand, brings into play the creativeness and intuitiveness of the forecaster and depends largely upon his experienced insight in perceiving and anticipating the nonadvective (and diabatic) processes in the atmosphere and in evaluating their effect upon relative-height change. The rational separation of the prognosis into its mechanical and nonmechanical operations (*viz.*, stages four and five), in addition to concentrating the forecaster's efforts in a routine way on these more or less intangible processes and the asso-

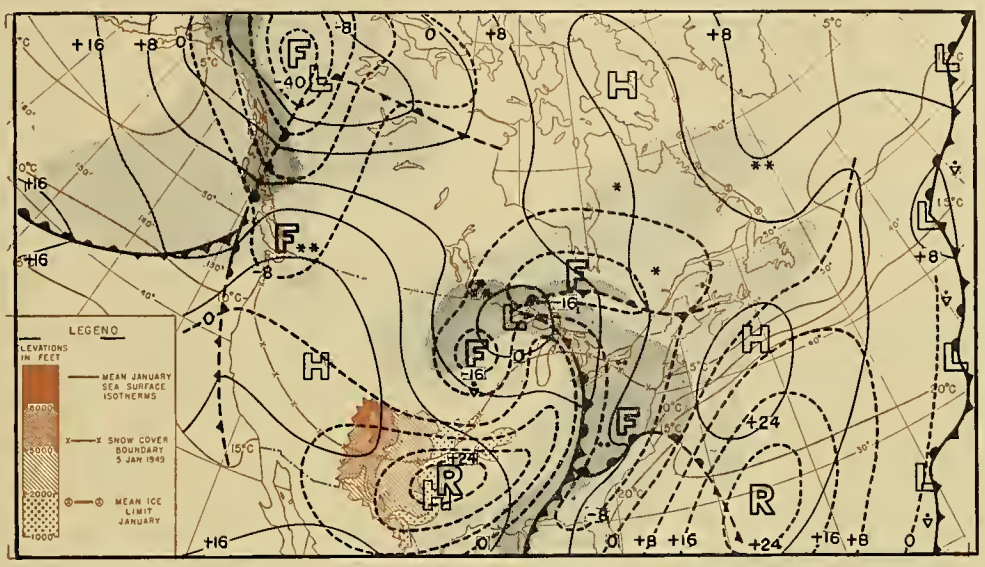


FIG. 8*a*

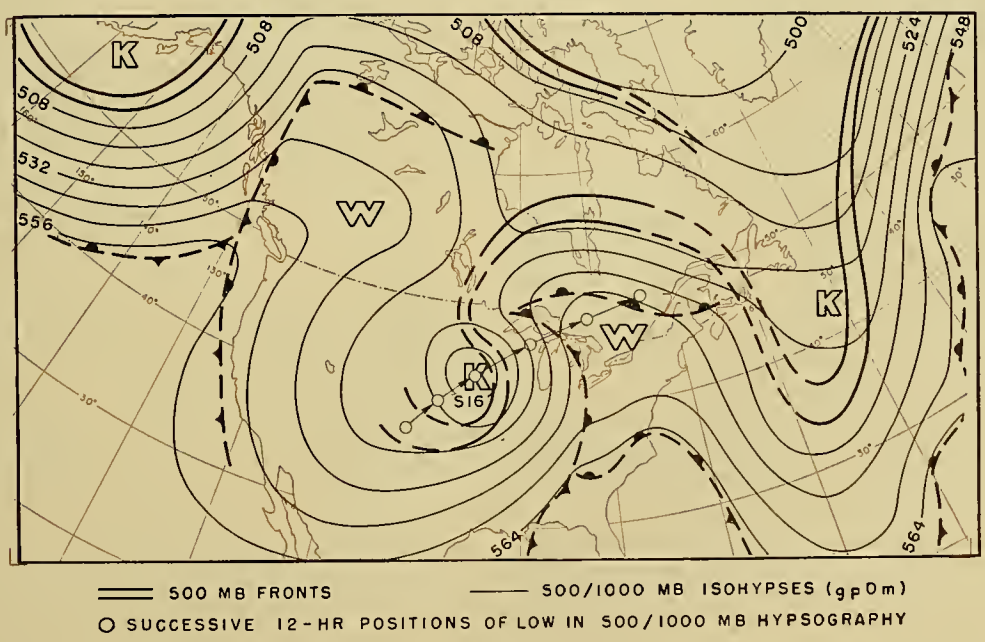


FIG. 8*b*

FIG. 8.—Maps for 1500 GMT, January 5, 1949, showing the relation of the nonadvective 500/1000-mb allohypsography for the past 24 hr to (*a*) the 1000-mb and 500/1000-mb hypsographies, (*b*) the areas of precipitation and altostratus-nimbostratus, (*c*) the extent of snow cover, (*d*) the sea surface temperature, and (*e*) the pertinent orography. This allohypsography has been based on streamline maps at 6-hr intervals, and instantaneous values of the *isobarically geostrophic advective change in the relative height* computed from the statical-kinematical tendency equation. (See [33], equation 11·61(1).)

ciated physical extrapolations, also provides for the acquisition of forecasting experience concerning such processes and extrapolations. Later verification will indicate the extent to which the nonadvective, diabatic processes have been correctly anticipated in the 24-hr prognosis: this verification is automatically accomplished at $t_0 + 24^h$, at which time is made the actual nonadvective relative allohypsography for the period from t_0 to $t_0 + 24^h$.

The final 24-hr prognosis of the relative hypsography is now obtained simply by adding graphically the prognosis of the advective relative hypsography for $t_0 + 24^h$, which has been just obtained by stage four, and the prognostic pattern of the nonadvective relative allohypsography for the period $(t_0 + 24^h) - t_0$.

With the completion of stages four and five, we can now proceed to utilize the temperature field for the prognosis of the surface map in a way not possible before the development of synoptic aerology, owing to the unrepresentativeness of the surface air-temperature observations. *As stage six, then, we exploit the prognostic relative hypsography for obtaining the prognosis of the surface map.* For example, as shown in Fig. 8*b*, frontal waves on the surface map should be associated with a warm tongue, or ridge, in the relative hypsography. An open frontal wave should be fitted to a broad and flat tongue (*e.g.*, Florida). A large spread between the prognostic relative isohypses describes a homogeneous air mass, which should be associated with the warm sector of an open wave (*e.g.*, Lake Ontario). The packing of the prognostic relative isohypses must occur in the frontal strip (*e.g.*, Gulf of Alaska, Mississippi Basin, Atlantic Ocean). An occluded frontal disturbance should be chosen for a narrow ridge with large amplitude in the prognostic relative hypsography (*e.g.*, the west coast).

Stage seven is the prognosis of the absolute hypsography, obtained by a simple routine procedure. The 24-hr prognosis of the relative hypsography is superimposed upon the prognostic surface map, upon which is now placed a hitherto unused transparent map overlay. By graphically adding these two prognoses, the prognostic absolute hypsography is drawn on the overlay.

Prognosis Beyond Thirty-Six Hours, Based on Extrapolation of Individual Upper Long Waves. The importance of the latitudinal variation of the Coriolis parameter for the explanation of the large-scale motions has been incorporated by Rossby [68] into a barotropic, approximately nondivergent, wave model. This physical model became a very simple instrument for the extended-period extrapolation of the baroclinic long waves in the upper westerlies. This extrapolation is based upon parameters determined at the "equivalent barotropic level." At this level it is then assumed that the extrapolation is governed primarily by the quasi-conservation of the vertical component of absolute vorticity. In spite of this somewhat unrealistic assumption, the simple prognostic rules developed from the application of this autobarotropic model have achieved partial success in predicting the propagation speed of the individual long waves in the upper westerlies and, to a lesser extent, their development. (This partial success initiated the use of such a model in making the mathematico-physical prognosis discussed in this volume by Charney on pp. 470–482 and by Fjørtoft on pp. 454–463.) In this subsection we very briefly consider the application of these rules, particularly those which have been rather well substantiated by synoptic experience.

Let us first consider the individual 500-mb maps, selected because they are, according to Charney [16, p. 147] approximately at the level of nondivergence. In the 500-mb westerlies one often finds *long waves* consisting of slowly moving cold troughs and warm ridges. Being often obscured on the surface map, they represent a family of *major waves* quite distinct from the rapidly moving *minor waves* or warm troughs and cold ridges which have their greatest amplitudes at the lowest level of the atmosphere and diminish with height.

The eastward velocity of propagation of the long waves is computed by means of Rossby's wave formula in the form

$$c = 2\Omega a \cos^3 \bar{\phi} \frac{L_s'^2 - L'^2}{(360)^2}, \tag{2}$$

where L' is the observed angular wave-length (expressed in degrees), and L'_s is the theoretical angular length which a wave should have in the atmospheric situation under consideration in order to be stationary. (The earth's parameters are represented by the usual notation, *viz.*, angular speed Ω, radius a, and mean latitude $\bar{\phi}$ of the waves.) Rossby's wave formula in the form

$$c = v_{x,5} - \frac{2\Omega a L'^2 \cos^3 \bar{\phi}}{(360)^2} \tag{3}$$

shows that, on the 500-mb map, a change in c can occur from a change in L' or from a change in ϕ (the latitudinal position of the maximum zonal wind) or $v_{x,5}$ (the geostrophic west-wind component), that is, in L'_s according to (2). In practice, the hemispheric chain of maximum westerlies is divided up into three parts, approximately 120° of longitude in length, such that waves within each part are roughly at the same latitude. Then $v_{x,5}$, which is measured at various latitudes across the maximum westerlies, and ϕ are space-averaged along each 120°-length of the maximum westerlies. To facilitate the application of equations (2) and (3), Byers has designed a nomogram [17] into which L', $v_{x,5}$, and $\bar{\phi}$ can be entered for finding c.

The short-range (*e.g.*, 48-hr) variations in these averaged values $v'_{x,5}$ and $\bar{\phi}$ of the maximum zonal windstream over its 120°-length are so small that they cause no significant change in L'_s, which is a function only of $v'_{x,5}$ and $\bar{\phi}$. A forecast of persistence for L'_s is therefore usually a good forecast. Hence, c varies primarily according to L'. However, it is generally difficult to predict the future changes in L', so that the forecaster must be content with using just the instantaneous t_0-value of c, over a third of the hemisphere, for the prognostic extrapolation of the long waves. The extent

to which he can predict the acceleration of propagation of the long waves depends on how well he can predict the changes in their wave length.

In an intense westerly jet stream (and also in the cases with waves of large amplitude), representative values of $v_{x,5}$ cannot be determined; the evaluation of c by Byers' nomogram should therefore not be attempted for such waves. The motion of the long waves in a narrow jet stream can be determined instead by averaging the speed profile from several cross sections, each eight degrees of latitude in width, and centered at the maximum speed of the jet stream. Next, α_i, the average direction, measured from the east-west axis (east = 0), of the isohypses at their inflection points, and ϕ_i, the corresponding average latitude are found directly from the 500-mb map. Then, the constant absolute vorticity path is evaluated; the wave length of this constant absolute vorticity path is equal to L'_s, the stationary-wave length. Therefore, for $L'/L'_s < 1$, the long wave under consideration will progress eastward during the prognostic period; and for $L'/L'_s = 1$, the wave will be quasi-stationary. When $L'/L'_s > 1$, L' will soon decrease; the decrease occurs first as a retrogression of the long waves and then as an increase in their wave number n.

The westward movement of the long waves is seldom a real phenomenon, but only an apparent, discontinuous one. It results from the fact that the major trough is transformed into a vanishing minor trough accelerating eastward simultaneously as a new major trough forms westward of the former position of the old major trough. After the retrogression, the final position of the major trough is at the distance of L'_s from the next major trough to the west. In the hemispheric chain of long waves, successive retrogressions of the major troughs generally occur downstream at short intervals of time.

At the end of this retrogression cycle, a new major trough often forms as an adjustment of L' in response to a change in L'_s, this adjustment appearing as an increase in n. It has been generally observed that an increase in a hemispheric chain of previously retrograding long waves leads to progression.

If the considerations outlined above could lead to a prognosis, three to six days in advance, of the major waves in the upper westerlies, it would establish the basic framework, so to speak, of the predicted future atmospheric state. But many assumptions still are involved in passing over from this more or less abstract prediction ("analysis-prognostic process") to the more detailed forecasting of the various weather constituents ("prognosis-forecast process"). These mental operations do, however, fall naturally into three straightforward parts: conjecturing qualitatively the most probable future (1) development of the minor waves in the upper westerlies, (2) sea-level paths of the cyclones, (3) anomaly patterns of precipitation and ground-surface temperature for the forecast period, or large intervals thereof, rather than for certain days, or particular times of the period.

Fultz [30] has suggested a certain life cycle for the minor waves in their conjunction with the major waves. These minor waves are associated with the cyclones. According to Klein [48] two of the prevailing sea-level paths of the cyclones converge in a region of confluence in the major wave. In this region, heavy precipitation is generally observed. In fact, from his investigation and from the experiences of others, Klein [48] has developed a schematic model of the precipitation anomaly pattern probably associated with a major wave. The ground-surface temperature anomaly pattern associated with the major wave has been studied by several investigators, whose results are reported by Bundgaard [33, Fig. 18·61·2].

THE PREDICTION OF VARIOUS WEATHER CONSTITUENTS

Introduction. The prediction of the interrelated weather constituents—temperature, cloud, precipitation—is naturally a composite process and should therefore be based mainly on the synoptic system of weather maps, actual and prognostic. But the map-made predictions of the weather cannot express the details of the local effects, since the map systems are intended to describe only in a coarse way the atmospheric state for the future—or even for the present. However, such effects can sometimes be applied locally for predicting an explicit weather constituent. In the first subsection we shall mention briefly and in a general way how objective techniques can be developed by methods of statistico-graphical integration. We then illustrate in a few words how temperature, cloudiness, precipitation, wind, etc., can be predicted by objective methods as well as by various physico-empirical considerations and synoptic rules of experience.

The Objective Method. The method of graphical integration provides a graphical solution of the form:

$$W = f(X_1, X_2, \cdots, X_i, \cdots, X_n), \qquad (4)$$

where W is the value of the weather constituent, (*e.g.*, rainfall, temperature) which is to be predicted and X_i are variables, measured at the initial time, which are believed to be related to W at some later period. The X_i are combined in successive pairs, each set of two variables being plotted on a scatter diagram with the X_i as coordinates and the values of the dependent variate, W, indicated beside each plotted point (see Fig. 9). Here, the small frequency charts represent the distribution of W in a small area on the graph. The data are analyzed by constructing $\overline{W}_i$-isograms of $a, b, \cdots, n$.

Each pair of X_i is analyzed in this manner, resulting in an equation of the form:

$$W = F(\overline{W}_1, \overline{W}_2, \cdots, \overline{W}_i, \cdots, \overline{W}_{n/2}), \qquad (5)$$

where each of the $\overline{W}_i$ is now a function of two X_i. The number of the new, $\overline{W}_i$-variables in (5) is about half the number of original X_i-variables in (4). This process is repeated, now using pairs of $\overline{W}_i$ as coordinate variables and, as before, analyzing each scatter diagram by considering the distribution of W itself. Obviously

this process may be continued until only one variable, W_f, is left on the right-hand side of (5).

The final variable W_f thus derived is a function of all of the original X_i. It has been obtained by a process which permits the distribution of the central tendency of the data to specify the form of the functional relationship, *thereby eliminating the major disadvantage of mathematical regression techniques, discriminant functions and the like, which require prior knowledge of, or assumptions regarding, the nature of the relationship between the independent variables and the dependent variate.* (Present mathematical regression techniques can treat only those relationships between variables of the

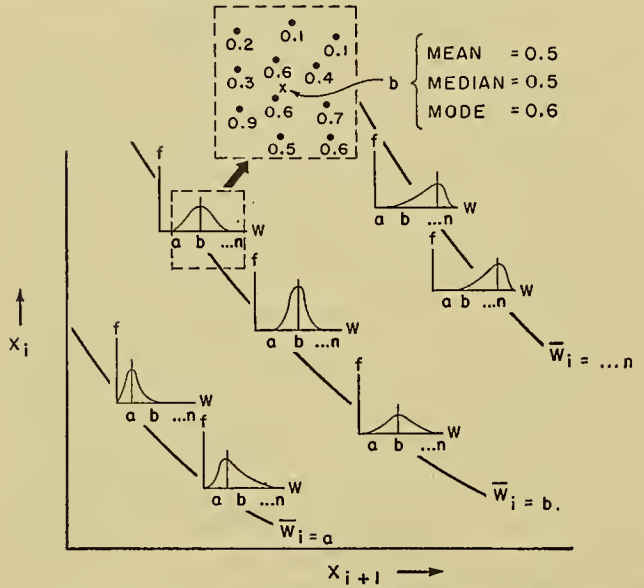


FIG. 9.—Schematic analysis of the scatter diagram. The dashed inset square at the top is an example of a small area of the diagram containing, for the purposes of this illustration, eleven values of a weather element W, say, rainfall amounts. For these eleven values of W, some measure $\overline{W}_i$ of their central tendency (*e.g.*, the mean), represented in this instance by the quantity $\overline{W}_i = b$, is entered at ×, the geometrical center of this small area. After such values of $\overline{W}_i$ have been entered over the entire chart, $\overline{W}_i$-isograms of these quantities, $a, b, \ldots n$, are drawn.

form $y = b_1x_1 + b_2x_2 + \cdots + b_nx_n$ or relations which can be transformed to that form. This permits the use of polynomial regression and sometimes exponential regression.) Practical application of the graphical integration process will be illustrated in the subsection on precipitation.

Temperature. The temperature ${}_gT$ at some point P on the ground can be predicted by starting from, say, the prognostic 700/1000-mb hypsography and by applying the indirect aerological methods as presented by Godske and others [33]. Relabelled (according to Table 17·50·1 in [33]), the prognostic 700/1000-mb isohypses give directly the ground temperature ${}_gT_s$ to which the predicted 700/1000-mb heights are reduced on the basis of the saturation-adiabatic lapse rate. Temperature corrections indicative of the *effective* 700/1000-mb lapse rate are added to ${}_gT_s$ by systematically considering at P the air-mass type (using Table 17·50·2 in [33]) and the pattern of the prognostic sea-level isobars (using Fig. 17·50·2 in [33]). Finally, additional corrections are applied to account for frontal surfaces, if any, predicted to be over P (using equation 17·50(1) in [33]) and anticipated ground inversions, if any, over P (using Fig. 17·50·1 in [33]). For predicting the rapid temperature changes occurring at frontal passages—one must take into account certain unrepresentative influences, such as the existence of cold-air films at the ground, pseudo fronts in advance of the real front,[9] local wind systems such as the land and sea breezes and foehn winds,[10] evaporation cooling in falling precipitation, and the predictable cloud distribution in space and time.

The local, diurnal, temperature minimum can be predicted quantitatively by means of formulas and nomograms which relate empirically the sunset observations of temperature, humidity, and wind at the place of prediction and at auxiliary stations some distance upwind to the heat loss during the ensuing night for synoptic situations in which no new air masses are to be expected. Such numerical methods have been developed by Kammermann [45], Dufour [21], Kessler and Kaempfert [46], Brunt [13], and Jacobs [44]. Geiger [31] has presented many detailed considerations for forecasting the night frosts, a problem still remaining after the minimum temperature has been predicted.

Speculating on the ways in which the total solar radiation received at the ground is dissipated, Gold [35], Neiburger [55], and Bundgaard [33] have presented quantitative methods for predicting locally the daily temperature maximum. These methods are applicable if, for at least the approximate half-day period from one hour after sunrise to the time of temperature maximum (normally 1300–1500 LMT), clear skies and no air-mass replacement are forecast locally. Under these conditions, then, the locality remains well within a homogeneous, extensive, and slowly moving air mass during this period, so that local advective changes are negligible.

General Hydrometeors. Extrapolated frontal positions at once indicate the future position of the areas of nimbostratus rain and altostratus cloud. The future location of drizzle and low stratiform cloud will extend through such parts of the extrapolated air masses where there will be cooling from below, showers and low cumuliform cloud where warming from below is anticipated. Using the predicted map changes of (1) surface temperature, (2) vertical motion, (3) upper-air humidity, and (4) upper-air stability, the regional forecaster can issue general predictions about nonfrontal hydrometeors: stratiform cloud—with or without drizzle, cumuliform cloud—with or without showers, thunderstorm, and hail. The change (3) enables us to decide whether or not the processes leading to hydrometeors, once started by a thermal influence (1) or a kinematical impulse (2), will culminate in the condensation phe-

9. Consult "The Instability Line" by J. R. Fulks, pp. 647–652 in this Compendium.

10. Consult "Local Winds" by F. Defant, pp. 655–672 in this Compendium.

nomena, while (4) enables us to predict the type of precipitation. The future location of precipitation due to upper-tropospheric instability will be found in the forward half of the extrapolated islands of cold air aloft (see Figs. 7 and 8). Upper-tropospheric instability usually occurs above a cold air mass within a nonfrontal trough in the rear of a cyclone on the surface map. At sea, where the cold air mass already contains convective clouds, the presence of upper tropospheric instability will give the clouds there maximum vertical growth. The same development may be found over land if the initial humidity supply and the heating from below are sufficient. When the low-level instability is missing, the upper-tropospheric instability can be seen to produce castellatus from layers or stripes of medium cloud. Showers and thunderstorms from such cloud systems are rather independent of the time of day or night and can be followed from map to map. On the other hand, in the rear half of the upper island of cold air, both the warm advection and the descent of the air contribute to the suppression of the unstable character of whatever cloudiness and precipitation there may be in that area. The forecaster should also take into account the orographic strengthening of ascending cloud masses due to the nature of the terrain and other more indirect orographic effects, such as foehn effects.

There is a tendency towards formation of cloud and precipitation in the entrance regions, whereas there is a tendency towards dissolution of hydrometeors in the exit regions. There is a tendency towards formation of hydrometeors in the areas where the isohypses are cyclonically curved and a tendency towards dissolution of them in regions of anticyclonic curvature. In poleward-moving air extending through a deep layer over the earth, cloudiness and precipitation may be expected in areas where the prognostic isohypses of the lower mandatory surfaces are either straight or cyclonically curved. This effect will be especially marked within tropical air, owing to its great relative and specific humidity already before its lifting and stretching. In an equatorward flow of similar vertical extent, cloudiness and precipitation are likely to occur only where the prognostic isohypses have pronounced cyclonic curvature.

The centers of nonadvective decrease of relative height correspond to regions where the air of the corresponding layer either is ascending or is losing entropy by radiation or precipitation. In both instances cloudiness and precipitation must be expected to be associated with these centers (see Fig. 8*a*). A detailed prediction of the air-mass hydrometeors, of great practical importance, must, however, be left to the area and local forecasters.

Warm Air-Mass Hydrometeors. Detailed predictions of warm air-mass hydrometeors—fog, stratus, and drizzle—cannot be based completely on the prognostic map system, since local effects are also important. Warm air-mass fogs, excluding the steam and frontal fogs, are abetted by the cooling of the air particles near the ground sufficiently below the dew point (Petterssen [59]) and are hindered by turbulence-producing wind. Upslope fogs may be predicted locally by applying a prognosticated lifting condensation level to the large-scale orographic ascent of the air which occurs with certain wind directions. For predicting local sea and coastal fogs, a problem more difficult than the regional problem of predicting tropical air fogs, the local forecaster must be familiar with the local changes in the sea temperature and wind as well as the local aerological observations, which are necessary to clarify the difference between the situation in which fog occurs and that in which the fog is lifted by turbulence so that stratus cloud appears. Valid for clear skies and slight winds (below 5.5 mph), Taylor's empirical diagram [77] for forecasting radiation fog at Kew Observatory has as its abscissa the evening temperature and as the ordinate the dew-point deficit. For the local prediction of advection-radiation fogs, George [32] constructed similar but more complete empirical diagrams which also take into consideration clouds, wind speed, wind direction, and air trajectories. Predicting the morning-time dissipation of fog and stratus by synoptic-local methods should take into consideration its linear relation to cloud thickness [83], in addition to other fog-dissolving influences such as wind speed and snow cover.

Cold Air-Mass Hydrometeors. Cumuliform clouds, cold-front and orographic showers, and showers superimposed upon a warm-front rain can all be predicted in a general way by the regional forecaster on the basis of his prognostic upper-air maps showing the conditional and potential lability of the air, its humidity, and the position of its ice nuclei. In fact, such shower areas also form a typical part of the cyclone models for the surface map. In nighttime and during winter, continental air acquires cold air-mass properties over sea, while in the afternoon and during summer, at the time of maximum ground heating, all air masses and especially maritime air acquire cold air-mass properties over land. For the diagnosis and prognosis of the early cumulus stage, we may mention that Beers [8] has applied a practical method of evaluating for each individual layer its circulational acceleration, a value which may also be interpreted as the geopotential difference of the ascending column from the descending one.

For very short-range forecasting (6-12 hr) of thunderstorms, Brooks [12] has found a correlation between thunderstorm occurrence and the direction of the lower tropospheric winds (ground to 1½ km), but no evident relationship between the speed of these winds and thunderstorm occurrence. However, strong winds above 3 km tend to inhibit thunderstorm activity. Brancato [11] states that large thunderstorms move with the 11,000-ft winds, while (small) nonfrontal storms seem to be steered by the 5000-ft winds. Humphreys [43] formulates the relation differently; he states that the thunderstorms move with the strata containing most of the cumulonimbi. But these conclusions are rendered indecisive by the report of Byers [15] that the thunderstorms move to the right of the winds aloft.

The formation of hail is favored by high humidity and strong upward motions which slow the fall of the hailstone and increase the equivalent cloud thickness

traversed by it. Hail starts rather suddenly and usually lasts from a few minutes to less than half an hour; it is mostly followed by rain, these two phases being partly superimposed. It falls along narrow streaks a few kilometers wide. The foregoing considerations show

for predicting the amount of winter rainfall occurring in Los Angeles during the 30-hr period beginning at $t_0 + 6^h$.

Figure 10 shows the results of the complete graphical integration process using the six variables listed in

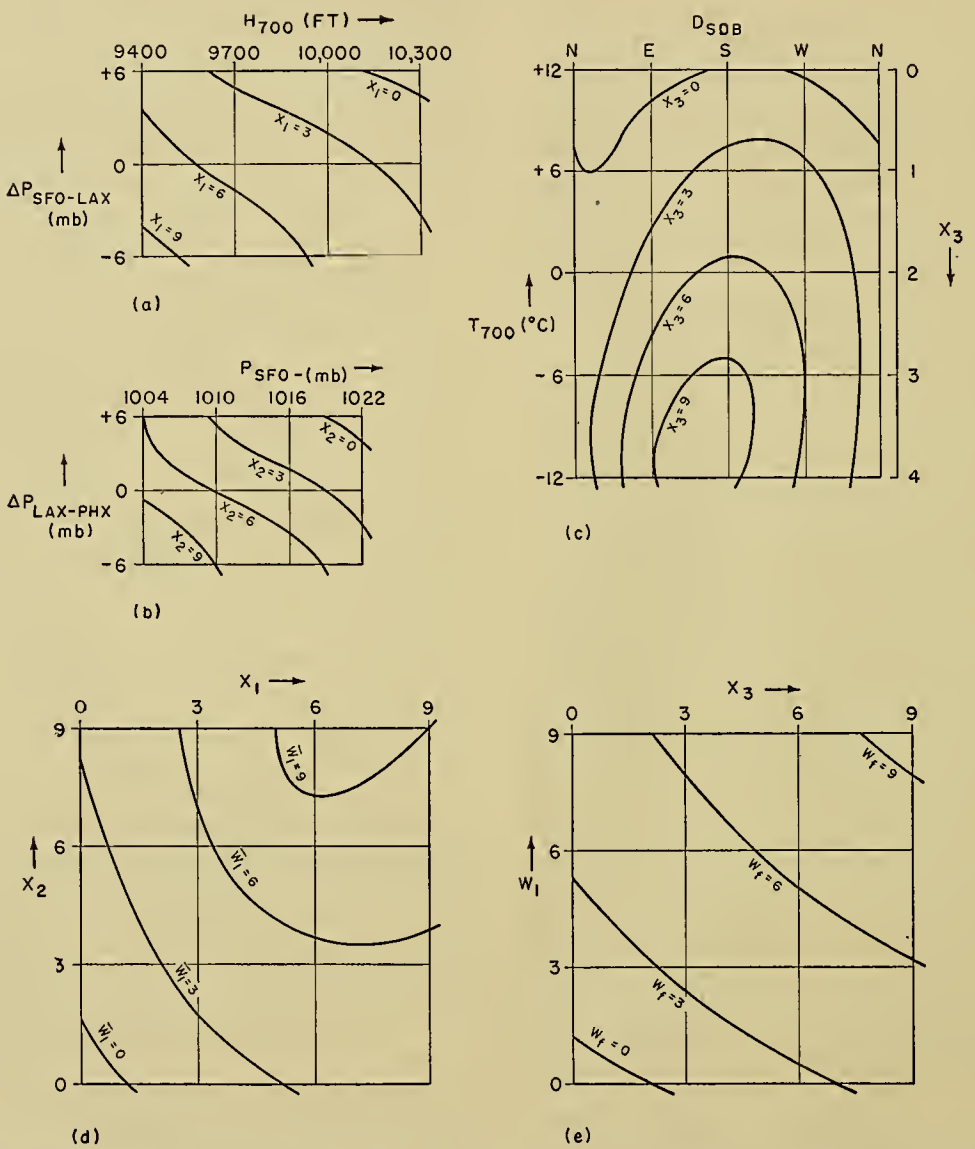


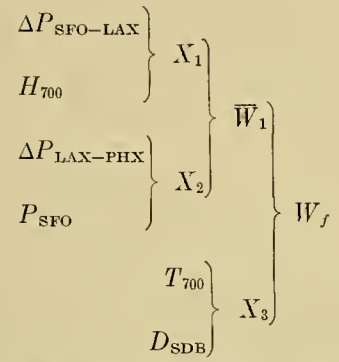
FIG. 10.—A nomogram for making an objective forecast of rainfall at Los Angeles. The rectangular coordinates of the scatter diagrams represent t_0-values of the six predictors listed in Table II. The solid curves are isograms of rainfall amount, adjusted to a scale of 0 to 10, during the 30-hr period following $t_0 + 6^h$. In part (*c*), the variable D_{SDB} loses its sensitivity as an indicator of the true wind field at low wind speeds, and thus was not used for speeds less than seven knots. These cases were plotted against T_{700} along the vertical axis to the right of the scatter diagram and analyzed separately to determine X_3. The solid curves in part (*e*) define a variable W_f, which is used in computing rainfall probabilities as given in Table III.

that the individual shower prediction will, at best, be possible only some few hours in advance for very small districts and with the assistance of *rareps* from specially trained observers, who reside within these districts in places with free horizon.

Amount of Precipitation. The application of objective techniques introduced in the first subsection can now be illustrated by Thompson's objective method [79]

Table II. The analysis of each part of this figure was carried out by first constructing isograms of rainfall amounts on parts *a–c* of Fig. 10 and then adjusting the isogram values to a scale 0 to 10. This latter device provides a uniform system of coordinates on all succeeding scatter diagrams *d–e* of Fig. 10, an operation which somewhat simplifies their use by the forecaster. The order in which the variables were combined is

illustrated schematically below:



A complete discussion of the physical reasoning involved in the selection of these variables has been given by Thompson.

TABLE II. PREDICTORS USED FOR OBJECTIVELY PREDICTING THE RAINFALL AMOUNT AT LOS ANGELES

Predictor	Abbreviation symbol for predictor	Part of Fig. 10 in which predictor is used
1. The sea-level pressure difference, San Francisco (SFO) minus Los Angeles (LAX)	$\Delta P_{SFO-LAX}$	*a*
2. The 700-mb height at Oakland	H_{700}	*a*
3. The sea-level pressure difference, Los Angeles minus Phoenix (PHX)	$\Delta P_{LAX-PHX}$	*b*
4. The sea-level pressure at San Francisco	P_{SFO}	*b*
5. The 700-mb temperature at Santa Maria	T_{700}	*c*
6. The surface wind direction at Sandberg (SDB)	D_{SDB}	*c*

The parameter W_f from Fig. 10*e* is used as the forecast criterion. Corresponding to each integral value of this criterion, the percentage frequency of rainfall is listed in Table III. This frequency is tabulated for

TABLE III. RELATION BETWEEN W_f AND THE PROBABILITY THAT RAIN WILL OCCUR IN THE INDICATED CATEGORIES
(*Abbreviated from Thompson* [79].)

W_f	No rain	0.01–4.00 mm	4.01–12.50 mm	12.51–38.00 mm	38.01 mm or more
0	100	—	—		
3	88	9	3	—	
4.1	*69*	*23*	*8*	—	—
6	28	36	28	6	2
6.6	*24*	*26*	*36*	*11*	*3*
9	—	—	25	55	20

each of five rainfall intervals, namely, no rain, 0.01–4.00 mm, 4.01–12.50 mm, etc. For example, suppose that by applying Fig. 10 the value of the forecast criterion has been found to be 4.1. Then, according to Table III, there are eight chances out of every hundred that the occurrence of rainfall will exceed 4 mm. If on the other hand $W_f = 6.6$, then the probability is $36 + 11 + 3 = 50$ per cent, or the chances are even that the rainfall will be greater than 4 mm. Finally we mention that objective methods for predicting precipitation also have been applied at Atlanta by Beebe [7] and at Washington, D. C. by Rapp [65].

Wind. The prognostic upper-air maps give a general picture of the future distribution of geostrophic and gradient winds aloft. The extent to which the wind deviates vectorially from the geostrophic wind has been investigated by Houghton and Austin [42], Machta [51], Bannon [6], Durst and Gilbert [22], Emmons [25], and Godson [34]. They found that, on the whole, the magnitude of this deviation is from about one-fourth to one-third of the wind speed. With increasing speed, this ratio first decreases and then increases. The minimum value of this ratio occurs with winds around sixty knots. At 700 mb the geostrophic speed is, on the whole, as accurate as the gradient speed. However the gradient wind-speed becomes more accurate than the geostrophic wind-speed whenever

$$KV_G^2 < 10^{-3} \text{ m sec}^{-2},$$

for example, at troughs with strong pressure gradients.

The geostrophic angular deviation of fast winds (for example, at 300 mb) is negligible. In this case, the absolute isohypses (or the absolute prohypses of the prognostic maps) may be considered as streamlines. For slow winds (*e.g.*, on the 700-mb map), there may be, under certain circumstances, appreciable differences in the direction of the absolute isohypses and streamlines. We shall now (1) indicate where on the constant-pressure map the forecaster should expect large, geostrophic angular deviations of the wind to occur, as well as (2) make some remarks on the nature of these deviations. In making an accurate upper-winds forecast from the prognostic constant-pressure maps, he must take (1) and (2) into account.

The normal equation for horizontal frictionless motion may be written as $V = a \cos \beta \, V_G$. (In this expression, $V = |\mathbf{V}|$ is the wind speed; $V_G = |\mathbf{V}_G|$ is the geostrophic wind speed; $a = f/(f + d\psi/dt)$, ψ being the wind direction, f the Coriolis parameter; and the angle β of geostrophic deviation is measured counterclockwise from $\mathbf{V}_G$ to $\mathbf{V}$.) For sufficiently small angles of geostrophic deviation the normal equation may be written approximately as $V = aV_G$. Assuming a constant, $\dot{V} = a\dot{V}_G$. But $\dot{V}_G \approx \partial V_G/\partial t + aV_G(\partial V_G/\partial s_G + \beta \partial V_G/\partial n_G)$. (In this differentiation $\sin\beta$ has been replaced by β, and the value of $\cos\beta$ has been taken as unity.) For this expression, s_G is measured in the direction of the geostrophic wind, and n_G normal to it. Thus, $\dot{V} = a\partial V_G/\partial t + a^2 V_G(\partial V_G/\partial s_G + \beta \partial V_G/\partial n_G)$. Introducing the tangential equation of horizontal motion $\dot{V} = f \sin \beta V_G \approx f\beta V_G$ for the left member and solving for β, we have finally

$$\beta = \frac{a\left(\frac{1}{V_G}\frac{\partial V_G}{\partial t} + a\frac{\partial V_G}{\partial s_G}\right)}{f - a^2\left(\frac{\partial V_G}{\partial n_G}\right)}.$$

This equation shows, first of all, that convergence or divergence in space and time of the absolute isohypses tends to produce positive or negative angles of geostrophic deviation, respectively. In other words, in regions of confluence (difluence), the air flows across the absolute isohypses toward lower (higher) height. Moreover, this equation shows that the angle of geostrophic deviation will be amplified or suppressed according as the horizontal geostrophic shear is anticyclonic or cyclonic, respectively. It can also be seen that, in the case of anticyclonic geostrophic shear at a constant-pressure surface, anticyclonic curvature of the path tends to magnify the horizontal shear effect and is therefore also conducive to relatively large angles of geostrophic deviation. (For anticyclonic curvature, the value of a is greater than one.) Relatively large deflection of the streamlines across the absolute isohypses (or prohypses in the case of prognostic maps) can therefore be expected in regions of anticyclonic horizontal geostrophic shear and where the curvature of the path is anticyclonic. Such conditions may be found in the right half of the confluent and difluent regions. On the other hand, the conditions for a relatively small angle of geostrophic deviation are (1) that the horizontal shear of the geostrophic wind be cyclonic or only slightly anticyclonic and (2) that, if the horizontal shear is anticyclonic, the curvature of the path be cyclonic or only slightly anticyclonic. (For cyclonic curvature, the value of a is less than one.) These conditions are found in the left half of the confluent and difluent regions. Analytic examples showing large geostrophic angular deviations of wind have been found by Gustafson [37, Figs. 1–3] (*cf.* also Petterssen [61, Fig. 2].)

The prediction of wind for a mandatory surface is thus comparatively simple once the pressure field is forecast; the prediction for intermediary pressures is achieved by interpolation between two mandatory surfaces. To forecast locally the wind *speed* aloft, say 12 hr in advance, the regional forecaster can, often with good results, extrapolate the quasi-conservative configuration of the isotachs. The crescent-shaped centers of maximum speed usually have their long axis of symmetry along the absolute isohypses. They move generally in the direction of the geostrophic wind. Their future direction of propagation is thus given by the prognostic absolute isohypses and their displacement by purely formal extrapolation. These efforts have revealed certain deficiencies, however, in the current techniques of observing the upper winds, especially at high levels with very large speeds. In the United States, these deficiencies are expected to be remedied with the introduction in the near future of an improved radar wind-finding instrument known as the AN/CPS-10. Finally, we mention that if no system of upper-air maps is available for predicting upper winds, we may apply indirect aerology, for instance, the method employed by Kibel [47], already introduced by Exner [27].

Closely connected with the prediction of wind is the problem of forecasting atmospheric turbulence. Since the amount of horizontal (vertical) shear of the actual wind is far higher (the same) for turbulent layers than (as) for normal ones, the forecaster should be on the lookout for pronounced clear-air turbulence in the regions adjacent to, but not within, the jet stream. The maximum speed of the wind gusts at the ground associated with thunderstorms can be forecast by the formula: C (in knots) $= 11\sqrt{\Delta\theta} + 15$, where $\Delta\theta$ is the difference (centigrade degrees) between the lowest wet-bulb potential temperature of the sounding and the wet-bulb potential temperature of the rising air above the condensation level [11].

Excessively strong winds occur in a small-scale model, the tornado (horizontal extent below one km), whose prediction is connected with the peculiar difficulty that it completely escapes detection on the synoptic map. For a long time to come, it will be generally impossible to forecast the time and position of occurrence of individual tornadoes. The forecaster must be content to predict only the general conditions favorable to tornado formation and then to forecast in a rather vague way that a number of tornadoes are likely to occur at scattered points in a region (say, 200 km square) and during a certain time period (of about one to two hours). These general conditions, shown in

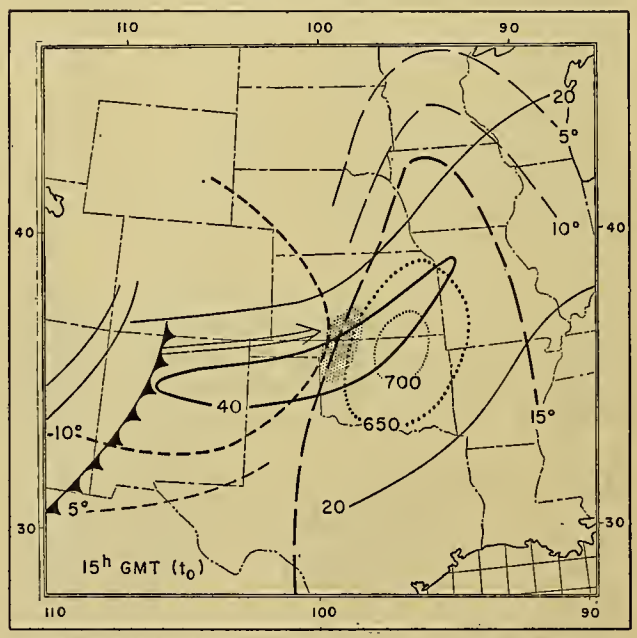


FIG. 11.—Characteristic features of the weather maps (at t_0) for the subsequent development of tornadoes in the shaded region at about $t_0 + 12^h$. The isohumes connect points of equal dew point; isotachs, points of equal wind speed.

Fig. 11 as they appear on the weather maps, are[11] [28, 74]:

1. A conditionally unstable, deep tongue of unusually dry air (namely, a 700-mb dew point of −10C

11. Consult "Tornadoes and Related Phenomena" by E. M. Brooks, pp. 673–680 in this Compendium.

or less) is superimposed on a horizontally narrow wedge of abnormally humid air at the ground (a dew point of +15C or more at the surface of the earth) which is possibly both conditionally and potentially unstable (hence, for the humid layer, a level of free convection at 650 mb, or lower).

2. Heading obliquely to the axis of the humid wedge, there must be a very narrow band of strong westerlies aloft (>35 knots at 600 mb).

3. The accurate analysis of forerunning upper cold fronts is also an important prerequisite for tornado forecasting. Forced lifting of the lower wedge of humid air must take place—usually as a result of the invading cold front to which the forerunning upper cold front or cold quasi-front is associated.

only in that part of a deep, subfreezing water cloud where supersaturation with respect to ice also occurred.

As shown by Godske and others [33], supersaturation with respect to ice exists whenever $T > -8D$, where the temperature T and the dew-point deficit D are expressed in degrees centigrade. On an aerological diagram, the icing layer is determined by the subfreezing cloud area enclosed by the temperature ascent-curve on the left and the $-8D$ curve[12] on the right, as shown in Fig. 12. Moreover, the intensity of the icing is indicated by the size of this enclosed area. The cloud type and precipitation, both observed at the surface of the earth, will show the degree of colloidal-thermodynamic instability and hence the type of icing—rime or glaze. In particular, whenever the ascent curves of tempera-

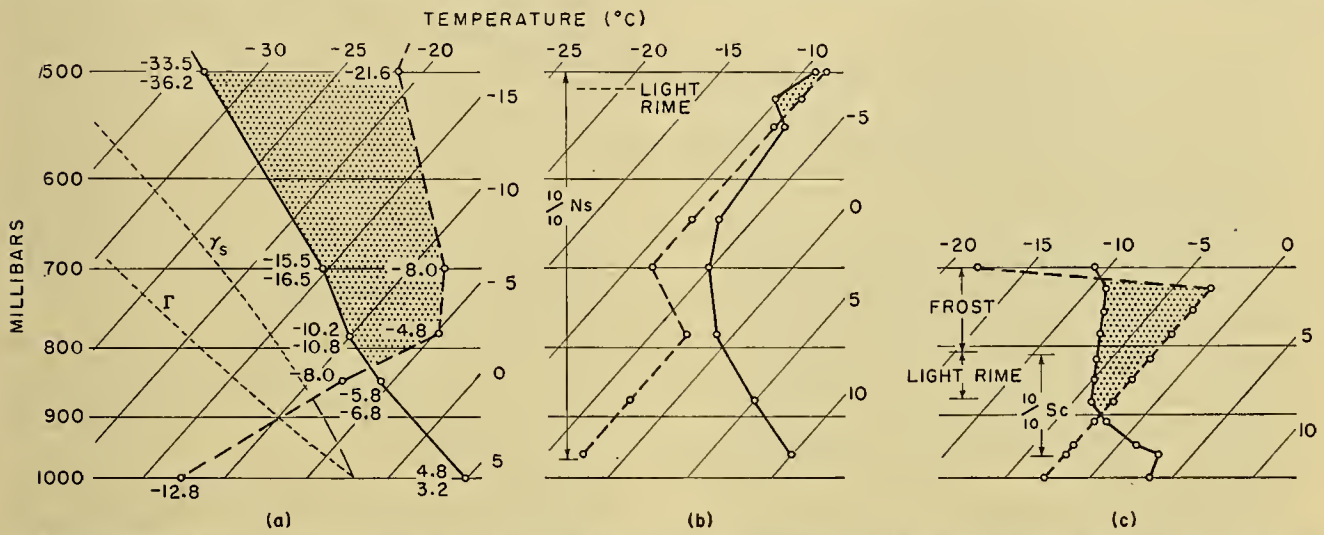


FIG. 12.—Three sets of ascent curves for temperature (solid) and—8D (dashed), plotted on a skew T, log p-diagram. (*a*) Radiosonde observations for 0300 GMT, April 25, 1949, Yakutat, Alaska. Image points labelled with values of temperature in degrees centigrade (upper left) and dew point; the —8D curve labelled with corresponding values of —8D. (*b*) Aircraft ascent for 1400 GMT, February 20, 1949, at 41.2°N, 56.2°W, with the observed icing and clouds plotted to the left. (*c*) Aircraft ascent for 1700 GMT, February 7, 1949, at 40.1°N, 46.6°W, with the observed icing and clouds plotted to the left.

An empirical rule for forecasting the movement of an individual tornado has been given by both Showalter [74] and Willett [82], namely that a tornado moves roughly with the direction and speed of the mT stratum just beneath the inversion. Others state simply that the tornado travels roughly parallel to the path of the accompanying cyclone.

The Prediction of Aircraft Icing. For icing on aircraft in flight, the subfreezing cloud containing water in the liquid phase must be supersaturated with respect to ice. Although supercooled cloud droplets occur in air subsaturated with respect to (a plane surface of) water, sublimation starts only after the moist air attains saturation equilibrium with respect to ice. In a layer subsaturated with respect to ice, the physically unstable, supercooled fog droplets, upon striking the leading edges of the aircraft, are induced mechanically to congeal as light rime ice which almost instantly evaporates into the air streaming very rapidly over the rime-forming surface of the aircraft. For example, the two aircraft soundings in Figs. 12*b* and 12*c* report icing

ture and dew point coincide, the $-8D$ curve must necessarily coincide with the 0C isotherm on the diagram. In the case where the ascent curves of temperature and dew point coincide in a subfreezing layer, the air would be saturated with respect to water, supersaturated with respect to ice. Only light rime icing would occur in the altostratus-nimbostratus cloud system of this layer; but moderate rime icing would be encountered in cumulonimbus virga which is located in this layer. Severe clear icing—for the same sounding as above—would occur in the stratocumulus virga, cumulus virga—particularly when transforming to cumulonimbus—and stratus. Whenever, on the other hand, the ascent curves of temperature and dewpoint do not coincide but the temperature curve lies to the left of the $-8D$ curve in a subfreezing layer, this

12. For most practical purposes, the dew-point ascent curve, which is frequently plotted on the aerological diagrams, could be replaced by the $-8D$ curve, since the latter, together with the 0C isotherm, may be used in place of the ascent curves of temperature and dew point for showing the vertical variation of the dew-point deficit

layer is supersaturated with respect to ice and probably subsaturated with respect to the cloud droplets. Only light rime icing will be experienced by aircraft entering this layer, if the clouds reported in them are altostratus, altocumulus cumulogenitus, or altocumulus virga; only light hoar frost will be sublimed on the aircraft, if the clouds in such a layer are reported as cirrus, cirrocumulus, cirrostratus, or cirrus nothus. In cloudless regions of this layer, no supercooled droplets will be present, but hoarfrost will also form on the aircraft through direct sublimation of the water vapor. Finally, in the case where the temperature curve lies to the right of the $-8D$ curve in a subfreezing layer, this layer is subsaturated with respect to both ice and plane water surface, so that no icing should be forecast here.

REFERENCES

1. Air Weather Service, AAF Headquarters, *Studies on Local Forecasting*, Ser. 600, 1945.
2. Angervo, J. M., "Einige Formeln für die numerische Vorausbestimmung der Lage und Tiefe der Hoch- und Tiefdruckzentra." *Ann. Acad. Sci. fenn.*, (A) Tome 28, No. 10, 45 pp. (1928).
3. —— "Über die Vorausberechnung der Wetterlage für mehrere Tage." *Beitr. Geophys.*, 27:258–311 (1930).
4. —— "Einige aus der Luftdruckverteilung herleitbare Gesetzmässigkeiten bei der Bewegung der Hoch- und Tiefdruckzentren." *Meteor. Z.*, 47:354–364 (1930).
5. Austin, J. M., "An Empirical Study of Certain Rules for Forecasting the Movement and Intensity of Cyclones." *J. Meteor.*, 4:16–20 (1947).
6. Bannon, J., "The Angular Deviation of the Wind from the Isobars, at Liverpool." *Quart. J. R. meteor. Soc.*, 75:131–146 (1949).
7. Beebe, R., "Forecasting Winter Precipitation for Atlanta, Georgia." *Mon. Wea. Rev. Wash.*, 78:59-68 (1950).
8. Beers, N. R., *Forecasting Thunderstorms by the Slice Method.* Wea. Inform. Div., AAF Headquarters, Rep. No. 798, 43 pp., 1944.
9. Bice, E. G., and Stevens, G. T., "Correlations of Isallobaric Patterns in the High Atmosphere with Those at the Surface." *Dept. Meteor. Univ. Chicago, Misc. Rep.* No. 12, 44 pp. (1944).
10. Bleeker, W., "The Verification of Weather Forecasts." *Meded. ned. meteor. Inst.*, (B) Deel 1, Nr. 2 (1946). (See Part B, pp. 5–7)
11. Brancato, G. N., *The Meteorological Behavior and Characteristics of Thunderstorms.* U. S. Dept. Comm., Wea. Bur. (Hydrometeor. Sect.) Misc. Rep., 14 pp., 1942.
12. Brooks, H. B., "A Summary of Some Radar Thunderstorm Observations." *Bull. Amer. meteor. Soc.*, 27:557–563 (1946). (See p. 558 and Figs. 3–4, p. 559)
13. Brunt, D., *Physical and Dynamical Meteorology.* Cambridge, University Press, 1934. (See pp. 123–132)
14. Bundgaard, R. C., "A Note on the Individual Changes in the Thickness of Isobaric Layers." *Cent. Proc. R. meteor. Soc.*, pp. 113-117 (1950).
15. Byers, H. R., "Nonfrontal Thunderstorms." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 3, 26 pp. (1942). (See p. 25)
16. Charney, J. G., "The Dynamics of Long Waves in a Baroclinic Westerly Current." *J. Meteor.*, 4:135–162 (1947).
17. Cressman, G. P., "On the Forecasting of Long Waves in the Upper Westerlies." *J. Meteor.*, 5:44–57 (1948).
18. Chromow, S. P., und Konček, N., (Translated to German by G. Swoboda), *Einführung in die synoptische Wetteranalyse.* Wien, J. Springer, 1940. (See pp. 436–444)
19. Dedebant, G., "Le champ du déplacement instantané des isobares." *C. R. Acad. Sci., Paris*, 185:359–361 (1927).
20. —— "Cinématique des centres isobariques." *C. R. Acad. Sci., Paris*, 194:1673–1675 (1932).
21. Dufour, L., "Notes sur le problème de la gelée nocturne." *Ciel et Terre*, 54:33–56 (1938); "La détermination du minimum nocturne." *Ann. de Gembloux*, pp. 415–430 (1938). (Extrait) (Also in *Inst. R. météor. Belg., Mém.*, Tome 9, 24 pp. (1938).)
22. Durst, C. S., and Gilbert, G. H., "Constant-Height Balloons—Calculation of Geostrophic Departures." *Quart. J. R. meteor. Soc.*, 76:75–86 (1950).
23. Elliott, R. D., "Study of Blocking-Action Weather Process." *Trans. Amer. geophys. Un.*, 29:773–776 (1948).
24. Elsasser, W. M., "Radiative Cooling in the Lower Atmosphere." *Mon. Wea. Rev. Wash.*, 68:185–188 (1940).
25. Emmons, G., *High Altitude Balloon Trajectory Study.* Progress Rep. 121.04. Res. Div., Coll. of Engng., N. Y. Univ., March 5–6, 1950.
26. Ertel, H., "Singuläre Advektion und Zyklonenbewegung." *Meteor. Z.*, 56:401–407 (1939).
27. Exner, F. M., "Grundzüge einer Theorie der synoptischen Luftdruckveränderungen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 115:1171–1246 (1906).
28. Fawbush, E. J., and others, "An Empirical Method of Forecasting Tornado Development." *Bull. Amer. meteor. Soc.*, 32:1–9 (1951). (See pp. 5 f.)
29. Ficker, H. v., "Zur Frage der 'Steuerung' in der Atmosphäre." *Metcor. Z.*, 55:8–12 (1938).
30. Fultz, D., "Upper-Air Trajectories and Weather Forecasting." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 19, 123 pp. (1945).
31. Geiger, R., *Das Klima der bodennahen Luftschicht, Lehrbuch der Mikroklimatologie*, 2. völlig umbearbeitete Aufl. Berlin, F. Vieweg & Sohn, 1942. (*Die Wissenschaft*, Bd. 78) (See pp. 367–383)
32. George, J. J., "Fog: Its Causes and Forecasting with Special Reference to Eastern and Southern United States, I–III." *Bull. Amer. meteor. Soc.*, 21:135–148; 261–269; 285–291 (1940). (See Fig. 1, p. 264; Figs. 3–5, pp. 266–268; Figs. 7–9, pp. 286–287; Table 4, p. 288)
33. Godske, C. L., Bergeron, T., Bjerknes, J., Bundgaard, R. C., *Dynamic Meteorology and Weather Forecasting.* Washington, D. C., Carnegie Institution (in press). (See equa. 11.61(1), Sects. 17·50, 18·40; Figs. 17·50·1–2, Sect. 18·45)
34. Godson, W. L., "A Study of the Deviations of Wind Speeds and Directions from Geostrophic Values." *Quart. J. R. meteor. Soc.*, 76:3–15 (1950).
35. Gold, E., "Maximum Day Temperatures and the Tephigram." *Prof. Notes meteor. Off., Lond.*, Vol. 5, No. 3, 9 pp. (1933). (See pp. 4 f.)
36. Guilbert, G., *Nouvelle méthode de prévision du temps*, Paris, Gauthier-Villars, 1909.
37. Gustafson, A., "On Anomalous Winds in the Free Atmosphere." *Air Wea. Serv. Bull.*, No. 1, pp. 21–29 (1950).
38. Heidke, P., "Über die Zweckmässigkeit einer objektiven Bestimmung von Erfolg und Güte der Wettervorhersagen." *Meteor. Z.*, 44:219–222 (1927).
39. Hess, S., "A Statistical Study of the Deepening and Filling of Extratropical Cyclones." *J. Meteor.*, 2:179–184 (1945).
40. Hesselberg, T., "Über den Zusammenhang zwischen Druck- und Temperaturschwankungen in der Atmosphäre." *Meteor. Z.*, 32:311–318 (1915).

41. Holzman, B. G., "The Separation of Analysis and Forecasting: A New Basis for Weather Service Operations." *Bull. Amer. meteor. Soc.*, 28:281–293 (1947).
42. Houghton, H. G., and Austin, J. M., "A Study of Nongeostrophic Flow with Applications to the Mechanism of Pressure Changes." *J. Meteor.*, 3:57–77 (1946).
43. Humphreys, W. J., *Physics of the Air*, 3rd ed. New York, McGraw, 1940. (See p. 359)
44. Jacobs, W. C., "A Convenient Minimum Temperature Diagram." *Bull. Amer. meteor. Soc.*, 21:297–301 (1940). (See Tables I–III, pp. 298–301; Fig. 1, p. 299)
45. Kammermann, A., "Die Vorausbestimmung des nächtlichen Temperatur-Minimums." *Meteor. Z.*, 3:124–128 (1886). (See p. 126)
46. Kessler, O. W., und Kaempfert, W., "Die Frostschadenverhütung." *Wiss. Abh. D. R. Reich. Wetterd.*, Bd. 6, Nr. 2, 243 SS. (1940).
47. Kibel, I. A., *Brief Instructions for the Diagnosis and Forecast of Temperature, Pressure, and Wind Velocity Aloft.* Moscow, U. S. S. R. Hydrometeor. Publ. House cen. Inst. Wea. cen. geophys. Obs., 1940. (Found in Kashin, K. I., *ibid.*, pp. 41 f., 1942)
48. Klein, W. H., "Winter Precipitation as Related to the 700-Mb Circulation." *Bull. Amer. meteor. Soc.*, 29:439–453 (1948).
49. Longley, R. W., "A Study of the Relationship between the 700-mb Flow and the Movement of Surface Pressure Centers." *J. Meteor.*, 4:202–204 (1947). (See p. 203)
50. Lucht, F., "Der Anteil der Stratosphäre an der Steuerung der Zyklonen und Antizyklonen zur Grund der Theorie der singulären Advektion erster Ordnung." *Meteor. Z.*, 58:11–23 (1941).
51. Machta, L., *A Study of the Observed Deviations from the Geostrophic Wind.* Sc.D. Thesis, Mass. Inst. Tech., 1948 (unpublished).
52. Mahrt, W., "Die Aufgaben der synoptischen Meteorologie im Reichswetterdienst." *Meteor. Z.*, 52:399–402 (1935).
53. Mügge, R., "Synoptische Betrachtungen." *Meteor. Z.*, 48:1–11 (1931).
54. Muller, R. H., "Verification of Short-Range Weather Forecasts (A Survey of the Literature)." *Bull. Amer. meteor. Soc.*, 25:18–27, 47–53, 88–95 (1944).
55. Neiburger, M., "Insolation and the Prediction of Maximum Temperatures." *Bull. Amer. meteor. Soc.*, 22:95–102 (1941).
56. Palmén, E., "Über die Bewegung der aussertropischen Zyklonen." *Acta Soc. Sci. fenn.*, Vol. 3, No. 7, 102 pp. (1926).
57. Penner, C. M., "Note on Radiative Cooling in the Free Atmosphere over Arctic Bay." *J. Meteor.*, 5:69–70 (1948).
58. Petterssen, S., "Kinematical and Dynamical Properties of the Field of Pressure with Applications to Weather Forecasting." *Geofys. Publ.*, Vol. 10, No. 2, 92 pp. (1933).
59. —— "Some Aspects of Formation and Dissipation of Fog." *Geofys. Publ.*, Vol. 12, No. 10, 22 pp. (1939). (See p. 15)
60. —— *Weather Analysis and Forecasting.* New York, McGraw, 1940.
61. —— "On the Sensitivity of the Wind Field Relative to Pressure Variations." *Tellus*, 2:18–23 (1950).
62. —— "Some Aspects of the General Circulation of the Atmosphere." *Cent. Proc. R. meteor. Soc.*, pp. 120–155 (1950).
63. Philipps, H., "Die Abweichung vom geostrophischen Wind." *Meteor. Z.*, 56:460–483 (1939).
64. Pogade, G., "Absterbande Warmsektorzyklonen." *Ann. Hydrogr., Berl.*, 66:343–347 (1938).
65. Rapp, R., "Forecasting Winter Precipitation Amounts at Washington, D. C." *Mon. Wea. Rev. Wash.*, 77:251–256 (1949).
66. Rodewald, M., "Die Guilbert-Grossmansche Regel in den Höhenwetterkarten." *Ann. Hydrogr., Berl.*, 65:335–337 (1937).
67. —— "Über die unvermittelt aufkommenden Nordweststürme in der deutschen Bucht." *Ann. Hydrogr., Berl.*, 65:337–340 (1937).
68. Rossby, C.-G., "Planetary Flow Patterns in the Atmosphere." *Quart. J. R. meteor. Soc.*, 66 (Supp.):68–87 (1940). (See p. 72, equation (11b))
69. Sawyer, J. S., "An Example of Cyclogenesis in Relation to Sutcliffe's Theory of Development." *Cent. Proc. R. meteor. Soc.*, pp. 107–113 (1950).
70. Schaffer, W., "A Rational Method of Delimiting Forecast Areas." *Bull. Amer. meteor. Soc.*, 30:224–228 (1949).
71. Scherhag, R., "Zur Theorie der Hoch- und Tiefdruckgebiete." *Meteor. Z.*, 51:129–138 (1934). (See p. 130)
72. —— "Die Verwendung der Höhenkarten im Wetterdienst." *Forsch. ErfahrBer. Reichs. Wetterd.*, Reihe B, Nr. 11 (1943).
73. Schwerdtfeger, W., "Untersuchungen über den Aufbau von Fronten und Kaltlufttropfen." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 3, 35 SS. (1948).
74. Showalter, A. K., *The Tornado: An Analysis of Antecedent Meteorological Conditions.* W. B. Publ., U. S. Weather Bureau, 1943. (See p. 19)
75. Starr, V. P., "On the Production of Kinetic Energy in the Atmosphere." *J. Meteor.*, 5:193–196 (1948).
76. Sutcliffe, R. C., and Forsdyke, A. G., "The Theory and Use of Upper Air Thickness Patterns in Forecasting." *Quart. J. R. meteor. Soc.*, 76:189–217 (1950).
77. Taylor, G. I., "The Formation of Fog and Mist." *Quart. J. R. meteor. Soc.*, 43:241–268 (1917). (See Fig. 15, p. 265)
78. Thomas, H., "Zur Prüfung der synoptischen Vorhersagen." *Meteor. Z.*, 50:29–31 (1933).
79. Thompson, J. C., "A Numerical Method for Forecasting Rainfall in the Los Angeles Area." *Mon. Wea. Rev. Wash.*, 78:113–124 (1950).
80. Vederman, J., "Changes in Vertical Mass Distribution over Rapidly Deepening Lows." *Bull. Amer. meteor. Soc.*, 30:303–309 (1949).
81. Wagemann, H., "Brauchbare Methoden zur Vorausberechnung von Wetterkarten." *Ann. Hydrogr., Berl.*, 60:136–151 (1932).
82. Willett, H. C., *Descriptive Meteorology.* New York, Academic Press, 1944. (See p. 280)
83. Wood, F. B., *The Formation and Dissipation of Stratus Clouds beneath Turbulence Inversions.* Mass. Inst. Tech. Prof. Notes, No. 10, 27 pp., 1937. (See Figs. 7–10)

OBJECTIVE WEATHER FORECASTING

By R. A. ALLEN

U. S. Weather Bureau, Washington, D. C.

and E. M. VERNON

U. S. Weather Bureau, San Bruno, Calif.

It is the purpose of this paper to outline the breadth of the field of study which may be called "objective weather forecasting," to describe in a general way some of the recent developments in this field, and to indicate the deficiencies and unanswered questions which have arisen in such work.

THE FORECASTING PROBLEM

Definition of Objective Weather Forecasting. In the history of weather forecasting, attempts have often been made to devise numerical and objective methods for producing the forecast. Thus Besson [2] in 1904 and Taylor [24] and Rolf [20] in 1917 produced graphical devices for representing lag relationships between selected weather variables. These studies, in common with others made in later years [4, 12, 14, 15, 27], have attempted to provide an equation or a graphical device of some form which would be useful in applying a particular relationship or combination of relationships to the problem of making a forecast. The distinction between an objective forecasting procedure and a procedure which depends on subjective judgments and subjective experience has not been sharply defined, nor is it intended in this paper to advocate a rigid definition. The purpose of this review will be served by defining an objective forecasting system as any method of deriving a forecast which does not depend for its accuracy upon the forecasting experience or the subjective judgment of the meteorologist using it. Strictly speaking, an objective system is one which can produce one and only one forecast from a specific set of data. From the practical standpoint it appears reasonable to include as objective, however, those forecasts which require meteorological training insofar as such training is standardized and is itself based upon a study of well-founded physical principles and atmospheric models which are commonly recognized from the facts of observation. It would be throwing away information of demonstrated value in forecasting if, for example, an objective forecasting system were not permitted to make use of isobaric patterns on analyzed maps because of the objection that they are arrived at subjectively. The test of whether a system is objective is whether different meteorologists using the system independently arrive at the same forecast from a given set of maps and data.

Goals of Objective Forecasting Investigations. The obvious ultimate goal of forecasting investigations is to enable the forecaster to increase the accuracy of forecasts made routinely. Contributions toward this end may be made in several ways. The forecaster may study the physical characteristics of the atmosphere, especially the dynamic relationships which have been derived on the basis of simplifying assumptions. Such study may enable him, in the course of analyzing given situations, to recognize processes in the real atmosphere which have been described analytically, and in such cases he will know better what to expect of the atmosphere in the immediate future. The success of this method of attack depends on the skill of the theoretical meteorologist in describing the real atmosphere when he sets up a model and makes simplifying assumptions, and on the skill of the forecaster in diagnosing the present sequence of events in the atmosphere, selecting the theoretical processes which are most nearly applicable, and judging what modifications are necessary in individual instances.

On the other hand, the forecaster may search for empirical relationships between observable characteristics of the atmosphere, and with little or no reference to the physical validity of the relationships, make use of them in forecasting. Many forecasters gain a high degree of skill after many years of experience because of this second factor, but skill obtained in this way is difficult to transfer from place to place or from individual to individual. It appears certain, furthermore, that some forecasters base forecasts in large part on hypothetical relationships that have neither a physical nor a statistical basis and that cannot even be expressed in objective or quantitative terms. In such a case, it is impossible to discover from data whether or not these relationships exist in the atmosphere.

Ideas for testing and possible incorporation into an objective system can come from several sources: by testing new theoretical concepts for their possible contribution to forecasting practice and providing objective ways to use the results; and by testing, combining, and systematizing the use of rules and principles which have been discovered or are already used by experienced forecasters.

The goal of objective forecasting is simply to eliminate as many as possible of the subjective elements which enter into the *application to forecasting* of the results of such studies. Objective forecasting is not so much concerned with the source of hypothetical relationships as it is with the practical value of the ideas and the extent to which they contribute to the accuracy of forecasts.

Objective forecasting studies and research projects which aim to develop objective methods or objective aids to forecasting are characterized by the use of historical data to demonstrate the reliability of forecasting rela-

tionships, and by the expression of the forecast itself in quantitative terms or at least in unequivocal terms. Fear has sometimes been expressed by forecasters that a result of the development of objective forecasting methods will be to supplant experienced forecasters by mechanical methods. It should be obvious, however, that the greater the reduction in the number of subjective and uncertain decisions required in the process of preparing the forecast, the more time will be available to the forecaster either for studying the effect of new and untried variables and the value of new principles, or for interpreting the forecast for the exceedingly diverse uses to which it is applied by the public.

From the standpoint of discovering and understanding relationships which hold in the atmosphere, forecasting investigations have been relatively ineffective because of their stress on lag relationships, and it seems clear that only a complete physical explanation of the atmosphere together with complete observational data will make it possible to produce perfect weather forecasts. Practically, however, uncertainties exist which make the maximum attainable accuracy something less than perfection [22]. The forecasting problem is thus, in essence, one of estimating what is likely to occur with any given state of the atmosphere and its environment. More precisely, the problem is to state the probability that any specified weather event will occur within any specified time interval.

The statistical or probability aspect of weather forecasting was recognized as early as 1902 by Dines [7], who pointed out the impossibility of knowing exactly what weather is going to occur and suggested that the laws of chance should be applied. Hallenbeck [9] in 1920 found an encouraging response from the public when he attempted the use of numerical probability statements as part of his agricultural forecasts. It seems to have been only recently, however, that this objective has been recognized by a large group of meteorologists and that attempts have been made to apply the methods of mathematical statistics or to develop new methods suitable for the estimation of forecast probabilities [5, 25]. Since the public generally has demanded categorical forecasts, attempts to express the "chances" of a weather event occurring have usually been frowned upon by forecasters. Nearly every decision the forecaster is called upon to make, however, involves weighing the chance as indicated by one set of factors against the chance as indicated by one or more other sets. Objective forecasting studies have not often provided final, conclusive evidence of the chance of occurrence of the weather event in question, but such studies have reduced the uncertainty to quantitative and understandable terms, and it is one purpose of such studies to determine the actual frequency with which any sequence of events which it may be desired to specify can be expected to occur.

Limitations of Objective Forecasting. Some of the objective forecasting methods which have been developed [1, 8, 17] have demonstrated a rather clear superiority to forecasts made by conventional methods under routine operational conditions. The question then arises, why have forecasters generally not seized wholeheartedly the opportunities offered by these methods of investigation for improving their own forecasts? There are a number of reasons, based partly upon misunderstanding of the methods and accomplishments of objective forecasting, but based more largely on the limitations of objective forecasting under the present organization of forecasting services.

One criticism often made is that objective methods do not take account of all of the pertinent variables nor of all the characteristics of the weather situation which help in judging the weather to come. This deficiency is certainly true of most, if not all, of the systems which have been published, but it is less a criticism of objective forecasting as such than it is an admission that items are used in forecasting which the forecaster cannot express objectively or quantitatively. The extent to which present forecasting knowledge is real knowledge as distinct from lore may be judged in part by the extent to which forecasters have been able to incorporate it into objective systems or have been able to write it down in a form which can readily be taught to students of forecasting. It is one of the goals of objective forecasting studies to collect and systematize this knowledge so it will be available for study and subsequent use by any forecaster, but major problems are encountered at this point. Even a relatively inexperienced forecaster has an ability to find in a given weather situation features which appear to be significant for the forecast, but which are extremely difficult to express in any objective way. Methods must be devised to take into account these more tenuous concepts and features of the atmosphere and to separate the real relationships from the fictitious.

The time required to work out a forecast by objective techniques is a limitation on the use of such techniques in many offices. The organization of forecasting services is generally centralized to reduce the work of plotting and analyzing the numerous maps and charts required for forecasting and to make it possible to utilize the best forecasters to cover the widest possible area. This generally results in the forecaster's being faced with tight schedules and gives him a totally inadequate amount of time to spend in actual preparation of the forecast. In spite of the success of simplified objective methods in equalling the accuracy of conventional forecasting procedures, it is suspected that forecasts which are improved to any great extent over present levels of accuracy will be produced, for the most part, only by methods which require a longer time to apply and which, therefore, cannot conveniently be used by forecasters at present.

One characteristic requirement of present-day forecasts which are issued to the public is that a definite bias must frequently be introduced into the forecast, or in other words, the forecaster must not state exactly what he thinks is most likely to happen. For example, the bare mention of rainfall in a public forecast will often be construed by the public as a forecast of rain no matter how the forecaster qualifies the statement. If the occurrence or nonoccurrence of rain is an un-

usually important matter because of previous extremely wet weather or extremely dry weather or because of an approaching national holiday, the forecaster will have to reach a decision not alone on the basis of the probability of occurrence of rain but in large measure on the anticipated public reaction to a rain (or no-rain) forecast. This type of situation is well typified in Gulf and Atlantic coastal areas when a hurricane is approaching. An objective forecasting system obviously cannot take into account these nonmeteorological factors. However, progress toward a rational solution from the standpoint of meteorology has been made by expressing the objective forecast in terms of the probability of occurrence of the specified event.

Another serious deficiency of objective forecasting systems which have been developed thus far is that the selected weather element is forecast only for individual times and places. Aviation forecasts in particular must usually cover a route of hundreds or perhaps thousands of miles, and forecasts generally must portray the expected weather trends through time intervals up to several days. These requirements impose a severe limitation on the extent to which objective forecasts have been applicable, although some progress has been made in solving this problem. New developments in the field of air traffic control suggest that probability forecasts may eventually be required for essentially all aviation purposes.

TYPES OF OBJECTIVE FORECASTS

A review of the literature on forecasting suggests three different methods of approach to the development of systems of objective forecasting. Although it is the purpose of this paper to discuss only one of these approaches, the other two are mentioned briefly in order to indicate in what way they offer the possibility of improved forecasts.

Numerical Calculation. The method of numerical calculation which was first attempted by Richardson [19] has recently become of practical importance through the work of Charney and Eliassen [6] and the development of high-speed electronic computers. This method of producing a forecast, based solely on the solution of simplified equations of the atmosphere, showed little promise of practical value so long as hand computation was required to obtain the solution. The value of the equations of motion for applied forecasting was limited to their ability to suggest variables which might then be found, empirically, to be of forecasting significance. The possibility now exists, however, that a prediction of the contour pattern at an upper level can be completed by a high-speed computer in time to be of use in preparing the weather forecast. This development requires careful consideration by those engaged in research on objective forecasting to insure that the information in such prognostic contour charts can be fully incorporated in objective techniques for preparing the weather forecast. Relatively little work has been done on the use of prognostic charts, and few forecasting systems can be evaluated in a way that would show the increase in accuracy to be had with a perfect prognostic chart or that give any hint of the possibility of using information from a prognostic chart in preparing the forecast. Studies of this kind have been made by Mook and Price [13] and by Klein [11].

Statistical Methods. Although all forecasting except that done by numerical calculation is in reality based on the statistical processing of data, there are differences in the extent to which a logical foundation of meteorological and physical principles is built up prior to the application of statistics. One possibility is to use statistical methods exclusively in the search for variables, retaining for prediction purposes only those variables which have a statistically significant relation to the element being forecast. In the field of long-range forecasting particularly, this approach has been appealing and to some extent perhaps necessary because of the absence of enough knowledge of the causes of long-term fluctuations in the general circulation to provide a useful mathematical-physical foundation. Among those leading in such studies have been Walker and Bliss (for a critique and bibliography of Walker's work up to 1936, see [16]).

More recently in the field of short-range forecasting, the statistical approach has been investigated at some length by Schumann [21] and Wadsworth [28]. Meteorologists have criticized this type of study on the basis that the statistical selection of variables for prediction without resort to prior meteorological knowledge is basically unsound [23], and this has led in some cases to rather sharp differences of opinion between meteorologists and statisticians. Sutcliffe, for example, takes the extreme view that meteorologists, by developing a physical foundation for all relationships used in forecasting, should drive statisticians out of meteorology. On the other hand, Professor Norbert Wiener discusses this question in an unpublished note [29], in which he points out that in meteorology the role of dynamics is to suggest what quantities one should expect to find interrelated, and that the role of statistics is to verify the correctness of the dynamics. Although this may be a somewhat oversimplified statement of the situation, it seems clear that any meteorological hypothesis which is claimed to be of use in forecasting must be tested statistically before its value will have been demonstrated, even though it may express a theoretical relationship based on what seem to be reasonable assumptions. Conversely, of course, relationships which are derived solely from empirical considerations, even though supported by a large number of cases, cannot be safely adopted as real (*i.e.*, practically certain to exist in future data) unless they can be shown to have a rational physical basis.

To the extent that the goal of a study is to increase the accuracy of forecasts, the proof of success will rest in the verification of the forecasts, and this is largely a statistical matter. In order that the interplay between the physical and statistical approaches can be effective, somewhat greater recognition needs to be given to the place of modern statistical methods in meteorology. At the same time, it must be recognized that statistical tools, for example, correlation or regression, can easily

be misinterpreted if the user is unfamiliar with their purpose or if he tends to rush into a statistical treatment without first considering the meteorology of the situation [10, 18].

A report by Wadsworth [28] illustrates some of the recent work in statistical forecasting; but this subject is to receive a more complete treatment elsewhere.[1]

Combined Deductive-Empirical Methods. The method of numerical calculation has the disadvantage that it is difficult (*i.e.*, it requires a large number of calculations) to permit actual atmospheric data to determine the coefficients of the forecasting equations used, and the equations which are presently available do not forecast the final weather elements. Other theoretical approaches to forecasting often are of little immediate value to forecasters because the application of the new concepts and new principles which result is not a straightforward procedure. The forecaster finds that he must expend considerable effort in applying the new principles to a variety of cases in order to learn what modifications to make and the conditions under which they apply. This situation appears to result in a gap between the development of new dynamic techniques and their application to routine forecasting practice.

On the other hand, statistical methods often depend too largely on empirical results and do not make adequate use of ideas which result from theoretical deductions. A new forecasting technique is required which will neither neglect the indications of theory nor require subjective interpretations and projections of weather maps and charts in order to arrive at a forecast. The methods of approach which have been under recent development in the United States will be described briefly in the following section. In this discussion, a distinction will be made between the terms "forecast method" and "forecast aid." The term "forecast method" will imply that the objective system produces a final forecast, and that in actual practice the forecaster should seldom if ever modify this forecast, no matter what he may feel about its chances for success. The term "forecast aid," on the other hand, will be used in the sense of a tool or an indication which the forecaster must consider and weigh subjectively with other available indications.

METHODS OF DERIVING OBJECTIVE FORECASTING AIDS

Specification of the Problem. A distinguishing characteristic of work which has been done on objective methods of forecasting is that the forecasting problem is attacked a little at a time. A problem is defined objectively by specifying a single weather element to be forecast, a single point or a specified set of points for which the forecast is to be made, and a single time or a time period to be covered by the forecast. The practical value of the result of the study depends in large part on the extent to which the specification of the problem meets practical forecasting requirements. The local forecast for a given city usually requires, for example, that the occurrence or nonoccurrence of precipitation be forecast by 12-hr periods or perhaps even by shorter periods. Thus an objective method which forecasts precipitation occurrence for a 24-hr period does not entirely meet the usual requirements, although it may nevertheless be of value in preparing a shorter-term forecast, and may be useful as a preliminary study. Limitation of the work to a single problem in this way is necessary in order that the weather element being forecast can be measured or objectively specified in some way. It imposes no limitation on the kinds of forecast problems that can be investigated, for the preparation of any type of forecast that can be objectively verified, such as area or route forecasts, cold-wave warnings, or the deepening of lows, can be studied by objective procedures.

Developing Hypotheses for Testing. The crux of the problem of increasing the accuracy of forecasts is that of discovering new and more exact relationships between observed parameters of the atmosphere and subsequent weather. The efficient application of such relationships to practical forecasting is the goal of objective forecasting, but without an underlying model of the atmosphere and its processes, based on the best available physical facts and theories, no amount of systematizing and testing will produce the desired results.

Numerous sources of ideas are available. In maximum-temperature forecasting, for example, equations are available for computing the effect of insolation on the maximum temperature [14] and, in theory, other effects can be evaluated. In practice, the most serious disturbing factors seem to be those for which quantitative data are scarce, such as albedo and outgoing long-wave radiation, or those which in turn depend on a forecast, such as cloud cover and temperature advection.

Recent developments in dynamic meteorology have provided new concepts which have not been adequately tested. Relationships involving wave length, zonal wind, and wave propagation in the upper air, and the various theories subsequently developed which have contributed to the understanding of large-scale phenomena in the atmosphere, need to be stated in ways that are amenable to objective evaluation in connection with short-range forecasting. This is the purpose of work begun recently at the University of Chicago, and a report on tropical cyclone forecasting [3] illustrates the method of attack. Such a report is only the beginning, however, and a large amount of work remains before the results can be objectively applied in forecasting.

A number of papers have been published within the last few years which describe forecasting studies wherein the ideas seem to have been only those suggested by a superficial study of the physical processes which are involved. A more logical approach has been initiated by the Honolulu office of the U. S. Weather Bureau, which, as a preliminary step in developing objective methods of rainfall forecasting, has prepared a lengthy report on the physical causes of rainfall on the Hawaiian Islands,

1. Consult "Application of Statistical Methods to Weather Forecasting" by G. P. Wadsworth, pp. 849–855 in this Compendium.

incorporating the ideas of all the forecasters as to basic causes or simultaneous relationships [26]. A study such as this is a major project in itself and is practicable only for a forecasting staff whose members are highly cooperative and willing to put aside work on other study projects to collaborate on an intensive study of a single weather element. The resulting report, aside from the immediate value which it may have for conventional forecasting techniques, furnishes a sound background of ideas for testing and development.

Another source of hypotheses which has not been sufficiently exploited is the "case study" approach. The preparation of "case books" consisting of detailed analyses of a number of individual weather situations such as cases of heavy snow at a city or in an area, or cold-wave cases, has sometimes been suggested as a useful aid for forecasting. Case studies alone may suggest *necessary* criteria, but further testing is necessary in order to establish sufficient criteria.

Testing Procedures. Although objective forecasting procedures are distinguished from conventional forecasting techniques chiefly by the fact that they are based on objectively expressed and tested ideas rather than on subjective and untested judgment, very little has been written about the testing procedures which are applicable. The reason for this appears to be the lack of any standard set of procedures which are sufficiently generalized that they can be advocated for use on all forecasting problems. A number of papers on objective forecasting have been published, most of them in the *Monthly Weather Review* since 1948, representing nearly as many different methods of procedure as there are papers. Other projects now under way will undoubtedly present improvements from the standpoint of more accurately representing the complex division of different weather types with which a forecaster must deal.

Brier [5] first presented a method of combining any number of meteorological variables by means of scatter diagrams to produce a weather forecast. The forecast can be expressed either in terms of the probability of occurrence of an event, or as a forecast of the magnitude of the weather element. The use of scatter diagrams in the search for forecasting relationships was not new, but Brier's study appears to be the first to combine a large number of variables into a simple forecasting procedure by this method. The use of scatter diagrams has since been widely adopted, and it has been a useful substitute for more rigorous methods of testing.

The statistical methods of correlation and regression are suitable for expressing forecasting relationships and testing the validity of those relationships if the assumptions which underlie their use are valid. Thus the relationship between temperature and predictors such as normal temperature, temperature of the approaching air mass, expected turbulence, and expected cloudiness may be expressed by a multiple regression equation. This is possible even if the relationships are nonlinear, provided the form of the relationship is known so that a transformation to a linear form can be made. If a long record of observations were available so that the present climatic era could be randomly sampled, valid tests of significance of the relationships could also be made.

For practical purposes, although rigorous statistical methods are useful in guiding the selection and processing of data, it is not often possible to make valid tests of significance, and the forecaster must depend upon an independent data sample to indicate the reliability of the forecasting relationship being tested. In investigations such as these, observational data may indicate how selected atmospheric events are interrelated, and thus stimulate the conception of new hypothetical relationships, or they may be used as evidence of the truth or falsity of a priori, objectively expressed hypotheses. In selecting data for testing purposes, the investigator must adopt the role of impartial umpire; otherwise he is likely to select or reject cases according to whether they do or do not support the relationship under test.

RECOMMENDATIONS FOR FUTURE RESEARCH

The development and use of objective forecasting methods encounters various unsolved problems such as inadequate data and insufficient time for the forecaster to work through the procedure. However, the basic problems which are encountered in common with any other method of forecasting are a lack of understanding of the physical processes in the atmosphere and an inability to apply to practical problems such knowledge as now exists. Most of the research in objective forecasting has been done by meteorologists whose primary interest is in forecasting and who feel the need for more systematic tools. Further progress could be expedited if a closer working relationship could be established between theoreticians and forecasters having an interest in the development of objective methods. As pointed out above, qualitative descriptions of how a forecaster can use the results of basic research are useful as a source of ideas, but assumptions must be made before the new parameters can be measured from observational data or from analyzed maps and expressed quantitatively, and the theoretician can best say which assumptions are most in accord with the theory.

Many of the objective forecasting methods developed in recent years make use of variables derived directly from station observations. The result is that changes in station location or the discontinuance of a station may invalidate a method until it can be tested by using a substitute variable. Claims have sometimes been made that the methods are more accurate when station data rather than values interpolated from analyzed maps are used. This subject needs more investigation, for if the claims are valid, it follows that administrators should be more slow to move or to discontinue stations, and further that much of the emphasis on centralized, accurate analysis is misplaced. The success of some of the objective methods which use station data is so great that this point cannot be lightly dismissed.

Graphical techniques such as the scatter-diagram technique used in searching for and testing relationships are unsatisfactory in spite of their success, because

of the considerable amount of subjectivity that may appear in the graphical analysis. More rigorous methods are needed, although it must be noted that difficulties seem to arise only when scatter diagrams are analyzed with too few data, or when there is no rational physical model underlying the selection of variables and the pattern of isograms drawn on the scatter diagram.

In spite of these difficulties, the methods illustrated in recent literature have had a rather astonishing success in producing useful forecasts, and it can be recommended that more forecasters attempt such investigations of their own local forecasting problems.

REFERENCES

1. Beebe, R. G., "Forecasting Winter Precipitation for Atlanta, Ga." *Mon. Wea. Rev. Wash.*, 78:59–68 (1950).
2. Besson, L., "Attempts at Methodical Forecasting of the Weather." Translation. *Mon. Wea. Rev. Wash.*, 32:311–313 (1904).
3. Boyce, R. E., and others, *Tropical Cyclone Forecasting.* Chicago, University of Chicago, Dept. Meteor., 1950.
4. Brier, G. W., *Predicting the Occurrence of Winter Time Precipitation for Washington, D. C.* U. S. Weather Bureau, Washington, D. C., 1945.
5. —— "A Study of Quantitative Precipitation Forecasting in the TVA Basin." *U. S. Wea. Bur. Res. Pap.* No. 26 (1946).
6. Charney, J., and Eliassen, A., "A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies." *Tellus*, Vol. 1, No. 2, pp. 38–54 (1949).
7. Dines, W. H., "The Element of Chance Applied to Various Meteorological Problems." *Quart. J. R. meteor. Soc.*, 28:53–68 (1902).
8. Gringorten, I. I., "A Study in Objective Forecasting." *Bull. Amer. meteor. Soc.*, 30:10–15 (1949).
9. Hallenbeck, C., "Forecasting Precipitation in Percentages of Probability." *Mon. Wea. Rev. Wash.*, 48:645–647 (1920).
10. Hess, S. L., "Reply to Mr. Namias' Comments on the 'Prognostic Usefulness of the Correlation between Pressure and Temperatnre Changes at 3 km.'" *Bull. Amer. meteor. Soc.*, 30:59–60 (1949).
11. Klein, W. H., "An Objective Method of Forecasting Five-Day Precipitation for the Tennessee Valley." *U. S. Wea. Bur. Res. Pap.* No. 29 (1948).
12. Landsberg, H., "On Climatological Aids to Local Weather Forecasting." *Bull. Amer. meteor. Soc.*, 22:103–105 (1941).
13. Mook, C. P., and Price, S., "Objective Methods of Forecasting Winter Minimum Temperatures at Washington, D. C." *U. S. Wea. Bur. Res. Pap.* No. 27 (1947).
14. Neiburger, M., "Insolation and the Prediction of Maximum Temperatures." *Bull. Amer. meteor. Soc.*, 22:95–102 (1941).
15. Nichols, E. S., "Notes on Formulas for Use in Forecasting Minimum Temperatnre." *Mon. Wea. Rev. Wash.*, 54:499–501 (1926).
16. Page, L. F., and others, "Reports on Critical Studies of Methods of Long-Range Weather Forecasting." *Mon. Wea. Rev. Wash.*, Supp. No. 39 (1939).
17. Palmer, W. C., "On Forecasting the Direction of Movement of Winter Cyclones." *Mon. Wea. Rev. Wash.*, 76: 181–201 (1948).
18. Panofsky, H. A., "Significance of Meteorological Correlation Coefficients." *Bull. Amer. meteor. Soc.*, 30:326–327 (1949).
19. Richardson, L. F., *Weather Prediction by Numerical Process.* Cambridge, University Press, 1922.
20. Rolf, B., *Probabilité et pronostics des pluies d'été.* Doctor's Thesis, Upsala, 1917.
21. Schumann, T. E. W., *Statistical Weather Forecasting.* Meteor. Res. Bur., Pretoria, Un. of South Africa, 1947.
22. —— "The Fundamentals of Weather Forecasting." *Weather*, 5:220–224, 248–253 (1950).
23. Sutcliffe, R. C., "Forecasting Research." *Meteor. Mag.*, 78:61 (1949).
24. Taylor, G. I., "The Formation of Fog and Mist." *Quart. J. R. meteor. Soc.*, 43:242–268 (1917).
25. Thompson, J. C., "A Numerical Method for Forecasting Rainfall in the Los Angeles Area." *Mon. Wea. Rev. Wash.*, 78:113–124 (1950).
26. U. S. Weather Bureau, *Physical Basis for Production of Rain in Hawaii.* U. S. Weather Bureau, Honolulu, T. H., August, 1950 (unpublished).
27. Vernon, E. M., "An Objective Method of Forecasting Precipitation 24–48 Hours in Advance at San Francisco, California." *Mon. Wea. Rev. Wash.*, 75:211–219 (1947).
28. Wadsworth, G. P., "Short Range and Extended Forecasting by Statistical Methods." U. S. Air Force, *Air Wea Serv. Tech. Rep.* No. 105–38 (1948).
29. Wiener, N., *Brief Note on the Relation between Dynamical and Statistical Meteorology.* Mass. Inst. Tech., Cambridge, Mass., 1944 (unpublished).

GENERAL ASPECTS OF EXTENDED-RANGE FORECASTING

By JEROME NAMIAS

U. S. Weather Bureau, Washington, D. C.

Absolute definition of the term "extended-range forecasting" is not possible and probably will not be for many years. As understood today it vaguely refers to those predictions which go beyond the limit imposed by ordinary standardized forecast techniques. The latter forecasts, disseminated by meteorological services the world over, generally cover periods of 24, 36, or at most 48 hr. With this in mind, "extended-range forecasting" presumably embraces any atmospheric prediction involving any time period beyond 48 hr and up to the geological epochs including the ice ages. The field of extended-range forecasting thereby becomes most inviting to people both within and without professional meteorology, particularly in view of the tremendous potentialities it offers in the direction of improving the lot of man.

Considering the extent and complexities of the problems posed by the vast spectrum of long-range weather forecasts, meteorologists have made small and slow progress. This spectrum may be divided arbitrarily into the broad bands: medium range (covering periods from three or four days to a week), monthly, seasonal, annual, decennial, centurial, and so on to millennial.

In spite of the fact that many workers have spent arduous years attempting to find solutions applicable to these periods, no system has evolved which even remotely approaches reliability. Unfortunately, the same comment can be made in regard to the short-range forecasts covering as little as 24 hr in advance. This sad state of affairs reflects the lack of basic knowledge of the workings of the atmosphere and points up the need for periods of stocktaking as exemplified by this Compendium.

But if objectives appear distant, it does not mean that progress has not been made. Indeed, medium-range forecasts covering periods from a few days to a week are already proving economically valuable in many countries. Forecasts of general weather conditions for periods a month in advance have shown promise. There are even some optimistic meteorologists who believe that seasonal, annual, or even decennial forecasts of some degree of reliability may be possible in the not-too-distant future. For the most part, this optimism cannot be traced to statistically significant results in the prediction of weather phenomena. If we look at the problem through the harshly revealing magnifying glass of statistics, it appears that, of the spectrum of extended-range forecasts cited, only the medium-range and perhaps monthly forecasts have demonstrated some degree of skill. The term skill is used here in its rigid statistical sense, that is, that the predictions are correct more often than are predictions made from many years of climatic records, or that forecasts must in the long run show a higher proportion of successes than the climatic probability of occurrence of the event being forecast.

The use of this yardstick makes it possible to restrict this article to only those two forms of extended-range forecasting, medium range and monthly, which in the knowledge of the author have met this criterion over a period sufficiently long to be considered significant. It must be admitted that this manner of selection may be somewhat unfair to those who claim ability to forecast for decades, centuries, and longer, for adequate meteorological records are as yet unavailable to verify these predictions.

THE MEDIUM-RANGE FORECAST PROBLEM

The historical development of modern short-range weather forecasting has been essentially inductive in nature. A multitude of observations taken at small space and time intervals are organized into an analysis designed to give physical meaning to the observations. This analysis is then compared with earlier analyses, generally in order to ascertain changes in circulation and weather over larger intervals of time and space, and by this method, essentially kinematic, a prediction is made possible. To the extent that the weather is determined by features of the weather map of roughly the dimensions of the cyclone and the anticyclone, the forecasts are successful. But the day-to-day movement of cyclones and anticyclones is often of roughly the same order of magnitude as their dimensions (their diameters). This rapid motion of weather systems, often coupled with erratic path, speed, and development, points up the difficulty of attempting to make forecasts for more than a day or two by using standard forecast methods which still rely, for the greatest part, on methods of interpolation and extrapolation.

Perhaps the first real forward step in crystallizing the problem of extended-range forecasting came with the discovery of the great "centers of action"—a term first applied by L. Teisserenc de Bort. These centers are brought to light by the statistical averaging of daily pressure maps during one month (or season) for many years. But if the averaging is carried out only for one month of one year (or even for only one week), one also obtains clear-cut large-scale features of the general circulation usually many times the size of ordinary migratory cyclones and anticyclones of the weather map. Since the time of Teisserenc de Bort (1855–1913) it has been recognized that it is the position and intensity of these centers of action that govern, or at least are associated with, the longer-period characteristics of the weather. The use of "govern" has been and still is objectionable to that diminishing group of meteorologists who view the centers of action only as a statistical average of randomly behaving cyclones and anticy-

clones which for certain periods happen to find some favorite life history and path.

In later years fresh emphasis on the importance of the centers of action in governing general weather conditions for long periods was afforded by Baur, whose work forms the basis for another article on extended-range forecasting elsewhere in this volume.[1] Baur recognized certain broad-scale (macroscopic) features of the general circulation which were effective in "steering" the individual cyclones and anticyclones along fairly well determined paths. He referred to these as *Grosswetterlagen*, a term now known the world over as denoting the broadscale weather situation as distinct from the day-to-day snapshot of the weather map. Moreover, it was Baur who was largely instrumental in associating the large-scale weather situation observed at the surface with the flow patterns in the middle and high troposphere and in ascribing to this latter flow pattern the property of steering. Contemporaneously, the Russian meteorologist, V. P. B. Multanovsky, had also discovered rather indirectly the large-scale weather situation in "natural periods." During these periods cyclones and anticyclones tend to follow in much the same tracks as their predecessors, that is, a series of cyclones (or anticyclones) during a natural period appear to move within a restricted region.

The work of these two investigators was certainly not unique in the meteorological world, for there may be found in the literature of the past half century considerable evidence that many weather forecasters had been toying with these ideas. But Multanovsky and, particularly, Baur became the most persistent advocates of the large-scale approach.

Despite the fact that the concept of *Grosswetterlagen* enjoyed increasing attention prior to World War II, no one had succeeded in developing a technique for predicting the strange and different large-scale weather situations which unquestionably showed up in the form of the natural periods of Multanovsky, in the steering patterns of Baur, or in the multitudinous systems of map typing proposed by various groups throughout the world.

Indeed it was not until the late 1930's that a physical theory of the evolution of different forms and positions of the great centers of action was proposed. This appeared in two papers: one by C.-G. Rossby and collaborators, "Relations between Variations in the Intensity of the Zonal Circulation and the Displacement of the Semi-Permanent Centers of Action" [17], and another by J. Bjerknes, "The Theory of Extratropical Cyclone Formation" [3]. From these epoch-making papers and considerable subsequent work, particularly by the American school of meteorologists, it has become increasingly clear that while earlier investigators were correct in stressing the broad-scale aspects of long-range forecasting, *the problem probably involves an even larger scale than they had assumed*, perhaps world-wide interactions. The recognition of the global nature of the problem arose essentially from two facts which in the past ten years have been documented with considerable empirical evidence: (1) that the behavior of any center of action depends upon that of others in the same hemisphere, regardless of how remote; and (2) that the speed of propagation of planetary wave energy from one system to another often takes place at a rate appreciably greater than the actual speed of the air particles in any observable atmospheric layer. The credit for the first discovery belongs largely to Rossby as indicated in his 1939 article [17] and in subsequent papers. To him also belongs much of the credit for proposing on theoretical grounds the second atmospheric characteristic of "dispersion," although this had been discovered earlier in forecasting practice by the author [14].

To the knowledge of the present writer there exists only one method of attack, to be described later, which utilizes directly these two concepts through the medium of hemispheric surface and upper-level charts. This method, first practiced in the United States, is now gaining ground in certain European countries, and during the past twelve years hundreds of related papers have appeared, some of which are summarized elsewhere in this Compendium (*e.g.*, see articles dealing with the general circulation).

Yet it must not be assumed that this vast research has led to amazing success in the prognostication of weather for long periods. To be sure, prediction is the ultimate goal here just as it is in all sciences. The point is that these global methods for the first time enable us to explain on a rational physical basis the evolution of large-scale weather types (*Grosswetterlagen*). The degree to which this "explanation" can be expressed in quantitative terms will determine the predictive skill of the method.

Once the two phenomena, hemispheric teleconnection of the centers of action and the dispersive character of planetary atmospheric waves, are recognized as facts, it becomes obvious that any system of extended-range forecasting based solely upon the data for a limited area of the hemisphere is bound to have imperfections. The degree of imperfection will presumably increase as the area considered in preparing the prediction is diminished. For this reason certain forecasting methods to be described and compared suffer inherent difficulties which place upon their maximum possible degree of skill a ceiling which is well removed from perfection. Among the methods to which these restrictions particularly apply are those involving symmetry points, singularities, regional weather types, and trends (kinematics).

The foregoing paragraph is not meant to indicate that these methods are of no value. Indeed, in our present imperfect state of knowledge of extended-range forecasting these tools, used judiciously, can assist materially in obtaining forecasts of some degree of skill. But these tools in themselves, without consideration of interhemispheric reactions, can hardly be expected to raise forecasting skill to a satisfactory level.

1. Consult "Extended-Range Weather Forecasting" by F. Baur, pp. 814–833.

METHODS OF MEDIUM-RANGE FORECASTING

Statistical Methods. Inasmuch as the physics of large-scale weather phenomena has been so imperfectly understood during the past half century, it seems natural that purely statistical approaches to the problem would be taken. These approaches have been many and varied. Some of them have been conducted with indiscriminate use of the correlation technique in the pious hope that the atmosphere would "speak for itself" and thus relieve man of the apparently superhuman task of working out its physical behavior. For the most part such hopes have been consistently thwarted by an atmosphere which apparently resists such crude forms of regimentation represented by the strait-jacket technique of simple lag correlation.

Other statistical, or chiefly statistical, techniques which possess some background of physical reasoning have enjoyed more success. Among the most widely known and practiced of these, and those to which we shall refer, are methods involving trends, singularities, symmetry points, weather types, and weather analogues. A brief (though necessarily incomplete) description of each of these methods will be given, followed by an attempt to evaluate and compare the various methods.

Longer-period trends seem to infiltrate many systems of extended-range forecasting. In their simplest form, it is the purpose of these methods to estimate the rate of change of a state of the general circulation, regionally or for several regions, and from this rate and the initial state to extrapolate the circulation for desired periods of time. Precisely this sort of thing forms the basis of perhaps 80 per cent of short-range forecasts that are issued today. The problem has been rendered more tractable in short-range than in extended-range forecasting by kinematic methods applied by Petterssen [16] and others, and more recently by the simple expedient of preparing synoptic charts every three or six hours. In this manner the tendencies are easily deduced and from these and from the continuity, extrapolations are easily made.

In extended-range forecasting, however, one is faced with the questions, first, What *is* the initial state of the large-scale circulation? and second, What is the tendency interval desired? Various methods have been proposed for determining the initial large-scale state of the circulation. All of these essentially involve some form of smoothing, the effect of which is to damp out the smaller-scale and presumably secondary irregularities of the circulation, chiefly migratory cyclones and anticyclones, and thereby bring into focus the great centers of action. Thus the Russian school plots composite charts on which are indicated singular features of the surface pressure distribution, particularly cyclones, anticyclones, ridges, troughs, etc., for certain "natural periods" during which the locations and tracks of these features are similar [15]. The areas persistently occupied or invaded by anticyclones or cyclones thereby become clear, and the composite chart gives a good clue to the dominating centers of action. In the U. S. Weather Bureau the smoothing is achieved by the procedure of averaging over intervals of time. In this manner five-day (or other period) means are prepared by averaging the pressures (or temperatures) interpolated from a grid of latitude and longitude intersections from twice-daily Northern Hemisphere charts at sea level and aloft. When isolines are drawn to these values, smooth patterns appear which usually leave no question as to the position and intensity of the centers of action. Other methods of smoothing are also possible and it may be that a simple arithmetic mean is not the best for the purposes of extended-range forecasting. Further research along these lines would appear to be desirable.

Of late some have advocated that a sufficient degree of smoothing for ordinary medium-range forecasting work is achieved by merely resorting to charts for higher elevations, perhaps 10 to 13 km. While it cannot be denied that these charts appear simpler in form than those in the lowest layers of the troposphere (except at low latitudes), there is some question as to what extent the greater simplicity arises from fewer and less accurate data which give freer rein to the analyst's imagination. Besides, the slow evolution so characteristic of means taken over longer time intervals is easily lost when each snapshot of the circulation at any level is inspected from day to day.

Having decided upon the initial state of the general circulation, the *Grosswetterlagen*, the next problem in the estimation of trends is to find the necessary "tendency interval" and methods of determining the tendency quantitatively. Here one rough method is to compare the last available daily chart with its preceding sequence, again using some integrating device like the mean or composite chart. The reason why such a method has some success in determining the trend of the large-scale features of the circulation is that there is appreciable serial correlation in meteorological data such as pressure. As a result, today's chart will bear a maximum resemblance to the mean chart whose period centers on the current day—provided the period of averaging is not too large. Thus the comparison of daily charts with mean charts may throw light on the larger-scale trends. Because of its subjective nature, this procedure is often not satisfactory and for this reason a more quantitative method of evaluating trends has been developed in the U. S. Weather Bureau. Inasmuch as this method is closely related to other statistical methods to be compared, and perhaps clarifies these methods, it will be described in some detail.

If we wish to treat the slow evolution of large-scale circulation patterns which, for example, appear at the 700-mb level, it would be helpful to be able to construct a tendency field to serve the same function as does the 3-hr tendency field of daily charts. The choice of a tendency interval, while somewhat arbitrary, must be small compared to the length of the period over which the means are taken and should preferably be centered around the mean. In the case of five-day means, for example, a tendency interval of two days might be a good choice. What is desired for each location, then, is the slope of a hypsogram representing a running five-

day mean of the 700-mb surface. The two-day tendency would then be the slope of the hypsogram from one day preceding to one day following the mean on the day the tendency was desired.

The use of the persistence correlation (or the short-range, 24-hr prognosis) makes it possible to estimate the values of the predictors for the tendency interval. The statistical method of doing this, described in more detail elsewhere [11], involves the computation of regression equations which express unknown (subsequent) values as a function of the last-known quantities and the normal values for the particular time and place. The basis of these computations rests in the observed approximately exponential behavior of the autocorrelation coefficient for pressure (and other continuous meteorological elements) as expressed by the equation:

$$r_n = r_1^n,$$

where r_1 is the simple linear one-day lag autocorrelation coefficient for the departure from normal of pressure, and n refers to the number of days between values being correlated. While r_1 varies with season, elevation, and geographical location, its value at mid-troposphere levels is surprisingly uniform (around 0.7). This uniformity makes tendency computations by this method practicable in routine forecasting procedure.

These height tendencies are computed for a desired grid of points, and from them a field of isallohypses (corresponding to isallobars) may be constructed. From the tendencies and mean contour fields it is possible to make kinematic computations of the movement and development of the principal singular features of the mean maps such as troughs, ridges, centers, and so on, for desired periods of time.

In short, the particular method just described, called the trend method, utilizes both the persistence of pressure and its tendency to be restored to certain normal values which vary from place to place and from month to month. It makes use of these two factors together with the past performance characteristics of the atmosphere in an attempt to evaluate longer-period trends of large-scale circulation features and results in tendency fields which apply to moving and developing centers of action. The use of climatological normals and extremes as a qualitative guide in projecting thickness patterns for extended-range forecasting work has recently been started by Sutcliffe and his colleagues in England [20]. It appears that the trend technique just described might be applied with advantage to thickness patterns as well as to contour patterns.

Another statistical method using a related technique is currently being used by German meteorologists. This method, credited to Baur, involves correlation tables. The ultimate purpose is to arrive at a forecast of the pressure distribution three days in advance. The method relies upon multiple correlations computed from a number of variables, important among which are the current day's pressure, pressure tendencies for different time periods, and the cloudiness. Tables giving the expected three-day changes in pressure are worked out for several stations in Europe. The grid of stations permits the construction of a prognostic chart of three-day isallobars. An attempt is then made to decide subjectively how the prognostic changes might come about from the initial state.

Clearly, this method has some features in common with the trend method of the U. S. Weather Bureau, described previously. The latest daily pressure and certain pressure tendencies used in the formulation of regression equations undoubtedly enter, though not necessarily explicitly, into both methods. But in the use of the multiple correlation technique it is of some importance to evaluate the contribution of each of the participating terms to the final correlation. To the knowledge of the present author, no figures on this point are available, and it seems quite possible, if not probable, that some of the terms used are not contributing to the goodness of the final correlation. The trend method, utilizing only autocorrelation and the normal restoring force, suffers from no such ambiguity. The contributions of its components are known through many statistical works [22]. Besides, the computed tendencies can be applied in an objective manner to the initial large-scale pressure patterns as reflected in the mean states.

Still another allied statistical procedure used in extended-range forecasting is the method of singularities. The primary contention of this method is that certain characteristic circulations and weather tend to recur in certain areas at almost the same time of the year, year in and year out. One might in this connection refer to the many papers of Schmauss (see, for example, [19]) which have appeared in German meteorological publications, and a summarization of some allied work by Brooks [4]. In the practical use of this method in the French Meteorological Service, normal curves of sea-level pressure for periods of five days are computed for a number of stations in Europe and the eastern Atlantic. In spite of the long periods of record from which they are constructed, these curves show certain irregular convolutions which are superimposed upon gradual seasonal trends. It is the contention of the followers of the school of singularities that many of these irregularities are genuine and would not be smoothed out by adding more years of record to the data. Hundreds of papers have appeared in the literature of the past twenty years in which attempts were made to endow these singularities with statistical significance. The arguments pro and con are still reverberating in the meteorological world, and in all probability it is safe to say that the last word on singularities has not been said nor will be for many years to come.

Returning to the practical use of the method, at least as the author saw it practiced in France, a plot is made of the observed five-day mean pressures for the past several months. This plot is made on a tissue and then superimposed on a normal curve with the same time scale. An approximate match is chosen chiefly by the positions of troughs and ridges in the curves, little consideration being given to amplitude or general mean level. In other words, parallelism is used as the principal indicator of singularity, and a slight sliding of the time scale generally becomes necessary to find the singu-

larities. Once the singularities have been determined, individual five-day mean curves are extrapolated more or less parallel to the normal curves (but the amplitudes may be suppressed or exaggerated according to the seasonal behavior relative to the normal). This procedure is followed for several points in Europe and the eastern Atlantic and results in the preparation of a five-day mean prognostic pressure-change chart which serves as an extended-range forecasting aid.

The similarity of this method to the trend method lies in the fact that it, too, makes some use of persistence and the normal. Persistence is incorporated by drawing and keeping the extrapolated five-day mean curves at the same general level (departure from normal) at which they have been operating. To the extent that singularities are real, the method also introduces the normal and its rate of change. It is difficult to compare trend and singularities methods as to efficacy, especially since one of them (singularities) is to a considerable extent subjective. At least it can be said that the trend method, aside from its greater objectivity, arrives at a method for predicting an orderly evolution of an existing large-scale state with a minimum of questionable basic assumptions.

Another statistical procedure—one concerning which considerable controversy has arisen—involves the concept of symmetry points. The strongest advocate of this method has been Weickmann who, along with his associates, has published many papers dealing with this subject (see, for example, [24, 25]). During World War II, investigations of the symmetry-point technique were made in Britain (*e.g.*, Walker [23] and the following discussion by C. E. P. Brooks, N. Carruthers, and others) and in the United States by Haurwitz [6]. Both of these groups could find little in the way of prognostic significance in the symmetry-point method. Their methods of testing undoubtedly involved the best available statistical procedures, and it would indeed be difficult to dispute the findings of these distinguished investigators who, while admitting the existence of symmetry points in meteorological data, could find no method of detecting them in advance. Presumably, if the faith of the users of this method is still unshattered, it is to be concluded that its use involves certain subjective elements which are as yet impossible to express quantitatively in a form suitable for testing.

In order to pursue this interesting topic further, it will be helpful to describe briefly the practical procedure currently being used by Weickmann's assistants working at Bad Kissingen. First, they work not with isolated points but with areas roughly 500-km square. The mean pressures of about ten stations in many such areas are plotted day-by-day in graphs. The graphs are then copied on transparent paper. By turning over the paper one obtains a mirror image and with the help of the original graphs tries to locate by inspection a point of symmetry. Cases of "inverted" symmetry, where the tissue must be folded along a horizontal line, are also sought. By working in this manner for several areas and then copying off portions of the observed curve as it would be reflected to produce the symmetry, an extrapolation of the pressure distribution into the future may be made.

Everyone must agree, along with Haurwitz, Brooks, and Carruthers, that symmetry points do exist in meteorological data. However, Baur [2], in a statistical study, claimed the existence of such symmetry can be entirely accounted for by the serial correlation (persistence) present in day-to-day pressures. To what extent symmetry points contain more information than a series of almost random data in which persistence operates is indeed a formidable question for its proponents.

On the other hand, no synoptic meteorologist who possesses the most fragmentary appreciation of art can deny that the flow patterns observed in mid-troposphere evolve in a manner suggestive of a vast attempt on the part of the atmosphere to arrange itself into symmetrical patterns. The currently popular concepts of the conservation of the total absolute vorticity about the vertical may be looked upon as nature's attempt to achieve symmetry. This attempt repeatedly shows up in the manner in which an upstream circulation subsequently becomes reflected some distance downstream in a manner suggestive of folding the zonal flow pattern along a meridian. The frequent failure of the atmosphere to achieve this symmetry in space is due to ever-present forces which work to destroy and change one portion of the pattern before there has been time for it to be reflected downstream.

The tendency toward symmetry described above applies to symmetry in space, whereas Weickmann's ideas are supposed to apply in time. Insofar as longitudinally symmetric patterns move steadily eastward or westward in an unchanging manner, they will produce symmetry of barograms. It may be that the method propounded by Weickmann is an attempt to capture the tendency toward spatial symmetry. If so, it would seem that it contains a germ of the truth, but this germ would be difficult to isolate in order to make diagnoses and prognoses.

Weather Type and Analogue Methods. The concepts of weather types and analogues probably arose very early in the history of meteorology. As with all systems of classification, the type method is designed to set up a certain partition of weather map sequences so that the differences between weather maps of one type are small compared to the differences between maps of different types. The work of Multanovsky [15] many years ago had indicated that the duration of a given type was not constant from year to year, and he referred to the particular length of operation of one sequence as the "natural period." But other typing systems have attempted to restrict the life history of a weather type to a certain number of days, it then being necessary for the type to change (or repeat) after this time interval. One such typing method is described at length elsewhere in this Compendium by one of the foremost exponents of the weather-type approach, R. D. Elliott.[2]

2. Consult "Extended-Range Forecasting by Weather Types" by R. D. Elliott, pp. 834–840.

The indefatigable work of Baur on the subject of types must also be mentioned, for he has probably done more work of this sort as it applies to European weather than anyone else [1]. For his types he has tabulated the seasonal frequency of occurrence, associated weather, and other statistics which are generally worked up by various typing schools.

Since both Baur and Elliott have articles describing their methods in this Compendium, we shall make only a few remarks on this general subject.

In the first place, the question of uniqueness of classification arises. Will two equally trained meteorologists give the same type designation to an observed sequence of maps, and furthermore, will they agree as to the dates of its beginning and termination? For the most prominent regional types it appears that there can be fairly good general agreement. However, there seem to be numerous "borderline" or transitional cases where there is disagreement. The dates of inception and termination of a type seem to be difficult to define. In the minds of some meteorologists, these difficulties have raised the question as to whether discrete types, with the necessary discontinuities between types, exist, or whether there is a more or less gradual transition between circulations. The final answer to this problem cannot yet be given. On the one hand the supporters of the type approach emphasize those periods in which a very rapid change of the circulation takes place and completely alters the weather regime in a short period of time, while its detractors point to many periods where the evolution of the general regional circulation is of a gradual nature for many weeks. Unquestionably the truth is composed of both these viewpoints.

When we come to the question of prognosis the lines become more sharply drawn. The exponents of weather types maintain that there exist methods of predicting the subsequent type. Statistical studies of the probability of one type following another and the relationship of subsequent types to the position and orientation of some center of action have been carried on for the purpose of prediction. Perhaps the safest method of prediction, though limited, is that used by the Multanovsky school—to wait until the first one or two days of a type have elapsed. By that time the type has been identified and one issues a forecast only for the remainder of the natural period during which the type exists. In the opinion of the present author, the weakest phase of the entire weather-type approach is the lack of any reliable methods of predicting what will be the coming type. This statement does not involve denial of the utility of weather types in diagnosis, or particularly in pedagogy, but it questions their utility for purposes of prognosis. This question was raised in the introduction to this article where the inadequacies of a regional approach, contrasted to a hemispherical approach, were cited.

The analogue method appears to crop up more or less periodically in the history of synoptic meteorology, showing perhaps a tendency to reach maximum popularity at the times of greatest stress (wartime). Wadsworth [22] was one of the principal exponents of this approach during World War II. The basic theory underlying it is entirely logical: If two maps or weather situations are found which are identical, then so are their physical properties, and therefore the subsequent evolution will be identical. Assuming no discontinuities in extraterrestrial influences, no one can take serious issue with this concept. The problem then resolves into one of finding the proper analogue in the archives of weather maps.

The proponents of the analogue method cite several advantages of its usage:

1. It is objective and does not rely upon complex and subtle reasoning inherent in physical methods which are far from perfected.
2. It is relatively fast in operation and forecasts can be turned out rapidly (indeed, analogues can now be selected by punch-card machinery).
3. Analogues automatically contain the climatological peculiarities of every location for which a forecast is required.

These three arguments are indeed potent and unquestionably have a strong appeal, particularly to administrative officials. To counter these arguments the opponents of the analogue technique point out the following:

For the analogue to be successful for a period of a few days or more the agreement between current and historical maps must be three-fold: (1) spatially similar for large areas covering a large portion of the hemisphere, (2) similar in the third dimension, and (3) similar in evolution. The necessity of considering large areas is a direct consequence of the rapid dispersion of planetary atmospheric waves indicated earlier. The other two considerations are obvious.

When these restrictions are placed upon the selection of an analogue it becomes virtually impossible to find a good one in the relatively limited archives of available weather maps. The very necessity of considering the upper-air flow patterns greatly shortens the available historical record. To this objection the proponents of the analogue method offer the rebuttal that it is possible to estimate with sufficient accuracy the appearance of the upper-air flow patterns from surface maps. They also maintain that while no analogies are ever complete, the discrepancies can be evaluated subjectively and allowance can be made for differences.

From the long-range view perhaps the most damaging blow to the analogue technique is that it is at best a substitute for understanding—and without understanding it is unlikely that meteorology can elevate itself above the low plateau of skill obtainable by analogue methods. Moreover, there is no evidence that this plateau is any higher than that of other more physical methods.

Physical Methods in the Circulation Prognosis. With the increase in upper-air data brought about largely by the progress in aviation there has come a better understanding of the general circulation. Of course the complete solution of the general circulation problem contains the key to long-range weather forecasting. Progress along these lines is described elsewhere in this

Compendium.[3,4] We shall refer to such studies only inasmuch as they have current and direct application.

The first milestone in the new aerological era was the increase of knowledge of the interrelationship between upper-level and sea-level circulations. Perhaps no one school of thought can be given principal credit for the discoveries, for many groups working along similar lines were converging to the truth. In this development, the centers of action observed at sea level became interrelated with the long waves in the middle and upper westerlies.

A clue that the movement and development of these upper waves might be predicted by physical methods was given by Rossby [17], who showed that such large-scale waves, when treated on the basis of conservation of the vertical component of total absolute vorticity, seemed to fit into a reasonable theory. This theory resulted in the now classical frequency equation from which is given the eastward speed c of a long wave in the upper tropospheric westerlies as a function of the zonal wind speed U, the wave length L, and the northward variation of the Coriolis parameter β:

$$c = U - \beta L^2/4\pi^2. \tag{1}$$

From this expression it develops that the stationary wave length L_s (*i.e.*, when $c = 0$) is a function only of U:

$$L_s = 2\pi\sqrt{U/\beta}. \tag{2}$$

In the practical application of this formula one is confronted first with the question of how the proper values of zonal speed U and the wave length L are to be computed. A great deal of work on these and other questions of application has been done at the U. S. Weather Bureau, and some of the more important results are described elsewhere in this Compendium.[4] These results, and indeed this entire article, should be considered a part of this treatment of extended-range forecasting. The principal results of these studies which relate primarily to the 700-mb level indicate that:

1. Zonal wind speeds for use in equation (1) are best obtained by computing the geostrophic wind in mid-troposphere averaged over large areas (perhaps of the order of 100° to 180° of longitude) for the latitude bands in which the waves are found. There may be, and usually are, two or more out-of-phase wave trains present in different latitude bands.

2. It is difficult to identify the primary long waves on a daily map because of the irregularities (small cyclone waves) embedded in the flow. The identification, as pointed out earlier, may be accomplished by some form of smoothing such as averaging over intervals of time, considering high tropospheric flow, or considering the patterns of isotherms associated with the contours. In the latter case those troughs and ridges in which isotherms and contours are definitely in phase are generally the primary long waves.

3. Equation (1) may then be applied by choosing a trough or ridge whose subsequent motion is desired and scaling off its distance from the next primary ridge or trough upstream.

When this is done for many cases it becomes clear that while there is reasonably good positive correlation between computed and observed values of c for periods up to three or four days, the computed motions generally give eastward speeds which are too small. For this reason empirical corrections must be developed for different areas and seasons as described elsewhere.[4] From such a coordination of theoretical and empirical data one may construct graphs (for different areas and seasons) which are useful in predicting the motion of long waves.

Another less restrictive method of using the vorticity principle in extended-range forecasting consists of constructing (on daily or mean charts) inertia trajectories based on the conservation of the vertical component of the total absolute vorticity. Here again, empirical, areal, and seasonal modifications are necessary for optimum success.

Both of the methods described above suffer from limitations imposed by the many assumptions underlying the development of the simplified vorticity equations. Among the most serious of these are perhaps the assumptions of no divergence, no solenoids, and no change in speed of air particles. It is indeed surprising that in spite of these and other restrictive assumptions, the waves behave in such good accordance with the theoretical-empirical formula. Part of this agreement must be ascribed to the quasi-barotropic nature of the atmosphere, and to fields of divergence which are normally operative in a fairly consistent manner within a given season over a given area.

However, the use of empirically derived wave-length formulas and constant vorticity trajectories leave numerous forecasting problems unsolved. One of the principal of these involves the development of new troughs and the disappearance of old ones. Wave-length expressions generally fail to distinguish between retrogression (westward motion of the long waves) and the possible development of a new trough which may increase the hemispheric wave number, thereby taking up the slack indicated by a wave length greater than the stationary wave length. On the other hand, computed vorticity trajectories may suggest a new trough development in an area currently occupied by a ridge or a flat westerly flow.

The problem of trough formation is closely allied with the process of "development" which has recently received considerable attention from British meteorologists, particularly Sutcliffe [20]. This approach makes considerable use of hemispheric thickness patterns for the layer 1000–500 mb. Naturally, these charts are similar to contour charts and thus bring to light the major long waves in the westerlies. Sutcliffe and his group attempt to apply wave-length concepts to them, at least qualitatively. But in addition, Sutcliffe has developed certain criteria for the deepening and filling of systems, formation of secondary disturbances, etc.,

3. Consult "The Physical Basis for the General Circulation" by V. P. Starr, pp. 541–550.

4. Consult "Observational Studies of General Circulation Patterns" by J. Namias and P. F. Clapp, pp. 551–567.

based upon the field of thermal vorticity indicated by the thickness lines. Further synoptic studies are being carried on by the British group to lend support to these criteria, and it seems possible that the vexing problems of trough development or disappearance may find partial solution in this work.

In the practical routine of extended-range forecasting in the U. S. Weather Bureau, the thermal field of the upper air is given careful consideration—not only in the thickness patterns themselves but also in charts showing departures from normal of both the thickness patterns and the temperatures at 700 mb. These thermal fields are also used as adjuncts in determining deepening or filling of individual long-wave systems.

These purely physical methods of prognosis are naturally far from perfected. For this reason it is necessary to use them in conjunction with the kinematic procedures described earlier. Just as with short-range forecasting, the evaluation of time derivatives becomes important. In a sense, to the extent that the tendencies are accurate, the kinematic methods represent an integration of physical processes. The arbitrary division into physical and kinematical tools is made only because it is possible to make partially successful extended-range forecasts with these kinematic methods even though the user has no idea of the physical processes involved.

Experience indicates that the highest skills are obtained by judicious combination of physical and kinematical tools. A specific example of such coordination may clarify this point. Let us say that there are two pronounced troughs in the upper-level westerlies whose wave length is about equal to the stationary wave length, but that the tendency fields suggest strongly that the troughs are spreading apart—perhaps the western one is retrograding while the eastern one is progressing eastward. Kinematic displacement computations would therefore suggest the incipience of a new trough, and the forecaster would be alerted to examine other evidence for further clues, such as computed vorticity paths, thermal fields, and cyclogenetic characteristics of the latest weather map.

The methods described above are meant to apply primarily to the construction of five-day (or longer-period) prognostic mean upper-level charts. From such upper-level prognoses the associated mean sea-level charts may be prepared by methods of differential analysis.

In the past ten years it has become increasingly clear to the author that *the most successful extended-range forecasting procedures must involve not only the current state (daily or mean) of the atmosphere, but also its manner and rate of evolution.* Methods of extended-range forecasting which do not incorporate in some manner the evolution must be very limited in scope and success.

Here we come to one of the principal dilemmas of the modern weather forecaster. How can he coordinate the tremendous mass of data in space and time necessary to incorporate evolution in a three-dimensional, globally interactive atmosphere? Yet he must do this to arrive at a regional prognosis which is internally consistent with other hemispheric features. The principal hope for the solution of this apparently superhuman task seems to lie in the development of new electronic high-speed computing machines from which may emerge at least a first approximation to the prediction.

THE RELATIONSHIP BETWEEN CIRCULATION AND WEATHER

The methods of prediction described above aim at forecasting features of the atmospheric circulation generally represented by pressure patterns at one or more levels. The tacit assumption is made that weather forecasts are by-products of the circulation prognosis. Precisely this assumption underlies the preparation and transmission of short-range prognoses as practiced around the world. While this assumption is to a considerable extent justified, controlled experiments indicate that even if forecasters were provided with perfect prognostic charts (analyses rather than prognoses) they would be unable to make perfect forecasts. In fact, in many situations forecasts of weather from such charts, whether a short- or long-range (mean) prognosis, leave much to be desired. For this reason it would appear that a disproportionate amount of research may be currently placed on improving circulation prognoses while very little is being done to interpret circulations in terms of associated weather.

This problem, equally important in longer-range work, appears to have received more attention than it has in short-range forecasting. For example, Baur [1] has made extensive statistical studies of the weather regimes associated with various large-scale circulation patterns over Europe, and similar studies for America have been carried on at the California Institute of Technology [5]. One of the most systematic studies relating pressure patterns to weather phenomena has been in progress for several years at the U. S. Weather Bureau. Here an attempt has been made to associate temperature anomalies and total precipitation observed over five-day and monthly periods to mean pressure patterns at sea level and, more especially, at 700 mb. The essence of the most promising methods for finding these relationships appears to lie primarily in the large-scale contour patterns of the lower troposphere.

Studies for different seasons and for many points over the United States indicate that regardless of how mean maps are made up (that is, the sequence and distribution of day-to-day patterns within the period), they determine to a large extent the average temperature and precipitation anomalies[5] for the period considered. This statement often comes as a surprise to many meteorologists in spite of the fact that it seems to have been recognized almost a century ago by the pioneers who discovered and studied the great centers of action. In part, the physical reality of the mean as it relates to average weather is due to the fact that events in time are serially correlated. Not only this persistence but

5. The term "anomaly," as used in this article, refers to departure from a long-term average over time and has nothing to do with the departures from means computed for latitude circles.

also a persistent recurrence of certain circulation types gives a decided character to both mean pressure patterns and associated weather phenomena. This relationship is rather fortunate in practical extended-range forecasting since it frees the forecaster from the seemingly hopeless task of giving a play-by-play description of the detail of the weather in order to make a statement of average conditions anticipated. Likewise the use of means makes it unnecessary to break up the period, if long, into a sequence of discrete types.

One rather simple technique of translating circulation patterns into weather anomalies consists of constructing isolines of departure from normal of the pressure or height. From the components of flow relative to normal, in conjunction with the distribution of normal surface isotherms or normal thickness charts of the lower troposphere, one can readily decide if a circulation is apt to bring warm or cold air to a locality and also get a rough measure of the degree of abnormality. Statistical refinements of this method as practiced in America and developed by Martin [9, 10] attempt to determine for various places and for different months the key regions where the anomaly of the contour pattern to a large extent governs the temperature anomaly. These studies bring into sharp focus the distant controls operating on apparently local weather. For example, for most stations in the eastern United States it turns out that surface temperature anomalies in winter are remarkably dependent upon the anomaly of 700-mb heights near southeastern Alaska! When the mid-tropospheric ridge in this area is strong (high positive anomaly) it sets in motion vast outbreaks of polar Canadian air which pour into eastern United States and produce low temperatures there. When the ridge is absent, a series of frontal waves move across the United States–Canadian border and a westerly flow of mild maritime Pacific air leads to warm weather over the east. These and other factors are now considered objectively so that it is possible to construct in a mechanical fashion the temperature anomaly anticipated with a given 700-mb pattern.

The relationship of precipitation to mean circulation is more difficult. Even here, however, a reasonable degree of success may be achieved by relating to the observed precipitation such factors as location and orientation of major trough and ridge systems of the mid-troposphere, anomalous components of wind, and curvature of isobars [7]. While a strictly objective method of doing this is obviously the goal, it appears to be difficult to design a system as effective as the objective temperature method.

When it comes to giving a day-by-day weather forecast beyond two or three days in advance, the current state of knowledge is indeed unsatisfactory. Most of the skill of extended-range forecasts appears to rest in the ability to predict general characteristics of a period and perhaps broad trends. The strain imposed by the requirement of a running description of detailed weather in temperate latitudes for, let us say, a week is much too great for all systems of extended-range forecasting developed so far. Consequently, the degree of skill when verified on a day-by-day basis falls off rapidly after two or three days. However, there is some encouragement in the fact that even on the sixth day some forecasts have demonstrated slightly greater success than climatological probabilities. But such forecasts can hardly be used to plan a picnic.

Despite this low level of skill an attempt is generally made to specify day-to-day conditions through the medium of the prognostic chart. In the U. S. Weather Bureau, prognostic charts for six days in advance are prepared—not with the belief that they possess striking accuracy, but rather with a view of expressing the anticipated general lines of air-mass, frontal, cyclonic, and anticyclonic activity. The prognoses elucidate further the anticipated mean state of the circulation and serve as a guide to general trends in temperature, rainfall, wind, etc. The charts are drawn by considering the mean prognosis and its steering influence, and are naturally constructed so as to be logically continuous with the short-range (*i.e.*, 30-hr) prognosis. Obviously, a healthy, vivid imagination and a firm hand on the part of the forecaster are helpful.

A more objective method of preparing such day-to-day prognoses for periods of a week or so in advance is supplied by the analogue method, for here, once the analogous period is decided upon, one can substitute (perhaps with small modification) observed maps for prognostic maps. The advantage of greater objectivity, while desirable and appealing, is unfortunately largely offset by inability to find in the archives of maps proper analogues, not only before the events, but many times even after the period is over.

Within the past few years the prognosis of mid-troposphere flow patterns for 48 or 72 hr in advance has become increasingly popular and these charts offer a starting point for a more extended prognosis. The methods for constructing these prognoses are for the most part those briefly discussed under physical methods, and involve discriminating use of vorticity principles, thermal fields, warm and cold advection, and kinematic (generally pressure-change) methods. As in the mean-map technique, there appears to be no especial reason why the forecast must be made in steps of 24-hr intervals. It would appear that synoptic meteorologists have too long considered the unit of 24 hr as mandatory in the preparation of a series of charts. The theory underlying modern extended-range forecast techniques is that the forecaster strikes out for a general state expected to be observed well ahead of 24 hr and then proceeds to determine how this general state will evolve. Experienced forecasters of all countries who have tried the step-by-step or extrapolation method of extended-range forecasting have long ago recognized that this method breaks down after about 48 hr. Probably the reason for this breakdown lies in the inability of man's brain to cope with the increasing complexity of atmospheric interactions. If and when electronic machines are used to grind out a reasonably good 24-hr mid-troposphere prognosis, meteorologists will have a good way of determining how far the step-by-step method of long-range prognosis can proceed.

THE EXTENSION OF FORECASTS TO LONGER PERIODS

The basis of the preparation of most medium-range forecasts is essentially synoptic in nature, although there is a substantially greater infiltration of statistical treatment than in short-range forecasting. As longer periods are considered, there appears to be greater and greater reliance upon statistical techniques at the expense of synoptic techniques. Yet the possible extension of synoptic techniques to forecasting problems for periods longer than a week has been explored very little. In view of the fact that some statistically significant success has been achieved in applying synoptic methods to monthly mean maps (described below) it would appear that meteorologists, particularly dynamic meteorologists, might be able to make rewarding contributions by directing some of their efforts to this problem.

As stated earlier, this article is concerned only with those long-range forecasting methods which can demonstrate skill in the rigid statistical sense over reasonably long periods of time. This means consistently predicting occurrences with greater success than the climatic probabilities. This criterion, while harsh, is after all the only rigid means of differentiating between a successfully operating forecasting method and a theory. When the field of long-range forecasting methods is scanned through this revealing eyepiece, there appears surprisingly little that can be discussed. In part, this reflects the unavailability of rigid statistical verifications of forecasts, but more probably it indicates the overwhelming complexity of the long-range forecast problem.

Public claims of success in long-range forecasting should be backed by statistically sound verification. It is high time that professional societies bring the spotlight upon those whose claims cannot be substantiated. Meteorology, no less than medicine, can ill afford to harbor quacks. This is not meant to detract from the painstaking work of those who have spent years, if not decades, in the quest for forecasting tools. Their work must be judged on other grounds than its immediate application to forecasting problems.

Before closing this article the author feels compelled to dwell a bit further on the question of the extension of synoptic-statistical techniques to longer-range forecasting. In 1942 he began an experiment in the preparation of thirty-day forecasts, the aim of which was to explore the possibility of extending mean circulation methods then used in five-day forecasting to monthly periods. Of course, it was soon recognized that the normal state of the circulation and its rate of change become increasingly important in such work. In essence the problem was found to be similar to the shorter (five-day) problem. At the start two highly pertinent questions arose:

1. Do the thirty-day mean circulation patterns determine the average weather conditions (anomalies) for the month? The answer is an unqualified yes. The scientific basis for this answer has been briefly discussed and is expanded upon elsewhere [7, 10, 13].

2. Is there a rational continuity to the great centers of action when treated on a thirty-day mean basis? Here again, experience gained during the last decade, in part reported on in the literature [8, 12, 13], indicates that there is a rational continuity when thirty-day charts are studied carefully with the aid of kinematical and physical techniques. This continuity is discussed twice monthly in a routine fashion by workers in the Extended Forecast Section of the U. S. Weather Bureau with as much conviction as are the movements of cyclones and anticyclones at daily map discussions.

The detailed method for projecting the centers of action into the future of the next thirty days is naturally too involved to be discussed here, but its basis may be briefly stated this way:

1. The past and current rate of evolution of mid-tropospheric flow patterns is assessed with the aid of the tendency method described earlier. These methods automatically introduce the effect of normals changing with time.

2. After kinematic displacements of the major features of the patterns (ridges, troughs, and centers) are made, it must be decided if these kinematic displacements are in harmony with vorticity (thus wave-length) principles. In other words, within certain speeds of the mid-troposphere westerlies the atmosphere prefers a certain spacing between ridges and troughs. If kinematically computed displacements introduce no severe contradictions into these physically harmonious upper-air patterns, the forecast is relatively straightforward. If contradictions arise—as is often the case—certain large-scale features, whose immediate behavior is more clearly outlined in their profiles and tendency fields, must be chosen as anchor points and the adjacent portions of the forecast molded into harmony with these.

Meteorologists can, and usually do, immediately raise embarrassing questions regarding this technique. In the first place, what reasons are there for applying physical reasoning (vorticity principles), designed essentially for instantaneous daily maps, to maps covering periods as long as thirty days? No satisfactory answer to this question can be given at present. However, it should be emphasized that the thirty-day mean is not composed of randomly distributed data. The daily circulations of which it is composed are themselves serially correlated and, moreover, during one month there are generally repetitive circulation types. Thus in a given area a circulation, which on a daily map introduces a certain field of vorticity attempting to modify its surroundings, recurs again and again during the month to perform a similar modifying function. Is it surprising, then, that when the mean chart is considered from a vorticity standpoint it, too, seems to behave as a physical entity? Another consideration involves an analogy to the statistical theory of turbulence. Here the equations of motion for laminar flow can be applied to an average fluid flow on which are superimposed random eddies, provided certain frictional stress terms are added. In the work with thirty-day means, these additional terms are, of course, not evaluated or known, but their omission may not be as serious as at first thought. It should be pointed out that they are

also omitted in working with daily charts which, in reality, are also mean charts, but for brief time intervals.

But these arguments can be of little avail until definite solutions, theoretical or empirical, are found. Ultimately empiricism and theory must converge to truth. It seems probable that in long-range forecasting, as in hydrodynamics, the most rewarding line of attack will be a happy combination of theory and empiricism.

By way of stimulating dynamic meteorologists to study the problem of evolution of mean states of the general circulation over periods as long as a month, the author desires to point out some indicative results of a statistical verification of thirty-day forecasts made twice monthly during the past five years and summarized by seasons.

1. The average correlation coefficients between forecast and observed patterns of *anomalies* of the mean monthly 700-mb contours have been positive during 19 out of 20 seasons over North America.

2. The forecasts of monthly temperature anomaly over the United States (verified at 100 points) have been superior to climatological probability forecasts for 19 out of 20 seasons.

3. The forecasts of total monthly precipitation over the United States (verified at 100 points) have been superior to climatological probability forecasts during 17 out of 20 seasons.

THE COURSE AHEAD

Viewed pessimistically, the present status of extended-range weather forecasting may appear to be hopelessly distant from perfection. It is in part this attitude which leads to the quest for stop-gap solutions. From the optimistic viewpoint, it appears that during the last decade notable advances have been made in understanding as well as in forecasting the general circulation and its weather. These advances lie principally in interrelations between the great centers of action and have been made largely through study of mid-tropospheric patterns. This progress holds out the hope that forecast skill may be on the lower branch of an exponentially increasing curve which may subsequently turn into an asymptotic approach to perfection.

Indeed, the problem of extended-range forecasting may be looked upon as the ultimate test of a complete theory of the general circulation. For this reason remarks made elsewhere in this Compendium pertaining to the avenues of improvement of knowledge of the general circulation are directly applicable to the subject of improvement in extended-range forecasting.

In addition to completing basic studies, meteorologists the world over would do well to improve the position of extended-range forecasting by consolidating their gains of the past score of years. Along this line, it would be desirable for those engaged in forecasting practice or research to make a greater effort to study and apply methods developed in other countries and to relate them to locally practiced techniques. This is a plea for greater international collaboration.

Another less basic but highly practical problem is a better method of representation of the tremendous wealth of data apparently necessary to make long-range forecasts. Similar questions involving time and space derivatives also arise in short-range forecasting.

Perhaps a healthier state of cooperation more conducive to progress could be developed between the so-called dynamic meteorologists and the practicing forecasters and empirical researchers. The ultimate goal of these groups is essentially the same—prediction, or understanding which leads to prediction. For this reason it is hard to reconcile attitudes of some dynamic meteorologists that application (*i.e.*, forecasting) is not their concern and that empirical, including statistical, knowledge is of little relevance if it does not fit certain classical concepts. Likewise, it is equally difficult to understand the philosophy of that group of practicing forecasters who look upon all theoretical work as completely removed from everyday problems of weather forecasting.

Given a prevailingly peaceful state of world affairs permitting international collaboration, the outlook for long-range weather forecasting appears bright. In the opinion of the author, an incurable optimist, not only an increase in accuracy but also in the time range of general forecasts is probable. Forecasts for at least a season in advance and possibly a decade are within the grasp of the present generation.

REFERENCES

1. Baur, F., *Musterbeispiele europäischer Grosswetterlagen.* Wiesbaden, Dieterich, 1947.
2. —— *Einführung in die Grosswetterkunde.* Wiesbaden, Dieterich, 1948.
3. Bjerknes, J., "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54:462–466 (1937).
4. Brooks, C. E. P., "Annual Recurrences of Weather: Singularities." *Weather*, 1:107–113, 130–134 (1946).
5. California Institute of Technology, Dept. Meteor., *Synoptic Weather Types of North America.* Pasadena, Calif., 1943.
6. Haurwitz, B., "Final Report on the Use of Symmetry Points in the Pressure Curves for Long-Range Forecasts." U. S. Air Force, *Air Wea. Serv. Tech. Rep.* No. 105–7 (1944).
7. Klein, W. H., "Winter Precipitation as Related to the 700-Mb Circulation." *Bull. Amer. meteor. Soc.*, 29:439–453 (1948).
8. —— "The Unusual Weather and Circulation of the 1948–49 Winter." *Mon. Wea. Rev. Wash.*, 77:99–113 (1949).
9. Martin, D. E., and Hawkins, H. F., Jr., "Forecasting the Weather: The Relationship of Temperature and Precipitation over the United States to the Circulation Aloft." *Weatherwise*, 3:16–19, 40–43, 65–67, 89–92, 113–116, 138–141 (1950).
10. Martin, D. E., and Leight, W. G., "Objective Temperature Estimates from Mean Circulation Patterns." *Mon. Wea. Rev. Wash.*, 77:275–283 (1949).
11. Namias, J., *Extended Forecasting by Mean Circulation Methods.* U. S. Weather Bureau, Washington, D. C., 1947.
12. —— "Characteristics of the General Circulation over the Northern Hemisphere during the Abnormal Winter 1946–47." *Mon. Wea. Rev. Wash.*, 75:145–152 (1947).
13. —— "Evolution of Monthly Mean Circulation and Weather Patterns." *Trans. Amer. geophys. Un.*, 29:777–788 (1948).

14. —— and Clapp, P. F., "Studies of the Motion and Development of Long Waves in the Westerlies." *J. Meteor.*, 1:57–77 (1944).
15. Pagava, S. T., and others, *Basic Principles of the Synoptic Method of Long-Range Weather Forecasting*, 3 Vols. Leningrad, U.S.S.R. Hydrometeorological Publ. House, 1940. (Translation prepared by Weather Information Service, Headquarters, Army Air Forces.)
16. Petterssen, S., *Weather Analysis and Forecasting*. New York, McGraw, 1940.
17. Rossby, C.-G., and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
18. Rossby, C.-G., "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2:187–204 (1945).
19. Schmauss, A., "Synoptische Singularitäten," *Meteor. Z.*, 55:385–403 (1938).
20. Sutcliffe, R. C., "A Contribution to the Problem of Development." *Quart. J. R. meteor. Soc.*, 73:370–383 (1947).
21. —— and Forsdyke, A. G., "The Theory and Use of Upper Air Thickness Patterns in Forecasting." *Quart. J. R. meteor. Soc.*, 76:189–217 (1950).
22. Wadsworth, G. P., "Short Range and Extended Forecasting by Statistical Methods." U. S. Air Force, *Air Wea. Serv. Tech. Rep.* No. 105–38, 202 pp. (1948).
23. Walker, Sir Gilbert T., "Symmetry Points in Pressure as Aids to Forecasting." *Quart. J. R. meteor. Soc.*, 72: 265–283 (1946).
24. Weickmann, L., "Wellen im Luftmeer, I. Mitt." *Abh. sächs. Ges. (Akad.) Wiss.*, Bd. 39, No. 2, 46 SS. (1924).
25. —— "Neuere Ergebnisse aus der Theorie der Symmetriepunkte." *Beitr. Geophys.*, 34:244–251 (1931).

EXTENDED-RANGE WEATHER FORECASTING*

By FRANZ BAUR

University of Frankfurt am Main

SCIENTIFIC BASES OF EXTENDED-RANGE WEATHER FORECASTING

The Basic Problems of Macrometeorology. *The First Basic Problem.* Macrometeorology forms the scientific basis for extended-range forecasting. The first basic problem of this new branch of meteorology is whether or not a real *Grosswetter* exists at all. In other words, can the observed longer periods of persistent cold or warm, dry or wet weather be attributed to some major variable influences, or are they merely the consequence of the so-called "persistence tendency" in connection with random developments and the annual variation of meteorological elements? Among the older school of meteorologists, there are, at least in Europe, those who are convinced that *Grosswetter* takes its course according to the principle: "small causes, large effects." They try, for instance, to explain the development of a severe winter by the fact that clearing takes place at the beginning of the winter after the first widespread snowstorm, thus causing the temperature to drop considerably. Hence the cold air is "maintained" so that with the next upgliding of warm air, a new snowfall will occur and thus the wintry cold will gradually be amplified.

Their reasoning is as follows: If it had "accidentally" been just two degrees warmer on that first day of winter, there would have been rain instead of snow. Thus the radiative heat loss during the following night would not have been so great. Less cold air would have formed and, finally, the winter would have been less severe. However, the following reasoning can likewise be applied to this case: During every winter in the temperate zone, there will be one case of clearing after a snowstorm. Whether the winter following a snowfall with subsequent clearing will, in the long run, turn out to be severe or mild depends on whether the probability of the occurrence of cold days is larger or smaller than normal as a consequence of governing influences. If there should be a severe cold wave, there are nevertheless sufficient possibilities for the cold to recede after some time, when there exist no governing influences that would cause a tendency toward new cold outbreaks and thus a greater probability of cold days.

The winter of 1876–77 in central Europe is an excellent example, showing that even after very severe winter days mild weather can recur even before the end of the winter season. The month of December 1876 started with especially mild weather; the departure of the temperature from a well-established average was +4.9C for the first half of December in Berlin. Then cooling, snowfall, and clearing occurred, and the daily mean temperature for December 24 dropped to −15.8C (corresponding to a departure from normal of −16C), and it remained below −10C until December 27. However, the daily mean temperature of December 30 was again 6.4C above normal, the month of January was 3.5C above normal, and the month of February was 2.2C above normal.

Such isolated examples, however, are not sufficient to enable one to decide whether *Grosswetter*, in the sense mentioned above, exists in reality. Rather, it is necessary to establish statistically from extensive observational data whether or not the fundamental probabilities of the occurrence of the meteorological phenomena that determine weather are (aside from the annual trend) constant or subject to variations.

Lexis' theorem [36] serves as the method of finding fluctuations of the fundamental probability. If the probability p of the occurrence of a phenomenon during an arbitrary number n of test series (each consisting of s independent trials) remains constant, the resulting distribution of the frequencies $m_1, m_2, \ldots m_n$ of the occurrence of the phenomenon in s trials is a so-called "Bernoulli distribution," whose standard deviation is $\sigma_B = \sqrt{sp(1-p)}$. If, however, the probability of the occurrence of the phenomenon remains constant from test to test within a series, but varies *from series to series*, then the observed standard deviation σ_h is larger than σ_B. In this case, the quotient $Q_1 = \sigma_h/\sigma_B$, known as *Lexis' ratio*, is greater than unity. If, however, the probability p varies *from test to test*, whereas $p_0 = (p_1 + p_2 + \ldots + p_s)/s$ remains constant from series to series, then $Q_1 < 1$. Lexis' ratio, therefore, is a criterion for determining whether or not the fundamental probability is subject to variations.

In its original form, however, Lexis' theory cannot be applied to meteorological observations, since in these —owing to their persistence tendency—the condition of the independence of consecutive terms is not fulfilled. Baur [12], therefore, introduced a new criterion, the "divergence coefficient of the second kind" which is defined by the quotient $D_1 = \sigma_h/\sigma_M$, where Markoff's standard deviation σ_M is used instead of Bernoulli's. By σ_M we mean the standard deviation of the distribution $w_n(x)$, where $w_n(x)$ designates the probability with which the phenomenon having the (constant) fundamental probability p will occur x times among the first n observations of a Markoff alternating chain (*i.e.*, an alternating series where each term depends on the preceding term). If the fundamental probability remains unchanged from series to series and from test to test, the mathematical expectation of D_1 again equals unity. If, however, the fundamental probability varies from series to series, but remains unchanged from test to test, then $D_1 > 1$.

* Translated from the original German.

Instead of counting the frequency of the occurrence in each series consisting of s tests, it is possible to determine the frequency of the occurrence in the first n tests, in the second n tests, etc., of all series, thus obtaining the standard deviation σ_H, as well as $Q_2 = \sigma_H/\sigma_B$ and $D_2 = \sigma_H/\sigma_M$, from the resulting distribution. If the probability p of the occurrence of a phenomenon varies systematically from test to test such that, for instance, the probability p of the third and tenth tests in each series is larger than that of the other tests, which would, meteorologically, correspond to an existence of true "singularities" according to Schmauss [50], then $Q_1/Q_2 = D_1/D_2 < 1$.

A study has been undertaken [20, pp. 924–928] involving nineteen series of meteorological observations of various elements, made at different locations (in different countries of the temperate zone) and during different seasons. Observations from January 1 to February 19 and from July 1 to August 19 for the fifty years from 1888 to 1937 (where $s = n = 50$) were used. The results showed that $D_1 > 1$ for all these series without exception (D is considerably greater than 1 in most cases) and Q_1/Q_2 was much larger than 1, the maximum of this ratio being 3.46. Two theorems follow unequivocally:

First Empirical Theorem: *A* Grosswetter *exists, in which there are governing complexes comprising variable conditions; because of these complexes, the probabilities of the occurrence of certain indices that characterize the weather for longer periods vary from year to year.*

Second Empirical Theorem: *These fluctuations of the weather-forming probabilities are, apart from the large annual trend, much more significant for the weather character than are the calendar probabilities.*

The first of these two theorems, that concerning the existence of *Grosswetter*, forms the logical foundation for the exploration of the problem of extended-range weather forecasting. If there were no governing complexes comprising variable conditions, and the *Grosswetter* evolved by chance and by a persistence tendency, then a reliable extended-range weather forecast would be impossible for all time, and a detailed study of this problem would be senseless.

The Second Basic Problem of Macrometeorology. Statistical evidence has been obtained that the fundamental probabilities of the meteorological elements change from year to year, and that there must be governing complexes comprising variable conditions which produce these variations of the fundamental probabilities. This evidence leads to the second problem: What are these governing complexes? Are they of terrestrial or extraterrestrial origin?

Whereas we shall see the various terrestrial influences on *Grosswetter*, we shall find that these influences are not sufficient to explain the large fluctuations of the probabilities of the occurrence of the indices that characterize the weather for longer periods of time. Therefore, we must look outside the earth for the governing complexes comprising variable conditions. This is now only an assumption. Evidence of the correctness of this assumption is found in the section on cosmic influences (p. 819).

Terrestrial Influences on *Grosswetter*. *Volcanic Eruptions.* The investigations by Abbot and Fowle [2], Humphreys [34], and Ångström [4] have shown conclusively that after violent volcanic eruptions the atmospheric transparency for solar radiation is decreased over wide regions. After the Katmai eruption in 1912, the direct solar radiation reaching the earth's surface was reduced by more than 10 per cent for several months. Furthermore, there is no doubt that, as a consequence of the decrease of insolation, the mean temperature for large regions is decreased by 0.5–1.0 centigrade degrees; and by as much as 2–3 centigrade degrees after especially violent eruptions.

According to Defant [31], the atmospheric circulation over the North Atlantic is disturbed through violent volcanic eruptions to an oscillation with a period of approximately 3½ yr, with a decrease of the meridional pressure gradient after the eruption, and an increase in the following two years. The disturbance, however, is rapidly damped.

These and similar examples cannot be considered the basis for extended-range forecasting. At best, they serve as a warning against the application of statistical results obtained in years without eruptions to a year during which there will be an eruption. At any rate, major volcanic eruptions that influence the general weather are too rare to provide an explanation for the fluctuations of the fundamental probabilities of the weather elements, mentioned in the opening section.

Ocean Currents and Ice Conditions. The number of investigations that have been made regarding the relationship between *Grosswetter* and preceding anomalies[1] of ocean currents and ice conditions is so great that it is impossible to mention all of them here. Most of these investigations were made in Europe; the reason for this is that the Gulf Stream, owing to the orientation of the west coast of Europe (from southwest to northeast), is much more important for the climate of western Europe than is the Kuroshio for North America. Furthermore, it is evident that the occurrence of ice in the North Atlantic plays an important role in the climate of Europe, since a large portion of the ocean ice of the north polar basin melts in the summer and is transported into the North Atlantic through the region between the northeastern tip of Greenland and Spitsbergen, as well as through the Davis Strait. By contrast, according to Schott [51], no ice passes from the polar basin through the Bering Strait into the Pacific Ocean.

It was initially hoped that a study of the influences of temperature and velocity anomalies in the Gulf Stream on European weather, as well as of the occurrence of ice near Iceland, might yield clues for extended-range weather forecasting. However, these hopes were not fulfilled. This statement is corroborated by the follow-

1. The term "anomaly," as used in this article, refers to departure from a long-term average over time and has nothing to do with departures from means computed for latitude circles.

ing examples: According to Pettersson [43], the temperature variations of the Gulf Stream are faithfully reflected in the simultaneous variations of the air temperature on the west coast of Scandinavia. Furthermore, Meinardus [38] showed that, during the years 1862–97, in 92 per cent of the cases the mean temperature for February and March in Berlin changed, relative to the preceding year, in the same direction as the mean temperature for the previous November, December, and January in Kristiansund had changed relative to the preceding year. This result seemed, indeed, useful for prognostic purposes and was mentioned to that effect in textbooks for several decades. However, a new investigation [11] showed that the correlation coefficient between these quantities was +0.73 for the period 1862–90, but was only −0.30 for the subsequent period 1891–1920.

For many years J. W. Sandström was of the opinion that a warm Gulf Stream would bring a mild winter to northern Europe. In the year 1939, however, the Gulf Stream temperatures reached a record height, but northern and central Europe experienced the coldest winter in 110 years.

A relationship between ice conditions and the general weather, particularly in northern Russia, has been shown by Wiese in several papers [56]. Undoubtedly, ice conditions and ocean currents are interrelated with the general atmospheric circulation in many ways. However, these interrelations exist mainly in the form of simultaneous correlations. The significant correlation coefficients between ice conditions and meteorological elements in subsequent time intervals, that were computed from long series of observations, were, at best, of the magnitude $|r| = 0.5$. Thus, the ocean currents and the ice conditions are never decisive, but only of contributory significance.

Motions of the Pole. The quasi-periodical displacements of the rotational axis of the earth which have been found by observations of polar motion are composed of a period of approximately 434 days (Chandler's period), caused by the unequal mass distribution of the solid terrestrial body, and of an irregular annual period caused by the air-mass displacements in the course of the seasons. The displacements of the rotational axis cause, in turn, changes in the centrifugal force and thereby a transport of air masses. These displacements of the earth's axis, however, are only very small. On the average, the amplitude of the oscillation amounts to 0.2–0.3 seconds of arc, 0.5 seconds of arc at the most. Thus, the pole deviates only about 9 m from its mean position. It has recently been found that over a forty-year period, a high pressure over northern Sweden was followed 434 days later by another high pressure with a relative frequency of 59 per cent [17, pp. 72–73]. Although this frequency, which was derived from 382 cases, exceeds the limit of chance, it is nevertheless much too small to be considered of practical importance for extended-range weather forecasting.

The Problem of Rhythms. *Pressure Waves.* The question regarding the existence of periods other than the well-known daily and annual periods of weather phenomena is of great importance to extended-range weather forecasting. It was hoped, above all, that periods or waves in the fluctuations of atmospheric pressure would be found, which, when extrapolated, would permit the computation of the future pressure trend for a given location. By using this process for several stations, one would be able to calculate the development of pressure patterns, at least for a limited number of days in advance. When V. Bjerknes and his co-workers [28, 30] showed that waves, subsequently developing into cyclones, can form along surfaces of discontinuity in the troposphere, a mathematical-physical basis was established for the problem of atmospheric pressure waves. Others, Exner [33, pp. 383–388] for instance, consider the thermal contrast of continents and oceans as a cause of the formation of forced oscillations. At any rate, regardless of the explanation offered for the formation of pressure waves, they can exist for a longer period of time only if they coincide approximately with the free oscillations of the atmosphere. Progress has also been made in the exploration of free atmospheric oscillations during the last two decades. While V. Bjerknes [30] neglected the vertical acceleration (quasi-static method) and considered the earth's surface as plane, Wünsche [58] treated the free oscillations of a compressible medium as a spatial problem in polar coordinates, and thus neglected neither speed nor acceleration in a vertical direction. However, his computations were still based on the assumption that the oscillations occur as isothermal changes of state in an isothermal atmosphere. This assumption naturally has an influence on the lengths of the resulting periods.

The short-period waves to which the cyclones owe their formation are of no importance for extended-range forecasting. It would be erroneous to attempt computation of the pressure distribution for the next two days from extrapolation of these short-period waves, since this can be achieved more easily by the methods of synoptic analysis. The assumption of J. Bjerknes and H. Solberg [29] that the rhythm of pressure changes is the result of four cyclone families circulating with more or less constant velocity around the pole has been disproven by observational facts. As regards the longer waves, there is still no convincing proof of their existence. During World War II, more than one thousand analyses of rhythms were made at the *Forschungsinstitut für langfristige Witterungsvorhersage* in Bad Homburg, Germany. Subsequent extrapolation of those resultant waves that were recognized as quasi-persistent on the basis of mathematical criteria, permitted prediction of the direction (sign) of atmospheric pressure changes to a definite date. These predictions were accurate more frequently than would have been possible by chance. However, the relative frequency of correct forecasts remained below 75 per cent and is therefore inadequate for practical purposes. If we determine how often in such analyses the individual trial waves prove to be quasi-persistent, we find frequency maxima for European stations at 7 and 8, 12 and 13, 20, 22 and 24, as well as 30 and 33

days. However, even the more than one thousand analyses are not sufficient to permit the definite statement that the frequency of a single wave exceeds the maximum limit of chance [20, pp. 933–934].

Periods of Several Years. Of the many oscillations with periods of several years that have been claimed in the past, only a small number have proved to be "probably real." They are a 2.2-yr period, a 3- to 3½-yr period, and an approximately 7-yr period. All these periods, however, are nonpersistent, in other words, they cease intermittently. With respect to temperature and pressure, the 3- to 3½-yr period is most marked in the region of the Malayan Archipelago and northern Australia. This period is probably a "free oscillation" (in a broader sense) between a tropical low-pressure area (India-northern Australia) and a subtropical high (South Pacific anticyclone near Easter Island), caused by the lagging of the temperature anomalies with respect to the pressure anomalies as a consequence of the inertia of the participating ocean currents. According to Berlage [25], this oscillation is propagated from the Malayan region over both hemispheres to the subpolar pressure troughs. The cause for the excitation of the Indian-Pacific oscillation has not yet been determined.

The oscillations with periods of 2.2 yr and 7 yr are probably free oscillations of the atmospheric circulation over the North Atlantic. These oscillations are similar to the 3-yr period found in the Malayan region. Whereas the 2.2-yr fluctuation is restricted to the North Atlantic and Europe, the 7-yr period has a world-wide distribution [20, pp. 937–940].

The 3-yr fluctuation in the Malayan region occurs so clearly at times that it can be obtained directly from the curve of the semiannual pressure averages. However, on the whole, these fluctuations of several years' period have such a small amplitude that they can be extracted from the observations only by means of mathematical processes (*e.g.*, periodogram analysis).

From periodogram analyses, a widespread 5- to 5½-yr oscillation has been found [17, pp. 93–96]. In reality, however, this oscillation has a variable period which only on the average amounts to 5½ yr and is identical with the double oscillation of the large-scale weather during a sunspot cycle (see section on cosmic influences).

Periodic Fluctuations of Climate. Rhythmic fluctuations of the meteorological elements having a period of more than 30 yr are called "climatic periods." For many decades, the existence of a 35-yr period was presented in all meteorological textbooks, and was known as the "Brückner cycle." This period is a classic example of the deceptions to which one may succumb by determining periods with primitive research methods and without adequate statistical tests for significance. It is to Wagner's credit that he proved [54] that the 35-yr period does not exist as a natural phenomenon, but was introduced by the methods of computation and smoothing used by Brückner.

For the demonstration of still longer periods, the data available today are insufficient.

Correlations between Consecutive Weather Anomalies. *Persistence and Repetition Tendencies.* The persistence tendency of the weather is caused by the fact that, for physical reasons, certain weather anomalies (*e.g.*, very high pressure or cold air) extending over wide regions, cannot disappear from one day to the next. However, this persistence tendency in its proper sense extends over a period of only a few days. When a certain anomaly in ten-day or monthly averages is found to be generally followed by one of the same sign more frequently than by one of opposite sign, it exhibits, fundamentally, not a persistence tendency but rather a repetition tendency, that is, the tendency of a situation which has existed for several days to be reinstated after a brief interruption. Important as the repetition tendency may be for the explanation of some phenomena, its importance must not be overrated. In the temperate zone, the repetition tendency is subject to strong local and seasonal differences. The dependence of the repetition tendency on the seasons is shown in Fig. 1.

Statistical determination of the local and seasonal differences of persistence and repetition tendencies and their physical explanation are indirectly of importance for extended-range weather forecasting. In the past, the persistence tendency of certain weather elements, such as pressure or temperature, was usually investigated by counting the frequency of sequences of the same sign or by computing correlation coefficients. Instead, the persistence and repetition tendencies of *Grosswetterlagen* should be subject to investigations in which statistics and synoptics supplement each other.

During certain parts of the year and in some regions, the probability of weather changes is, under certain premises, greater than is the persistence tendency during other seasons. As can be seen from Fig. 1*a*, such a tendency toward a major change in the synoptic situation exists in Europe in the last third of June. Connected with this tendency is the fact that for fourteen years (of the 102-yr period 1848–1949), in which the temperature of the first half of June in Berlin was more than 2C above normal, the following midsummer (July and August) was for the most part abnormally cool in central Europe, and that in thirteen (93 per cent) of these years the precipitation was more than 15 mm above normal (Table I). Since the fundamental probability of a positive precipitation departure of more than 15 mm in midsummer is 38 per cent, the maximal chance limit of the relative frequency of such midsummers for a random choice of 14 cases is 81 per cent if, as is customary, the limit of the range of chance is so chosen that the probability of exceeding this limit is $\epsilon = 0.0027$. The observed relative frequency of 93 per cent of wet midsummers is, therefore, significantly above chance.

On the other hand, when the temperature during the first half of July is 2C above normal in central Europe, there follows in about three-fourths of the cases a warm and dry late summer (August and September). This varying repetition tendency can probably be explained physically as follows: In June, when the subtropical high-pressure belt over the North Atlantic

lies still farther to the south, and the daily insolation is still increasing, an overheating of central Europe causes an irreversible pressure fall over the continent and a corresponding pressure rise over the North Atlantic. In July, however, when the Azores High is farther to the north and the daily insolation is decreasing, the pressure fall caused by overheating is no longer irreversible and, during the following August and September, it is possible for high-pressure cells to split off repeatedly from the Azores High and to migrate toward central Europe.

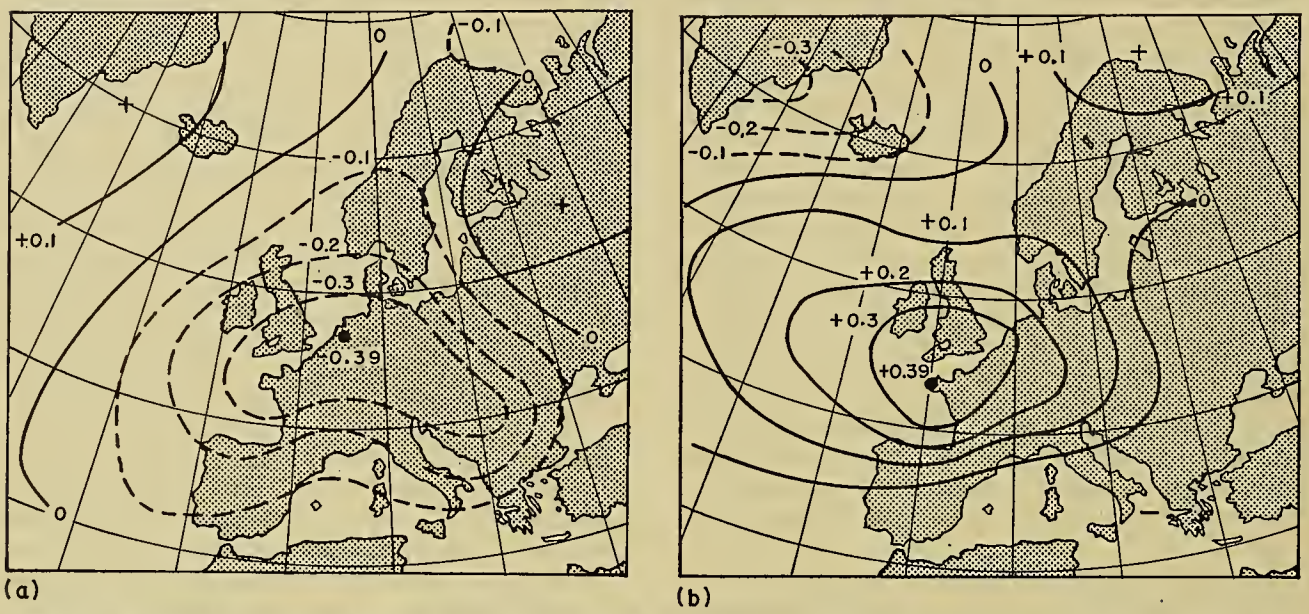


FIG. 1.—Isopleths of the correlation coefficients between the mean barometric pressure of the preceding 10-day period for various places in Europe and the mean barometric pressure at Potsdam for the following 10-day period for the years 1893–1932 (number of cases $N = 40 \times 10 = 400$), (*a*) when the ten preceding 10-day periods end on the days between June 15 and June 24, (*b*) when the ten preceding 10-day periods end on the days between July 26 and August 4.

TABLE I. LARGE POSITIVE TEMPERATURE ANOMALIES IN BERLIN IN JUNE AND SUBSEQUENT PRECIPITATION AND TEMPERATURE ANOMALIES IN CENTRAL EUROPE

Year	Temperature departure in Berlin for June 1–15 (cent. deg.)	Precipitation departure in central Europe for the following midsummer (average of 16 stations) (mm)	Temperature departure in central Europe for the following midsummer (cent. deg.)
1855	+3.1	+34	−0.1
1858	+4.5	+43	−0.2
1866	+3.0	+26	−1.5
1877	+3.3	+42	+0.4
1889	+6.5	+16	−0.9
1896	+3.3	+25	−0.7
1897	+2.1	+22	0.0
1910	+5.2	+44	−1.3
1915	+3.4	+17	−1.4
1917	+2.7	+16	+0.2
1930	+3.1	+60	−0.6
1937	+4.1	−16	+0.2
1940	+2.8	+43	−1.5
1948	+3.9	+45	−0.4

Also of importance for the physical explanation is the fact that in many cases a repetition or change tendency exists only for anomalies of one sign. During the period 1848–1947 for instance, when the temperature in the first half of December in central Europe was more than 3C above normal, the following midwinter (January and February) was too warm in all 13 cases. However, if the first half of December was below normal by just as many degrees, the following midwinter was too cold in only 4 out of 14 cases; in 10 cases, the midwinter was warmer than normal.

The considerations of this section can be summarized as follows:

THIRD EMPIRICAL THEOREM: *The repetition tendency of weather anomalies varies with the seasons and, as a rule, is not the same for positive and negative anomalies.*

Correlations between Successive Weather Anomalies in Distant Regions. A large part of the nonsimultaneous correlations of weather anomalies in distant regions can be traced to the persistence tendency of the general atmospheric circulation, as has been shown by Schell [46]. However, these correlations are, for the larger part, significant only in the tropics and subtropics and in the Southern Hemisphere. In the temperate zone of the Northern Hemisphere only a few exceed the limit of chance. Here, those correlations that point toward a *change* of the circulation are more important; in this respect, the "general circulation" is less important as a rule than the "special large-scale circulations" into which the general circulation of the Northern Hemisphere resolves (see p. 825).

Particularly, the atmospheric circulation over the North Atlantic is, to a great extent, a closed system that follows certain rules of its own (see p. 817). Generally, the pressure gradient between the Azores and Iceland determines the strength of the zonal circulation over the North Atlantic Ocean. According to Baur [7], however, the preceding zonal pressure gra-

dients as well as the temperature difference between the eastern and western parts of the circulation system are decisive for the mean pressure changes that are observed over the Azores and Iceland during periods of approximately a month's duration. Thus, from these pressure gradients and temperature contrasts, at least for the spring, the change of the pressure difference between the Azores and Iceland can be computed from one month to the next with an error of approximately six-tenths of the standard deviation. These correlations are presented in Table II, in which the correlation coefficients that exceed the upper limit of chance are italicized.

Cosmic Influences on *Grosswetter.* *Unimportance of Moon and Planets.* No one has ever proved in a single case to date that during or after a given arrangement of the moon or the planets any weather phenomenon occurred more frequently or less frequently than would have been expected from chance. On the contrary, proof has been established from extensive data and by means of statistical criteria [14; 17, pp. 93–96] that certain coincidences claimed by "astro-meteorologists" lie well within the range of chance.

The influence of the moon upon the atmosphere is restricted to producing atmospheric tides, which represent diurnal pressure variations that amount to only $\frac{1}{1000}$ of the changes associated with weather developments. Likewise, there is no notable influence of the planets on terrestrial weather. The hypothesis by See, according to which the planets are supposed to influence the sun's activity by their power of attraction upon a hypothetical stream of matter directed toward the sun, is no longer tenable. According to Waldmeier [55], each sunspot cycle, from one sunspot minimum to the next, represents a separate phenomenon that can be caused only by processes in the sun's interior (eruption hypothesis).

Relationships between Grosswetter *and Sunspot Cycles.* Köppen was the first to show the existence of an average period of eleven years in the air temperature. However, according to more recent investigations [3; 5; 17, pp. 97–105], no significant parallelism or antiparallelism of any one meteorological element to the variation of sunspots exists in the temperate zone, and even in the tropics the relationship is generally quite poor. There is a better-than-chance correlation only between the fluctuations of the level of Lake Victoria in East Africa and sunspots, but even this correlation is not persistent. The correlation coefficients between the annual means of relative sunspot numbers and the annual means of the water level of Lake Victoria (mean of maximum and minimum readings at Kisumu according to *World Weather Records*) are:

+ 0.87 for 1899–1924,
+ 0.07 for 1925–1943,
+ 0.58 for 1899–1943.

The utilization of these weak correlations for prognostic purposes does not seem feasible. However, several investigations [9, 18, 19] have revealed that over the whole earth, particularly in both temperate zones, a double, sometimes a triple, fluctuation of nearly all meteorological elements occurs within a sunspot cycle. The world-wide distribution of this fluctuation is shown by the curves of Fig. 2. The bottom curve represents the dependence of the departure of the mean annual temperatures in Apia, Samoa, and Colombo, Ceylon, on the year's position in the sunspot cycle. This curve shows that a single oscillation prevails in the tropics, but that even here a double oscillation is superposed. The temperature maximum does not coincide with the sunspot minimum, as would be the case with pure antiparallelism, but occurs two years before. Summarizing all phenomena, we find that about 2–2½ yr before the sunspot extremes, the general (zonal) atmospheric circulation is increased. The subtropical high-pressure belts are stronger and displaced poleward in midsummer, at least in the North Atlantic and European region. Moreover, the pressure difference between winter and summer over the Asiatic continent is diminished, and central Europe has dry midsummers as a rule. In years of sunspot extremes (*i.e.*, minima as well as maxima) and shortly before and afterward, the following phenomena are observed: The planetary circulation is weakened, the mean pressure in the high-pressure belts is lowered, the difference between winter and summer pressure in Asia is increased, and the

Table II. Correlation Coefficients of Zonal Pressure and Temperature Differences with Monthly Changes of North Atlantic Meridional Pressure Difference* (1874–1923)

	Station pairs	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Pressure difference	Rome—Ponta Delgada	+0.36	*+0.48*	+0.37	*+0.53*	*+0.59*	*+0.46*	+0.35	+0.32	+0.16	*+0.48*	+0.21	+0.27
	Indianapolis—Ponta Delgada	*+0.42*	*+0.52*	*+0.66*	*+0.56*	+0.41	*+0.44*	*+0.48*	+0.29	+0.23	*+0.46*	+0.41	+0.34
	Haparanda—Stykkisholm	−0.39	*−0.58*	*−0.55*	*−0.58*	−0.26	*−0.59*	*−0.58*	−0.19	−0.30	*−0.43*	−0.35	−0.16
	Jacobshavn—Stykkisholm	−0.32	*−0.45*	*−0.47*	*−0.47*	−0.06	−0.36	−0.23	−0.31	*−0.50*	−0.39	*−0.44*	−0.36
Temperature difference	Tromsö—Western Greenland	−0.16	*−0.42*	*−0.44*	−0.41	−0.21	*−0.59*	−0.16	*−0.43*	*−0.53*	*−0.51*	−0.39	−0.34

* The pressure and temperature differences for the station pairs at the left for the indicated month are correlated with changes in pressure difference between Ponta Delgada and Iceland from the indicated month to the next; correlation coefficients that exceed the scope of chance are italicized.

frequency of severe winters in central Europe, as well as in central and eastern North America, is increased.

Figure 3 shows that in central Europe, during the period 1755–1949, winters with a negative temperature departure of more than 1.8C occurred particularly around the time of sunspot maxima and minima. In central Europe, during the intervals between 0.5 yr before and 1.0 yr after a sunspot maximum, nine out

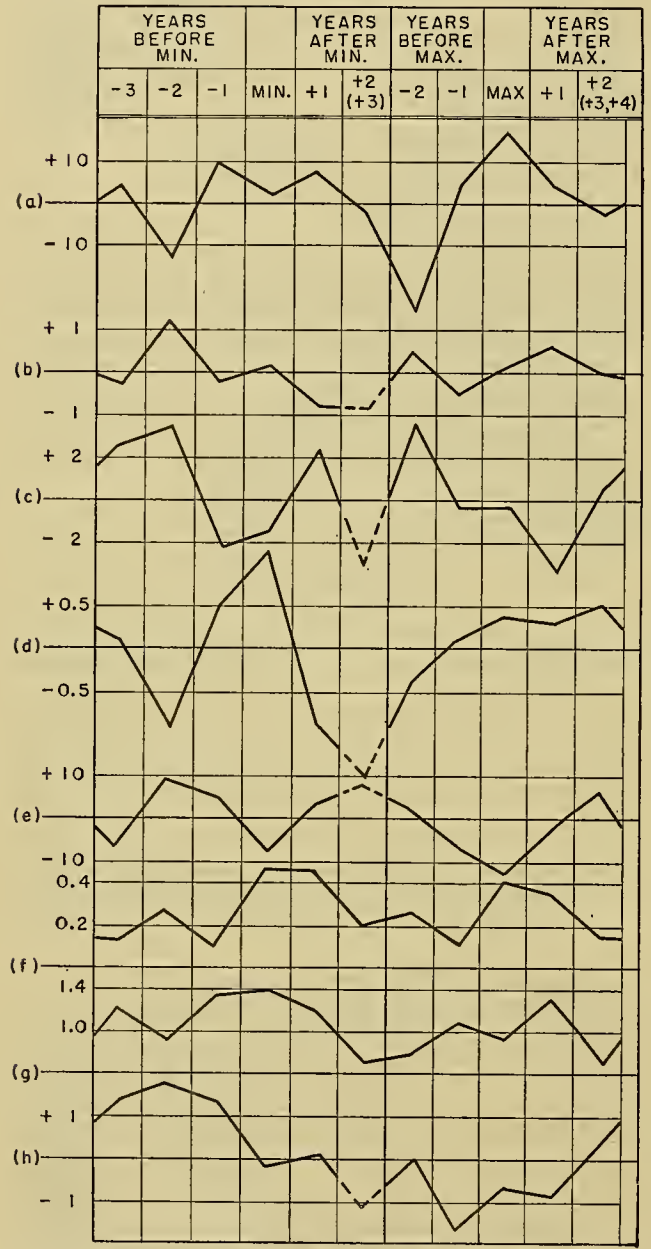


Fig. 2.—The average course of some meteorological elements during a sunspot cycle. (The dashed portion of some curves indicates that these cases, because of the relative brevity of the interval, are based on only few years of observations.) (*a*) Departure of precipitation (in mm) in central Europe during the midsummers, 1803–1943; (*b*) pressure departure (in mm *Hg*) in Berlin for the midsummers, 1875–1944; (*c*) departure of the pressure difference (in mm *Hg*) between the Azores and Iceland for the winters, 1866–1940; (*d*) departure of the difference between winter and summer pressures (in mm *Hg*) for ½ (Barnaul + Irkutsk) + ½ (Lahore + Karachi) for 1881 to 1940; (*e*) departure of the mean annual pressure for ½ (Capetown + Adelaide) (in 1/1000 in.) for 1865 to 1940; (*f*) yearly number of severe winter months (| departure | $>1.6\ \sigma$) in central Europe, 1752–1947; (*g*) yearly number of cold winter months (| departure | >4.5F) for Chicago + St. Louis, 1847–1946, (*h*) departure of the mean annual temperature in $\frac{1}{10}$C for ½ (Apia + Colombo), 1890–1945.

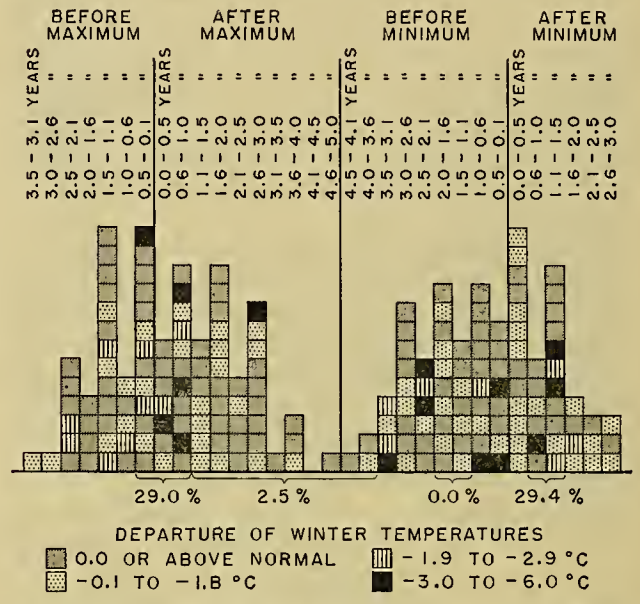


Fig. 3.—The position of the winter (January) with respect to the nearest sunspot extreme and the temperature characteristics of the winter in central Europe for the period 1755–1949 (for lack of space the individual years have not been identified). The frequencies below the abscissa refer to winters with negative temperature departures of more than 1.8C.

of thirty-one winters (29 per cent) were too cold by more than 1.8C, whereas only one such winter (2.5 per cent) occurred in the intervals from 1.1 yr after a sunspot maximum to 3.6 yr before a minimum. The probability that this difference in the relative frequencies is due to chance (assuming a constant fundamental probability) amounts to 0.0018 (computed according to R. A. Fisher's *H*-method). This probability is below the residual probability of the usual limits of the chance range. Thus a physical relationship must be assumed to exist here.

It has been shown in a similar manner that the relative frequencies of various *Grosswetter* phenomena in different seasons and in different parts of the world exhibit fluctuations within the solar cycle, a fact incompatible with the assumption of a constant fundamental probability. For this reason, there can no longer be any doubt that the *Grosswetter is dependent on the solar cycle* [18; 19; 20, pp. 962–966]. It must be emphasized, however, that this dependency does not appear in the primitive form of a linear relationship with sunspots.

Relationships between Grosswetter *and Variations in Solar Radiation.* It seems surprising, at first, that precipitation and temperature in the tropics show mainly a single oscillation within a solar cycle, whereas in the temperate zones, a double or higher multiple oscillation can be observed. This can, however, be easily explained if we consider that, on the one hand, the sun can exert an influence upon the weather only by way of radiation and that, on the other hand, the sun's radia-

tion is an extraordinarily complex phenomenon. Those parts of the sun's radiation which emanate from the sun's corona and the uppermost chromosphere fluctuate with the sunspots, that is, on the average, with an eleven-year period. These radiations (corpuscular rays and electromagnetic waves) are, however, with one exception, absorbed in the ionosphere and hence, according to our present ideas, are of no influence upon the weather. Only the ultrashort electromagnetic waves with a wave length $\lambda < 15$ m penetrate to the earth's surface. Since, according to Kiepenheuer [35], ultrashort waves act upon colloidal processes, it is possible that the single fluctuation of precipitation in the tropics is caused by a solar influence on the condensation and sublimation of the water vapor, whereas the temperature fluctuation is secondarily produced by cloudiness and precipitation. In agreement with this assumption is the fact that the correlation of sunspots with temperature is of opposite sign and of less significance than is the correlation of sunspots with the amount of precipitation or with the water level of large inland lakes.

The radiation emanating from the photosphere, that is, the heat radiation, shows—as has been demonstrated several times [7, 9, 26]—no single oscillation in or out of phase with the sunspots. However, when the different mean annual trends in the time intervals 1919–23 and 1924–30 are eliminated [9], the curve of the solar constant for the period from July 1918 to February 1937 shows minimum values at about the time of sunspot extremes, and maximum values 1½ to 2½ yr before the sunspot extremes [19; 20, pp. 967–968]. These fluctuations agree quite well with the double oscillation of the meteorological elements as presented in the preceding section. However, it must be conceded that the double fluctuation in the most recently revised monthly means of the solar constant no longer appears so clearly [1].

An increase in the total energy radiated from the sun would mean an increase in the heat supply, particularly for low latitudes during winter. The temperature difference between the tropics and the polar region would thereby be increased, resulting in an increase of the general circulation of the atmosphere and a consequent increase of the pressure gradient between the subtropics and the subpolar low-pressure belt. The reality of the fluctuations of the solar constant is corroborated by the fact that during midwinter (January and February) in the period 1919–37, a strong positive correlation existed between the mean solar constant and the simultaneous mean values of the pressure differences between Valencia (Ireland) and Spitsbergen and between Rome (Italy) and Haparanda (Sweden) [17, pp. 97–105]. These correlation coefficients were +0.69 and +0.65, respectively, both of which exceed the maximum chance value.

The objection has been raised that the variations of the solar constant are too small to cause fluctuations of the general atmospheric circulation. However, according to the Stefan-Boltzmann law, a variation of an amplitude of 1 per cent of the solar constant (assuming incoming and outgoing radiation to be equal) causes a temperature change (temperature departure) of 0.7C which refers to the entire earth's surface. It is therefore quite possible that, owing to the inequalities of insolation (as a function of the geographical latitude) and the different heat capacities of land and ocean, the temperature change caused by a one per cent change of insolation may amount to considerably more than 0.7C in some regions.

Results of recent investigations [35, 37] suggest that the fluctuations of the atmospheric circulation within a sunspot cycle are caused not so much by the changes of the entire heat radiation as by the fluctuations of the sun's radiation in the ultraviolet. This idea is supported by the fact that the difference between the areas of the sun's faculae F and those of sunspots L, both relative to their long-established mean values F_0 and L_0, undergoes the same double fluctuation within a solar cycle as does the atmospheric circulation [19]; this is shown in Fig. 4. Radiation emanating from the

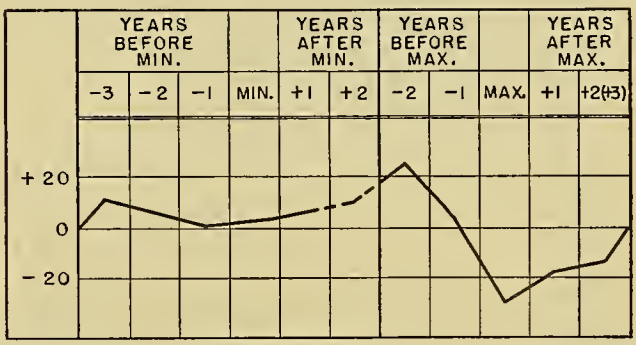


FIG. 4.—The mean course of the difference between the equivalent annual means of the areas of sun faculae (F) and sunspots (L), $100(F/F_0 - L/L_0)$, for the sunspot cycle, from 1875 to 1940.

faculae is known to be particularly rich in ultraviolet. This relationship can be explained as follows: In the upper mixing layer above the stratospheric inversion layer, temperature differences are set up by differences in the absorption of ultraviolet radiation. These temperature differences are connected with pressure differences of the same sign in a manner similar to that found in the middle troposphere. If this assumption holds true, an increased ultraviolet radiation results in an increased pressure gradient from the equator toward the pole in the upper-air layers, because—disregarding the time around the summer solstice (see p. 824)—the ultraviolet radiation as well as the total radiation affects primarily the lower latitudes. Because of the increase of the pressure gradient aloft, the planetary (zonal) circulation increases, and the poleward border of the subtropical high-pressure belt is displaced toward the pole.

Solution of the Second Basic Problem. It follows from the discussion in the foregoing sections that weather-governing complexes, the existence of which has been proven (first empirical theorem), are not to be found in terrestrial processes. The terrestrial processes are either too rare (strong volcanic eruptions) or too weak (motions of the pole) to influence the course of the weather. There are, indeed, purely terrestrial regulatory

phenomena such as the alternation between meridional and zonal circulation in the Northern Hemisphere (see p. 827) and the southern oscillation in the Southern Hemisphere. However, these regulatory phenomena merely shape the macrometeorological events and cannot be regarded as governing complexes of conditions. If they alone were decisive, obvious periods should occur in the course of the weather, whereas in reality one must search for such periods by means of "methods for the discovery of hidden periodicities." The solution of the second fundamental problem is given by the empirical facts briefly sketched in the two preceding paragraphs from which the following conclusions can be derived:

Fourth Empirical Theorem: *The fluctuations in the solar radiation represent complexes of conditions that govern the course of the large-scale weather.*

for the latitudes 50°N, 60°N, and 70°N, where R is the mean radius of the earth, ϕ is the latitude, and $|\Delta p|_\lambda$ is the pressure difference from one extreme value to the next along a latitude circle. The mean zonal air motion (eastward *and* westward) is expressed by the mean of the meridional pressure gradients at sea level from 40°N to 75°N, averaged over all meridians,

$$M_{40-75} = \frac{1}{n}\sum_{1}^{n}\frac{\Sigma\,|\Delta p|_\phi}{L},$$

where L is the length of the meridional distance between 40°N and 75°N, $|\Delta p|_\phi$ is the pressure difference from one extreme value to the next along the meridian between 40°N and 75°N, and n is the number of meridians for which $|\Delta p|_\phi$ is computed [22]. It should be noted that the mean gradients were first determined for each single day and the monthly means computed

Table III. Monthly Means of the Daily Zonal and Meridional Pressure Gradients in the Northern Hemisphere

Month	Zonal pressure gradients* (mb per 1000 km)									Average of all meridional pressure gradients (mb per 1000 km)		
	1937			1938			1939			1937	1938	1939
	50°N	60°N	70°N	50°N	60°N	70°N	50°N	60°N	70°N	40°–75°N	40°–75°N	40°–75°N
January				10.7	12.4	10.5	10.7	10.5	12.0		11.5	11.1
February				10.5	11.8	10.5	12.2	11.9	8.9		10.7	11.9
March				9.7	10.6	10.6	9.8	10.8	9.7		10.6	10.3
April				8.8	10.2	9.8	8.8	10.4	8.9		9.8	9.0
May				8.0	9.0	8.1	7.8	8.9	7.6		8.2	7.8
June				6.9	7.6	7.4	6.8	7.5	7.4		7.2	7.0
July				6.0	7.0	7.3	6.1	7.3	7.0		6.4	6.5
August	6.0	7.4	7.9	(6.2)	(7.8)	(7.5)				7.3	8.1	
September	7.8	10.0	9.6	(8.0)	(9.8)	(9.0)				8.6	9.0	
October	9.0	10.3	10.0	(8.7)	(11.8)	(12.7)				9.6	10.8	
November	10.1	11.3	10.7	(10.1)	(13.1)	(12.0)				10.3	11.5	
December	11.2	12.3	9.4	(11.2)	(12.2)	(13.1)				10.5	11.8	
Mean (Aug. 1937–July 1939)							8.8	10.1	9.5			9.4
Mean (50°N–70°N)								9.5				

* Numbers in parentheses are uncertain.

Therefore, an intensive promotion of solar physics is one of the prime prerequisites for progress in long-range weather forecasting.

Morphology of the *Grosswetter.* *Forms of the General Atmospheric Circulation.* Twenty years ago it was recognized that it is erroneous to consider the west-east circulation as the essential feature of the "general atmospheric circulation" in the temperate zone [8; 33, pp. 215–218]. Without the meridional movements there would not be the exchange between the warm subtropical and cold polar air masses that is the most important effect of the general circulation in the temperate zone. The quantitative equivalence of zonal and meridional circulation in the lower troposphere is shown in Table III. Here, the mean meridional air motion (poleward *and* equatorward) is expressed by the mean of the zonal pressure gradients at sea level,

$$Z(\phi) = \frac{\Sigma\,|\Delta p|_\lambda}{2R\pi\cos\phi},$$

from them afterwards. Comparison shows that the differences between the mean zonal and the mean meridional gradients by months are much less pronounced than the seasonal trend of both these indices.

Sometimes, however, large differences between these indices occur *on individual days.* Thus the mean zonal gradient at 60°N amounted to only 8.5 mb per 1000 km on January 11, 1949 but, on the same day, the average of all meridional gradients from 50°N to 65°N was 15.0 mb per 1000 km. The mean gradient for meridians with west-east zonal circulation only, between 50°N and 65°N, (that is, two-thirds of all meridians on that particular day) was 17.5 mb per 1000 km. Table IV shows the pronounced daily fluctuations of the circulation in addition to the seasonal trend. On some winter days, values are observed that almost equal the summer normals and on many summer days the gradients correspond closely to the mean values of winter.

The recently available daily circumpolar 500-mb con-

tour charts show that a remarkable meridional circulation extends also to the upper troposphere of the middle latitudes. Even at the 5- and 10-km levels, there is, only in exceptional cases and then at the most as far south as 55°N, an uninterrupted west-east circulation around the pole (Fig. 5). In most cases, the isobars at the 5-km level seem to "meander" (Fig. 6), the poleward convexities representing warm ridges, the poleward concavities representing cold troughs [52, 53].

TABLE IV. DAILY VALUES OF THE MEAN ZONAL GRADIENT ALONG THE 60°N LATITUDE CIRCLE DURING THE FIRST SIX MONTHS OF 1949*
(In mb per 1000 km)

Day	January	February	March	April	May	June
1	12.1	13.7	13.1	*7.0*	**10.8**	7.3
2	12.6	*10.2*	13.1	8.1	7.3	9.1
3	9.4	10.7	12.4	11.1	9.5	8.6
4	12.6	13.2	13.7	11.9	8.8	7.9
5	12.6	12.7	13.7	11.4	7.7	8.5
6	10.6	13.1	*10.1*	9.7	10.6	7.0
7	11.1	11.2	12.5	9.1	*7.2*	*6.4*
8	**15.5**	**13.9**	**14.0**	9.0	7.4	7.6
9	15.1	9.2	10.5	10.0	9.0	7.8
10	9.9	13.5	*10.0*	**12.4**	8.1	7.3
11	8.5	13.6	11.3	9.9	7.9	6.9
12	10.5	11.8	12.1	7.8	10.0	8.2
13	7.4	10.4	12.3	7.9	8.2	8.0
14	*7.3*	*9.0*	11.6	7.1	8.0	7.4
15	10.5	10.3	11.9	*6.4*	7.2	8.1
16	11.3	11.6	12.7	7.4	9.1	**11.7**
17	12.8	12.4	11.9	8.1	**10.7**	10.4
18	13.4	**14.8**	15.3	8.8	9.4	11.6
19	13.4	13.9	**16.0**	8.9	7.2	9.5
20	11.3	11.7	10.8	6.5	*6.3*	8.3
21	14.5	12.0	10.3	7.6	8.2	7.4
22	**17.6**	12.8	11.0	**10.5**	8.0	*7.2*
23	14.3	8.8	12.0	8.3	8.2	7.4
24	14.0	*8.7*	9.7	9.1	8.4	**9.1**
25	13.2	10.8	10.2	9.9	10.0	6.9
26	11.9	9.8	9.0	8.5	8.2	8.4
27	11.9	10.7	8.7	8.3	8.1	7.3
28	*11.5*	**13.8**	7.7	*7.3*	**10.2**	*4.3*
29	13.1		8.5	7.4	8.6	8.2
30	15.6		7.2	8.1	8.7	7.7
31	**15.9**		7.7		9.0	
Monthly means	12.3	11.7	11.3	8.8	8.6	7.0

* The principal maxima are in boldface; the principal minima are in italics.

The sinuosities of the isobars are called waves by some meteorologists. One can, however, argue that this designation is misleading because it is suggestive of progressive waves. Experience has taught us that the ridges and troughs at the 5-km level are, in general, rather stationary with respect to their geographical longitude (as in Figs. 7 and 8). Zonal movements occur principally when two pressure ridges on either side of a trough join on the poleward side of the trough thereby cutting off a low-pressure cell, or when two troughs on both sides of a wedge join on its equatorward side thus cutting off a high-pressure cell. In some cases there may be longer zonal movements of individual pressure centers in the upper troposphere. An impressive example of such a case is the high-pressure cell at the 5-km level that within five days (November 24–29, 1949) moved from 70°N and 172°W to 78°N and 55°W along an unusual path. This is also an example of the fact that relatively warm anticyclones in the upper troposphere can occur at rather high latitudes even during the winter season. (Regarding zonal displacements of wedges and troughs, see p. 830.)

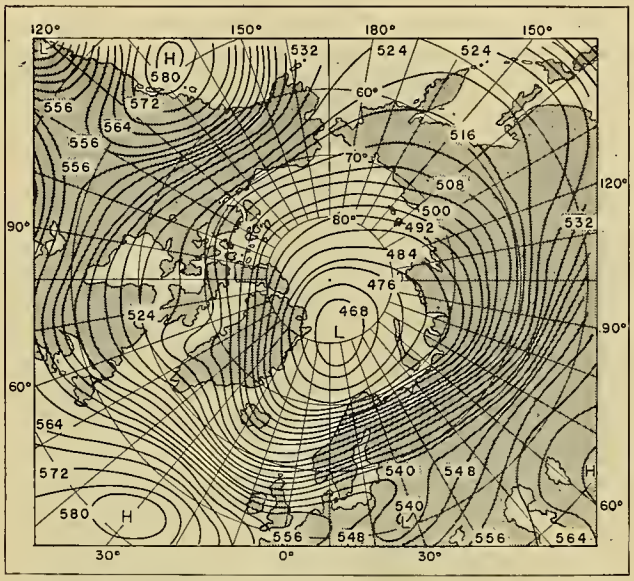


FIG. 5.—Contours of the 500-mb surface (in tens of dynamic meters) for January 11, 1949 at 0200 GMT. Example of a circumpolar zonal circulation.

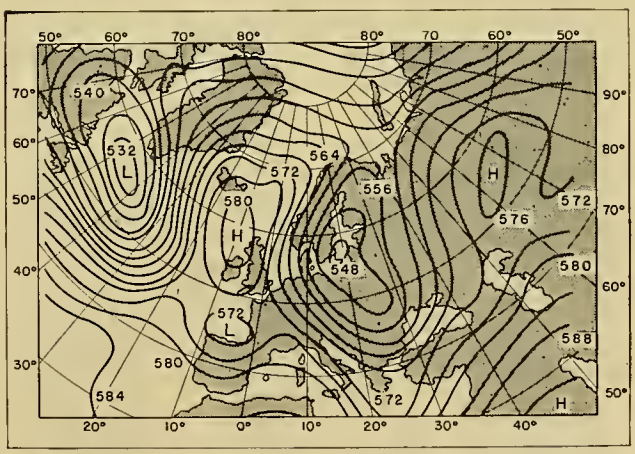


FIG. 6.—Contours of the 500-mb surface (in tens of dynamic meters) for June 23, 1949 at 0200 GMT. Example of a circulation in the form of elongated meridional circulation cells (*Zirkulationsstreifen*).

The centers of high- and low-pressure cells at the 5-km level for the period from June 21 to July 5, 1949 are shown in Fig. 7. In each of the regions marked I and III there is a moving low-pressure center (the one of region III splits at times). Regions II and IV each contain a high-pressure center and show, in particular, the small zonal displacement. On June 27 and 28, a stronger zonal movement occurred. to the northeast for region

II, and to the northwest for region IV, thus producing a new high-pressure cell in region V, which, according to the investigations of Palmén and Rossby [53], remained there until July 6.

Fig. 7.—Centers of anticyclones (●), ridges (○), and cyclones (+) at the 5-km level from June 21 to July 5, 1949.

Assuming that the meridional heat transport is the main effect of the general circulation, we can distinguish the following principal forms of the general circulation [8; 17, pp. 49–55].

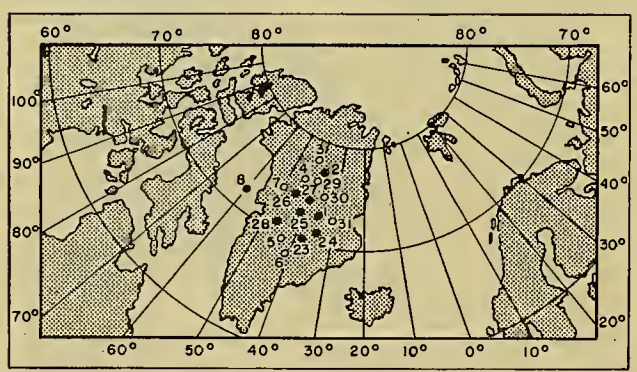


Fig. 8.—Centers of anticyclones (●) and ridges (○) at the 5-km level from July 23 to August 8, 1949.

1. *An unordered heat exchange.* This occurs in west-east-travelling highs and lows of the lower troposphere with isobars running approximately from west to east at the 5-km level (zonal circulation). In this case, the southern portion of the troposphere (in the Northern Hemisphere) is warm, the northern portion is cold. As long as the flow equilibrium between warm and cold air remains undisturbed, a "frontal zone" ascending poleward exists in the troposphere with a strong meridional temperature gradient [47, 52, 53]. Above this frontal zone, the isobars of the upper troposphere are crowded in a comparatively narrow belt, so that the west-east current reaches very high speeds (jet stream). The maximum of this zonal movement is found where the tropopause slopes down most strongly toward the pole. This form of circulation (Fig. 5) rarely develops as an upper circumpolar low as far south as the 55th parallel, as is so in the case illustrated in Fig. 5, and, according to our present experience, it usually develops in only one or two quadrants, and never in the entire temperate zone around the whole hemisphere. Even in the rare case represented in Fig. 5, the jet stream does not run uninterruptedly around the entire hemisphere, but parts of it are dissipated by divergent currents.

2. *The meridional circulation strips.* An ample heat exchange is effected in these strips; quasi-stationary warm tropospheric anticyclones that reach up to the stratosphere and are displaced far poleward alternate with cold tropospheric troughs that reach far equatorward and likewise extend to high altitudes (Fig. 6).

Whereas these two principal forms occur during all seasons, the two following secondary forms occur only in certain seasons:

3. *The winter anticylcone in the lower troposphere* [52]. Below the circumpolar low at high altitudes there often exists during the winter an anticyclonic circulation system in the lower troposphere over the interior of the Arctic. When these arctic air masses penetrate to the south, additional cooling occurs through radiation over the continent, and wide regions of stationary cold anticyclones are thus formed. The frontal zone between these cold air masses and warm air to the south, and the resulting westward drift, are displaced far toward the equator. In Europe this displacement extends to the Mediterranean area, in North America to the northern edge of the Gulf of Mexico.

4. *The arctic anticyclone of the stratosphere in midsummer.* According to Baur and Philipps [23], the total insolation during the summer solstice is strongest at the pole, and thus the ultraviolet radiation is most likely also to be a maximum there. For this reason, in the upper stratosphere a pressure gradient from the pole to the equator exists during this short season, in strong contrast to midwinter conditions.

Figure 9 shows the average topography of the 41-mb surface (at approximately 22-km height) as it exists, according to Scherhag [47], in July over the Northern Hemisphere. In the Arctic this high pressure extends in some years to the earth's surface, as in Europe from June 5 to June 13, 1947 (Fig. 10) before the extremely dry summer.

Fifth Empirical Theorem: *On the average for all meridians and over long periods of time, the zonal and meridional components of the atmospheric circulation in the lower troposphere are nearly equal in middle latitudes of the Northern Hemisphere.*

Sixth Empirical Theorem: *In the lower and middle troposphere on both hemispheres, there are always long*

zones, extending over many meridians, in which warm subtropical air borders on cold polar air (polar front). In the upper troposphere above these zones, there is always a narrow belt with extremely strong westerly winds (jet stream), which reach their maximum velocity at the tropopause level. However, neither the polar front nor the jet stream runs uninterruptedly around the entire Northern Hemiphere.

Seventh Empirical Theorem: *A zonal circulation exists only on rare occasions over all meridians of the middle latitudes. In most cases, regions with a prevailing zonal circulation and regions with a prevailing meridional circulation exist simultaneously.*

Types of Grosswetter. Since in middle latitudes of the Northern Hemisphere the character of the circulation is seldom uniform over the entire hemisphere, it is necessary to consider a division into separate circulation cells, which Seilkopf refers to as "special circulations" [52]. Furthermore, it is necessary to define an observational unit which lies between the daily weather situation of a smaller region and the world weather situation over the Northern Hemisphere or even the entire earth. This unit is the *Grosswetterlage*. Because of the existence of two continents (one of which has a large longitudinal extent) and two oceans on the Northern Hemisphere, there usually exist four or five troughs and the same number of ridges in the Northern Hemisphere. Therefore, the most suitable regional delineation of *Grosswetterlagen* results from a division of the total circulation of the Northern Hemisphere into five (partly overlapping) special circulation regions [17, pp. 58–62]:

North America	(160°W–60°W)
North Atlantic	(70°W–0°)
Europe	(30°W–45°E)
Asia	(45°E–150°E)
North Pacific	(150°E–135°W).

At some future time it might perhaps be useful to combine the overlapping North Atlantic and the European regions into a single circulation region.

The temporal limits of a *Grosswetterlage* result from the following definition [13]: A *Grosswetterlage* is the mean pressure distribution (at sea level) for a time interval during which the position of the stationary (steering) cyclones and anticyclones and the steering within a special circulation region remain essentially unchanged. Here "steering" means the direction of propagation of 24-hr isallobars. The reference to the pressure distribution at sea level is necessary in order to enable one to apply this definition to observations of earlier years when there were no daily aerological observations. On the other hand, by requiring constancy of the steering, the pressure distribution in the middle troposphere (5-km level) is indirectly taken into consideration.

The mean pressure distribution during an arbitrary five-day period cannot be considered as a *Grosswetterlage*. Rather, it is possible that by such an arbitrary formation of means, two different *Grosswetterlagen* might cancel each other.

Baur [17; 20, pp. 920–923] established 17 types of *Grosswetterlagen* for the circulation region of Europe—25 types, if finer subdivisions are included. For the circulation region of the North Atlantic he introduced 16 types of *Grosswetterlagen*. The 25 European *Grosswetter* types can be grouped into the following divisions:

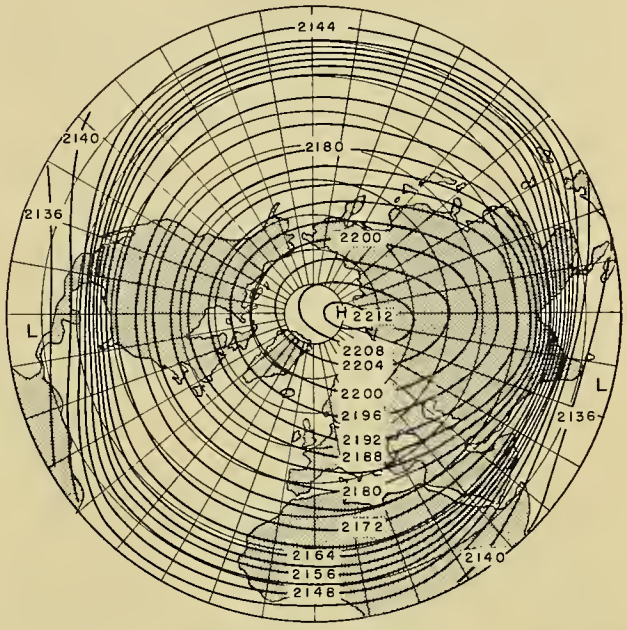


Fig. 9.—Average topography of the 41-mb surface for July in the Northern Hemisphere. (*After Scherhag* [47]).

1. A high in the northwest region (between Greenland and Scandinavia).

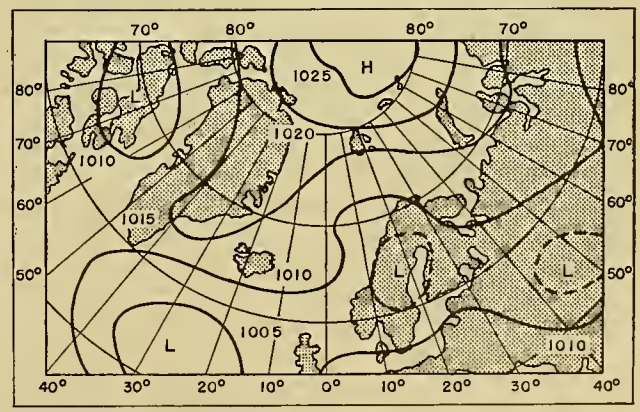


Fig. 10.—Mean pressure distribution (in mb) at sea level in the north polar region for June 4–12, 1947.

2. A high in the west and southwest (between the Azores and England).
3. A continental high.
4. A high in the southeast and south.
5. Westerly cyclonic flow.
6. Central cyclonic situations.

The 16 types of the North Atlantic *Grosswetterlagen* are grouped into zonal, meridional, and mixed situa-

tions. (For further details, see [13; 17; 20, pp. 920–923].) For the North American circulation region, Krick and Elliott have established a system of weather types.[2]

Considering the pressure distribution at the 5-km level, we can classify the *Grosswetterlagen* in all special circulation regions of the temperate zone in the Northern Hemisphere according to the following seven fundamental types based on the circulation:

1. Three types with zonal circulation:
   - I. High pressure in the south; low pressure in the north; westerlies and jet stream reach farther south than normally.
   - II. High pressure in the south; low pressure in the north; northern boundary of high pressure either in normal or more northerly position; westerlies and jet stream far to the north.
   - III. High pressure in the north; low pressure in the south.
2. Four types with meridional circulation:
   - IV. Subtropical meridional flow; high pressure in the east; low pressure in the west.
   - V. Polar meridional flow; high pressure in the west; low pressure in the east.
   - VI. Meridional ridge.
   - VII. Meridional trough.

These seven types occur in all special circulation regions of the temperate zone in the Northern Hemisphere, but with different frequencies.

Schematic illustrations of the pressure distribution and air flow in the middle troposphere occurring with these seven types are shown in [21].

Annual Variation of Grosswetterlagen. *Frequency Maxima above Chance Limit* (*Weather Key-Days*). The establishment of types of *Grosswetterlagen* and the determination of the corresponding weather represent only the first steps toward rational macrometeorological research. Further steps must include the following: (1) determination of the seasonal trend of *Grosswetter* types, (2) investigation to determine whether certain *Grosswetterlagen* occur at certain times of the year more frequently than according to chance, (3) determination of the persistence and repetition tendency of individual *Grosswetter* types and their dependence on the seasons and on the general atmospheric circulation, and (4) investigation of the frequency and the circumstances of transition from one *Grosswetterlage* to another. There are many problems to be solved which require the closest coordination of theory, synoptics, and statistics based on the probability theory. A clear-cut solution of these problems is of incomparably greater value for extended-range forecasting, particularly for medium-range forecasting, than are the many harmonic and rhythm analyses still undertaken at several institutes.

The determination of the annual course of the relative frequency of European *Grosswetter* types in a 63-yr period showed [16] that in Europe one or more types of *Grosswetterlagen* occur in 22 intervals of several days each during the year with so high a relative frequency that it cannot be explained as chance fluctuation of a uniform distribution over the entire year. The number 22 represents only the present state of our knowledge; it is possible that this number will be increased with the further increase of observational data. As an example, Fig. 11 gives a portion of the annual course of

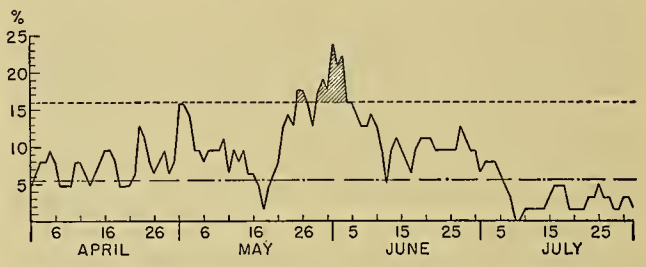


FIG. 11.—Seasonal variation of the relative frequency of the *HN*-situation from April 1 to August 1 for the period 1881–1943. The dash-dot line represents the annual mean value of the relative frequency of the *HN*-situation among the other *Grosswetterlagen*, the dashed line represents the upper limit of the range of chance of this relative frequency, where the range of chance is defined in the usual manner as the range which contains 99.730 per cent of all values.

the relative frequency of the "*HN*-situation," a *Grosswetterlage* which is characterized by a "steering high" or "central high" over the northeastern Atlantic Ocean between 30°W to 5°E and 58°N to 75°N. As can be seen from this figure, the relative frequency of occurrence on May 25 and 26, as well as on May 29 to June 3, exceeds the maximum chance limit. In such a sector of better-than-chance occurrence, the day of the greatest relative frequency is called a "weather key-day."

The general physical explanation of the annual course of the relative frequency of the European *Grosswetter* types by means of the annual course of insolation and the general circulation of the atmosphere is given elsewhere [17; 20, pp. 920–923].

It must be emphasized, however, that the relative frequency of a given *Grosswetter* type, at least in Europe, never reaches 50 per cent on any day of the year, and, even when several similar *Grosswetter* types are combined into one group, it never exceeds 70 per cent (see second empirical theorem).

EIGHTH EMPIRICAL THEOREM: *The occurrence of certain* Grosswetter *types is connected with the annual course of the general atmospheric circulation in such a manner that during certain parts of the year certain types occur more frequently than according to chance. However, they do not occur so frequently that, without the aid of other indications, an extended-range forecast could be based on them.*

Weather Key-Days and Weather Development. The weather key-days indicate, so to speak, the rhythm of the normal seasonal trend of the atmospheric circulation in the special circulation region under consideration. The term "rhythm," however, is not to be interpreted as a periodic or quasi-periodic process, but its meaning in this connection is only a defined dependence upon time. The basic cause of this rhythm is the annual course of the incoming and outgoing

2. Consult "Extended-Range Forecasting by Weather Types" by R. D. Elliott, pp. 834–840 in this Compendium.

radiation. The irregularity of the rhythm probably has its cause in the irregular distribution of land and water and in the variation of the period of the possible free oscillations. This latter variability results from the continuous change of temperature in the course of the year. The occurrence or nonoccurrence of the atmospheric phenomena pertaining to the weather key-days of the year is therefore the expression of the circulation character of the season or year in question. It is therefore probable that the weather on weather key-days is connected with the later weather development.

In fact, in Europe series of better-than-chance connections have been found between characteristic weather anomalies occurring around weather key-days and the subsequent or later meteorological situation [15; 17, p. 133]. For example, September 29 is a weather key-day for Europe. On this day, as well as on the preceding and the following day, certain European anticyclonic situations within the period from 1881 to 1943 reach a better-than-chance frequency [16]. From September 23 to October 1, central Europe has in 65 per cent of all years a so-called *Altweibersommer* with little cloudiness, infrequent precipitation, and relatively high midday temperatures. Weather of this character occurs in high-pressure regions, stretching from west to east, with a zonal circulation over northern Europe, as well as in the region of high pressure with the center over eastern Europe accompanied by a meridional circulation. In the latter case, the pressure at Moscow is considerably above normal. As is shown in Table V,

Table V. Pressure Anomalies During the European "Altweibersommer" and Temperature Character of the Subsequent Winter

(Example of an unequivocal, significant relation)

Year	Pressure departure (in mb)			Temperature departure of the subsequent winter in central Europe (cent. deg.)
	Sept. 21 to 30		Sept. 1 to 30	
	Moscow	Berlin	Moscow	
1884	+5.3	+3.7	+3.5	+0.9
1901	+8.3	+2.4	+4.7	+1.4
1902	+2.3	+9.3	+0.9	+0.7
1904	+16.7	+1.5	+8.8	+1.1
1909	+5.6	+2.0	+5.5	+2.2
1912	+7.2	+9.7	+2.8	+1.5
1913	+4.8	+6.0	+1.3	+0.8
1920	+12.3	+5.1	+3.2	+2.4
1926	+4.3	+1.1	+0.1	+1.6
1929	+6.3	+5.9	+1.3	+2.0
1937	+3.7	+2.3	+0.9	+1.2
1938	+12.4	+3.9	+5.9	+1.2
1947	+5.7	+0.4	+1.2	+2.2

in all the years of the period from 1881 to 1948 during which the mean pressure between September 21 and September 30 at Moscow was 2.0 mb above normal and, furthermore, the mean pressure departure was positive for Berlin from September 21 to 30 and for Moscow from September 1 to 30, the average temperature of the subsequent winter in central Europe was, without exception, greater than 0.6C above normal. Since the basic probability of such winters is 41.8 per cent, the upper limit of the scope of chance, within which the relative frequency can fluctuate by chance alone (if one defines the chance limit as on p. 817) lies at 86.5 per cent for $n = 13$. We can therefore conclude from the 13 cases (100 per cent) of mild winters shown in Table V that a physical relationship exists between the given pressure anomalies in September and the temperature character of the following winter in central Europe. This relationship is founded on the laws followed by the change between meridional and zonal circulation in the North Atlantic and European circulation regions.

Ninth Empirical Theorem: *Unequivocal relationships which lie outside the scope of chance exist between the weather anomalies about the time of the weather key-days and the subsequent* Grosswetter.

The Change between Meridional and Zonal Circulation. The persistence of each of the seven types of special circulation, described on page 826 is limited. For example, if with low pressure over the North Atlantic and high pressure over Europe and North America, subtropical meridional flow (Type IV) prevails for a relatively long period over western Europe and polar meridional flow (Type V) over eastern North America, then over western Europe and the eastern Atlantic, warm subtropical air is continuously transported toward the polar regions, while over eastern North America and the western Atlantic, cold polar air is transported toward the subtropical regions. Thereby the temperature contrast between subtropical and polar regions is diminished and with it the fundamental cause of the meridional circulation, so that eventually a zonal circulation replaces the meridional circulation. On the other hand, in the absence of a meridional exchange the temperature difference between subtropical and polar regions is re-established by the radiation difference between these two regions. With purely zonal circulation the temperature contrast between low and high latitudes eventually becomes so great that it forces the establishment of a meridional circulation. This re-establishment of a meridional circulation is explained as follows: The increase in the meridional temperature gradient causes a strong concentration of isotherms connected with a strengthening of the jet stream aloft. The large-scale lateral mixing processes (and frictional forces) on both sides of the vigorous westerly flow cause the formation of cyclonic vortices on the equatorial side and of anticyclonic vortices on the polar side of the band of west winds [44, 53]. Thus pressure troughs and ridges are formed, and a meridional circulation is again started.

In this manner, a continuous change between increased and diminished, and between predominantly zonal and predominantly meridional circulation, takes place. However, this change in circulation types does not occur in the same manner over the entire temperate zone of the Northern Hemisphere and, in general, not simultaneously either. This follows from the fact that, aside from the very rare case of a zonal circulation over all meridians, one and the same circulation type never exists in all special circulation regions (see seventh empirical theorem). The significant correlation coefficients in Table II confirm the fact that

circulation changes within limited regions—of the size of perhaps a continent or an ocean—actually exist and are caused to a large extent, though not exclusively, by the distribution of pressure and temperature within these regions and those immediately adjacent to them.

CRITIQUE OF THE METHODS OF EXTENDED-RANGE FORECASTING

Medium-Range Forecasts. It is useful to distinguish between medium-range forecasts (for two to five days) and extended-range forecasts (for six to ten days, for months, or for seasons). The methods of synoptic meteorology play a decisive role in medium-range forecasts. This is possible today, since weather maps of the entire Northern Hemisphere are available every day owing to the improvement of the communication system. However, it is necessary to supplement synoptics by meaningful statistics, partly to enlarge and consolidate our experience, and partly to create objective bases which could be of direct aid to forecasting. The application of rhythm analyses and symmetry points is of little use for the medium-range forecast, since the shorter pressure waves are mostly nonpersistent [20, pp. 933–934]. These methods are therefore discussed in the section on long-range forecasting.

Analogue Methods. An aid for the medium-range forecast can be found in the study of the subsequent development of past cases with both similar pressure distributions and similar preceding *Grosswetter* developments. It is to be noted, however, that such subsequent developments have a definite seasonal trend with frequency maxima in certain rather narrowly limited portions of the year, just as have the types of *Grosswetterlagen* and the repetition tendency. Therefore, only those days of previous years should be used for comparison whose date is not more than five days removed from the day in question. As a consequence of this restriction, really useful and significantly similar cases can be found only on rare occasions. This method can be used regularly only when weather map series for 150 yr are available.

Combination of Synoptics and Statistics. The only hopeful method of obtaining reliable medium-range weather forecasts is the combination of synoptics and statistics. By synoptics we mean here not only the chartwise representation and analysis of weather and *Grosswetterlagen*, but also the thorough consideration of subsequent developments from a physical viewpoint. For this purpose, it is necessary to develop the present synoptics, from which forecasts can be made only up to 36 hr in advance, to *Grosswetter* synoptics. Such *Grosswetter* synoptics deal with the physical aspects of the sequence of *Grosswetterlagen*, the forms of transition from one type to another, and the mutual influence of *Grosswetterlagen* in neighboring circulation regions. With such *Grosswetter* synoptics, the preceding development which has led to a certain *Grosswetterlage* must be taken into account to a much greater extent than has been the case with daily synoptics.

The multitude of the phenomena and problems that must be taken into account make it absolutely necessary that statistics be used as a broad empirical basis. What a theoretically designed experiment is to the physicist, clear statistics are to the *Grosswetter* investigator. Investigations of special situations which are favored in the usual synoptics give results which only appear to be proofs and have no value in promoting knowledge of macrometeorology. Such investigations can be compared to experiments in physics in which uncontrollable secondary effects are not eliminated. Clean-cut statistics must be preceded by definite physical or theoretical formulations of the problems; these statistics must comprise all available observations pertaining to the problem and they must be compared with suitable control samples to show whether or not the correlations are statistically significant. Such investigations can serve a double purpose: furthering our knowledge in a manner similar to a well-designed experiment as well as serving as a basis for forecasting, provided their results are unequivocal and significant (with a significance level of 0.27 per cent).

Forecasting the Sea-Level Pressure Distribution for Two to Three Days. The combination of synoptics and statistics makes it possible today to predict in many cases the approximate pressure distribution for the second and third day in those regions for which the necessary data are available. Scherhag [49] has shown how to determine the sea-level pressure distribution for the following day from the movement and intensity change of 3-hr and 24-hr isallobars, from the direction of the isopleths of geopotential, and finally from the divergence and convergence of the isopleths on the 500- and 225-mb charts. Furthermore, one can compute approximately the change in thickness of the layer between the 500- and 1000-mb surfaces from the surface pressure distribution; by adding the prognostic thickness chart to the prognostic surface pressure chart, the prognostic 500-mb chart for the following day can be obtained. In turn, from this 500-mb chart a forecast of the surface pressure distribution can be obtained for the second consecutive day. Finally, a prognostic chart of the surface pressure for the third consecutive day can be obtained by repeating this procedure with a consideration of the "steering" of the isallobaric maxima and minima in the upper troposphere by means of the 96-mb chart [17, p. 126]. Of course, the forecast becomes more uncertain with each successive day because of the propagation of errors. For this reason, it is necessary to supplement the prognosis statistically.

If we transform the hydrodynamic equation of motion so that it includes surface friction, and add to it the equation of continuity, the first law of thermodynamics, and the equation of state, a system of equations is obtained from which the following equation for the pressure change can be derived, as shown in [24]:

$$\frac{\partial p}{\partial t} = \frac{p}{RT}\left[\frac{\kappa}{1-\kappa}\frac{dq}{dt} + \left(gv_z - \frac{RT}{1-\kappa}\frac{\partial v_z}{\partial z}\right) + v\left(\frac{dv}{dt} + cv\right)\right] - \frac{p}{1-\kappa}\,\mathrm{div}_2\,\mathbf{v},$$

where the symbol $\mathrm{div}_2\,\mathbf{v}$ stands for the *horizontal* divergence of the velocity $\mathbf{v}$, $\kappa = R/c_p = 0.2884$, dq/dt is

the amount of heat received per unit time by a unit mass of a moving particle, c is a constant depending on external friction, and the other symbols have their customary meanings.

The replacement of the terms on the right-hand side of this equation by quantities which can be statistically interpreted—for these quantities the reader is referred to the original report [10]—leads to the following variables which are used as headings in a table of multiple correlations:

X_1 = pressure in mb reduced to sea level, on the clue day (day of forecast) at 1400 (4 classes),
X_2 = 24-hr temperature change in centigrade degrees from the previous day at 1400 to the clue day at 1400 (4 classes),
X_3 = 24-hr pressure change in mb from the previous day at 1400 to the clue day at 1400 (4 classes),
X_4 = cloudiness in tenths at 1400 on the clue day (4 classes),
X_5 = 7-hr pressure change in mb from 0700 to 1400 on the clue day (5 classes).

In Table VI, a section from a multiple-correlation table is given whose headings include these five variables. The numbers in the blocks are the 48-hr pressure changes from 1400 of the day of the forecast to 1400 two days later, as well as the date of the clue day (in parentheses). All observations refer to Potsdam (near Berlin). The table is based on observations from June 21 to August 9 of the years 1893–1937. Of the 1280 blocks of the table, 698 are covered with a total of 2250 entries. The boldface numbers of the table represent the mathematical expectation of the 48-hr pressure changes, in mb. Of course, in practice, only those squares are to be used that are most likely to lead to an unequivocal result, as is the case with the blocks of Table VI, marked by a black star. If one block contains a pressure change whose sign is different from all the other values of that block, this possibility could be excluded if the weather situation on the day in question (for instance July 12, 1928, in Table VI) were entirely different from that on the day of the forecast. For such comparisons, the method of similar (or, in this example, opposite) cases can be applied even at the present time.

The same table can be used for several climatically coherent places. However, different tables must be computed for different climatic regions and different seasons. The tables can be supplemented by the meridional and zonal pressure gradients or other physically important quantities. This, however, is possible only if observations over a longer period at several climatically similar stations can be combined in one table, such that the number of blocks, which increases with every new variable, can be filled with sufficient observations.

On days when sufficiently reliable results cannot be obtained from a multiple-correlation table, it is better not to draw prognostic pressure charts for the second and third consecutive days.

Mean-Circulation Methods of Forecasting. A judicious combination of synoptics and statistics is represented in the American investigations which can be classed under the name of "mean-circulation methods of extended forecasting" [32, 40, 41, 45, 53, 57]. With respect to the nature and results of these methods, the reader is referred to the article by Namias in this Compendium.[3] Only a few critical remarks will be made in

Table VI. Extract from a Multiple-Correlation Table for the Prediction of 48-hr Pressure Changes at Potsdam
(X_1 = 1019.5 to 1036.5 mb)
(X_2 = +4.0 to +11.8 cent. deg.)
(X_3 = 0.0 to 3.2 mb)

X_5	X_4	X_0
+6.8 to +1.3	0 to 4	—
	5 to 7	—
	8 to 9	—
	10	—
+1.2 to 0.0	0 to 4	−2.8 (8-9-01) +4.8 (7-4-23) −0.5 (8-2-26) **+0.5**
	5 to 7	−1.5 (8-9-14)
	8 to 9	+0.4 (7-9-14) −4.0 (7-11-28) **−1.8**
	10	—
−0.1 to −1.2	0 to 4	−2.8 (8-8-93) +0.4 (7-12-28) −2.7 (7-20-99) −4.8 (7-11-29) −0.8 (7-7-04) −14.8 (7-23-31) −3.3 (6-27-08) −5.3 (7-9-32) −7.3 (6-30-14) −13.5 (7-17-34) −4.1 (7-15-21) **−5.4** ★
	5 to 7	−4.9 (8-8-05) −0.3 (7-7-33) **−2.6**
	8 to 9	+1.7 (7-1-06) −6.0 (8-3-07) **−2.1**
	10	—
−1.3 to −2.5	0 to 4	−11.7 (6-23-97) −10.7 (6-26-24) −1.2 (7-23-08) −10.7 (7-26-27) −2.9 (7-3-17) −12.5 (7-4-32) **−8.3** ★
	5 to 7	−5.6 (6-23-20)
	8 to 9	−12.9 (6-22-11)
	10	—
−2.6 to −7.0	0 to 4	−7.2 (7-21-23)
	5 to 7	—
	8 to 9	—
	10	—

3. Consult "General Aspects of Extended-Range Forecasting" by J. Namias, pp. 802–813.

order to include these valuable investigations in the total structure of macrometeorology.

The most important aspect of these investigations is the world-wide point of view of the general circulation, on which they are based. This is *Grosswetter* synoptics in the true sense of the term. This method of consideration, comprising the entire Northern Hemisphere, is also applied in Rossby's formula [45]:

$$c = U - \frac{\beta L^2}{4\pi^2},$$

where c is the eastward velocity of the perturbation (trough or ridge), L is the wave length, U is the zonal velocity of the westerlies, and β is the rate of change of the Coriolis parameter with latitude. This formula (as derived from simplifying assumptions) is of such fundamental importance for the mean circulation methods that it is necessary to conduct rigorous statistical investigations to determine the extent to which it is confirmed by observation, and the conditions under which systematic deviations occur which can be eliminated by computation.

Clapp [41] made such an investigation using five-day mean 10,000-ft charts covering the period from January 1941 to April 1943. He computed the correlation coefficients between the observed displacements (for half a week and for a full week) of troughs and ridges at 45°N and those computed from Rossby's formula. Clapp used as a representative value of U, the zonal westerlies, the mean west to east geostrophic wind between 40°N and 50°N over the entire range of longitudes from about 130°W to 30°W, known as the 10,000-ft index I_3. He furthermore defined the wave length L as twice the difference in longitude between a given trough and the ridge to the west of it. Eliminating cases of new trough development, Clapp obtained the following correlation coefficients r in N cases:

	Full week		Half week	
	r	N	r	N
Winter	*0.66*	37	*0.67*	36
Spring	0.44	35	*0.51*	35
Summer	0.35	29	0.42	29
Fall	0.42	35	*0.63*	35

For the displacements during a full week the correlation is better than chance (with a significance level of 0.27 per cent) only in winter; however, for displacements during half a week, three of the four correlation coefficients are significant. We can conclude from this, that the Rossby formula, in spite of the simplifications made in its derivation, furnished an approximation of actual conditions. However, the correlations are too small to permit predictions of the displacement of troughs and ridges on the basis of this formula. For this reason, Namias [40] employed additional criteria and methods for the forecasting of circulation patterns.

According to Rossby's formula, a westward displacement (retrogression) of a trough or a ridge occurs when the wave length, that is, the distance between troughs, exceeds a certain value depending on the season and the geographical latitude; in other words, when an increased zonal circulation exists. This result is in apparent contradiction with the concept prevailing among central European investigators, according to which a westward displacement of steering centers, particularly high-pressure centers, at the 5-km level is to be expected for physical reasons, notably when a pronounced meridional circulation prevails on both sides of the steering center [47]. Nevertheless, both concepts are mutually consistent. It is entirely possible that a strong meridional circulation on both sides of a ridge at high altitudes is associated with a large distance between pressure ridges (Fig. 12*a*), whereas for a shorter

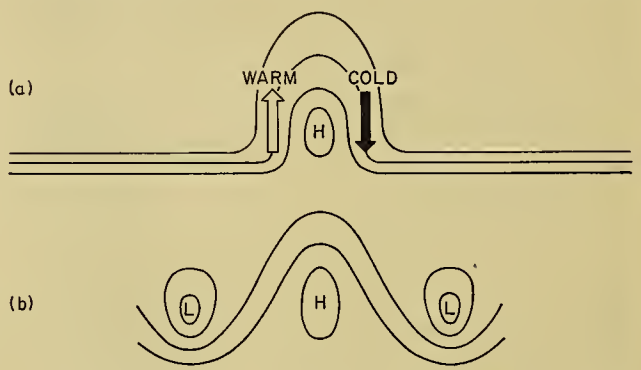


FIG. 12.—Schematic representation of the 500-mb contour lines with (*a*) alternating strong zonal and strong meridional circulation on both sides of a ridge, and large distance between ridges, and (*b*) prevailing meridional circulation, but not so concentrated locally and not so strong as in (*a*) and small distance between ridges.

distance the strong meridional circulation will, perhaps, occur less frequently or not at all (Fig. 12*b*). Further investigations are needed to establish the relative frequency with which a strong meridional circulation on both sides of a steering center in the upper troposphere is associated with an intensified zonal circulation at a relatively large distance from this steering center.

Observation of the large-scale meridional and zonal temperature contrasts is of great importance for the theory as well as for the practice of extended-range forecasting. Namias [39] has called attention to the very high correlation between the meridional temperature difference (35°N–55°N) and the west-east component of the wind speed at the 3-km level. Scherhag [48] has shown by a selected example that, even for monthly means, a strong negative pressure anomaly near Iceland corresponds to a strong meridional temperature gradient over the Great Lakes region of North America. Both these results are in agreement with the high correlation obtained by Baur [6] of the monthly mean temperature difference between New England and eastern Greenland with the simultaneous monthly mean pressure difference between the Azores and Iceland, for nearly all months of the year. For January, the correlation coefficient for the mean monthly temperature difference between New Haven and ½ (Jacobshavn + Upernivik) and the simultaneous mean monthly pressure difference between Ponta Delgada and Stykkisholm amounts to +0.61. However, this

correlation of the monthly means can be interpreted contrariwise: When the pressure over Iceland is very low, western Greenland lies within a cold northern flow, whereas to the west of the above-normal Azores High, warm air flows from the south to the north. In turn, with a strong air-mass exchange between high and low latitudes the temperature difference will diminish. The general circulation shows, after a certain duration of intensified air movement, "fatigue phenomena" which finally lead to a change in the circulation pattern (see Table II and [7]).

In order to develop circulation methods so that they can be used as a basis for reliable extended forecasts; a transition is necessary from the statistics of averages, now predominantly in use, to statistics of selected cases which take into account especially the physical processes, and the special characteristics of the particular case. Examples of "selective statistics" are contained in Tables I and V, which, to some extent, form parts of a multiple-correlation table. Table VI also forms part of "selective statistics."

Long-Range Forecasting. Forecasts for more than five days are not possible on a synoptic basis. The multitude of all possible developments no longer permits consideration of the development in all physical details. Here, statistics must be employed if we are to make statements about the future. However, two prerequisites must be fulfilled: (1) the choice of statistical methods must be guided by a physical formulation of the problems and by physical considerations, and (2) long-range forecasts should be given, at least publicly, only if better-than-chance statistical relationships are obtained that ensure the success of the forecast with a probability of over 92 per cent. It must be borne in mind that the damage which might result from an incorrect forecast is greater the longer the time interval for which the forecast is made. Therefore all demands for regular issues of monthly and seasonal forecasts must be declined in the present state of our knowledge. In order to be on the safe side, it is advisable to issue long-range weather forecasts only if at least two arguments are available that fulfill the above-mentioned prerequisites and lead to the same expected weather character without contradiction by any other statistical argument.

Extrapolation of Periods. The extrapolation of periods is more helpful for long-range forecasts than for medium-range forecasts. Tests of over a thousand harmonic analyses at the *Forschungsinstitut für langfristige Witterungsvorhersage* over a number of years have shown, however, that the percentage of correct forecasts is too small for these extrapolations to be used as a basis for long-range forecasts. Even the use of rigorous mathematical criteria permits only the decision as to whether the period has been persistent to date, but it does not yield any clues to its future behavior. Rhythm analyses cannot be used as the basis for any meteorological forecasts as long as we have no physical, statistically sound criteria that would indicate the continuation of the period with an expectancy of over 92 per cent during the forecast period.

Symmetry Points. There are even stronger objections against the use of so-called symmetry points for the pressure curve. In the first place, it was shown that even in a model sample based on chance and persistence tendency only, symmetry points occur with the same degree of approximation as they do for the actual pressure curve [20, pp. 924–926 and 934–936]. In the second place, the existence of a symmetry point can be detected with some degree of certainty only if about 20 days have elapsed. In the third place, even from a very good symmetry during 20–25 days, no sufficiently reliable conclusion can be drawn as to whether and how long the reflected curve of the pressure or any other weather element will persist. As an example, the symmetry point for the pressure curve of Potsdam on December 26, 1932, may be cited. The correlation coefficient of corresponding days from the first to the 25th day before and after the symmetry point was $+0.75$; from the 26th to the 50th day it was -0.46.

Regression Equations. It is necessary that regression equations which are to be used for long-range forecasts yield a multiple correlation coefficient greater than 0.80, since only then will the standard error of estimate of the correlation be less than $\frac{6}{10}$ of the standard deviation. Furthermore, the basic single correlations must be stable, that is, they must be approximately the same for subdivisions of the total period.

Multiple Correlation Tables. Since most of the *Grosswetter* correlations are not linear, it is very rare that multiple-correlation coefficients are obtained for sufficiently long periods which are larger than 0.80. Selective statistics again appear much more promising in this case, since fundamentally they yield parts of multiple correlation tables. As long as their computation is guided by a physical approach to the problem, unequivocal, better-than-chance results can be used as an empirical basis for the further development of the theory as well as a basis for long-range forecasts.

REFERENCES

1. Abbot, C. G., *Ann. astrophys. Obs. Smithson. Instn.*, Vol. 6. Washington, D. C., 1942. (See Table 27) (Also, personal communications through December 1945.)
2. —— and Fowle, F. E., "Volcanoes and Climate." *Smithson. misc. Coll.*, Vol. 60, No. 29, 24 pp. (1913).
3. Alter, D., "Application of Schuster's Periodogram to Long Rainfall Records, Beginning 1748." *Mon. Wea. Rev. Wash.*, 52:479–483 (1924).
4. Ångström, A., "Teleconnections of Climatic Changes in Present Time." *Geogr. Ann., Stockh.*, 17:242–258 (1935).
5. Baur, F., "The 11-Year Period of Temperature in the Northern Hemisphere in Relation to the 11-Year Sun-Spot Cycle." *Mon. Wea. Rev. Wash.*, 53:204–207 (1925).
6. —— "Statistische Untersuchungen über Auswirkungen und Bedingungen der grossen Störungen der allgemeinen atmosphärischen Zirkulation, III." *Ann. Hydrogr., Hamb.*, 54:227–236 (1926).
7. —— "Der gegenwärtige Stand der meteorologischen Korrelationsforschung." *Meteor. Z.*, 47:42–52 (1930).
8. —— "Die Formen der atmosphärischen Zirkulation in der gemässigten Zone." *Beitr. Geophys.*, 34:264–309 (1931).
9. —— "Schwankungen der Solarkonstante." *Z. Astrophys.*, 4:180–189 (1932).

10. —— "Die Störungen der allgemeinen atmosphärischen Zirkulation in der gemässigten Zone." *Meteor. Z.*, 54: 437–444 (1937).
11. —— "Zur Frage der Beziehungen zwischen der Temperatur des Golfstroms und dem nachfolgenden Temperaturcharakter Mitteleuropas." *Meteor. Z.*, 54:188–189 (1937).
12. —— "Über die grundsätzliche Möglichkeit langfristiger Witterungsvorhersagen." *Ann. Hydrogr., Hamb.*, 72: 15–25 (1944).
13. —— *Musterbeispiele europäischer Grosswetterlagen.* Wiesbaden, Dieterich, 1947.
14. —— "Ein Beitrag zur Klärung der Frage eines Einflusses der Planeten auf die Witterung." *Z. Meteor.*, 2:47–55 (1948).
15. —— "Was kann die Wissenschaft verantworten, über das voraussichtliche Witterungsgepräge des kommenden Winters zu sagen?" *Naturwiss. Rdsch.*, 1:256–261 (1948).
16. —— "Zur Frage der Echtheit der sogenannten Singularitäten im Jahresgang der Witterung." *Ann. Meteor.*, 1:372–378 (1948).
17. —— *Einführung in die Grosswetterkunde.* Wiesbaden, Dieterich, 1948.
18. —— "Die doppelte Schwankung der atmosphärischen Zirkulation in der gemässigten Zone innerhalb des Sonnenfleckenzyklus." *Meteor. Rdsch.*, 2:10–15 (1949).
19. —— "Zurückführung des Grosswetters auf solare Erscheinungen." *Arch. Meteor. Geophys. Biokl.*, (A) 1:358–374 (1949).
20. —— "Die Erscheinungen des Grosswetters," *Lehrbuch der Meteorologie*, J. v. Hann und R. Süring, Hsgbr., Bd. 2, 5. Aufl. Leipzig, W. Keller, 1951.
21. —— "Winter Temperature in New England Following Anomalies of Circulation over the North Atlantic in November." *Bull. Amer. meteor. Soc.*, (1951) (in press).
22. —— u. a., *Mitteleuropäischer Witterungsbericht.* Bad Homburg, Forschungsinst. f. langfristige Witterungsvorhersage, 1937–39. (See also "Der Mitteleuropäischer Witterungsbericht." *Meteor. Z.*, 55:142–147 (1938).)
23. Baur, F., und Philipps, H., "Der Wärmehaushalt der Lufthülle der Nordhalbkugel im Januar und Juli und zur Zeit der Äquinoktien und Solstitien." *Beitr. Geophys.*, 42:160–207 (1934).
24. —— "Die Bedeutung der Konvergenzen und Divergenzen des Geschwindigkeitsfeldes für die Druckänderungen." *Beitr. Phys. frei. Atmos.*, 24:1–17 (1938).
25. Berlage, H. P., Jr., "Über die Verbreitung der dreijährigen Luftdruckschwankung über die Erdoberfläche und der Sitz des Umsteuerungsmechanismus." *Meteor. Z.*, 50:41–47 (1933).
26. Bernheimer, W. E., "Über den angeblichen Zusammenhang der Sonnenstrahlung mit der Fleckenhäufigkeit." *Meteor. Z.*, 47:190–191 (1930).
27. Best, N., Havens, R., and LaGow, H., "The Pressure and Temperature of the Upper Atmosphere," in *Upper Atmosphere Research Report IV.* Naval Res. Lab., Washington, D. C., 1947. (See pp. 111–115)
28. Bjerknes, J., "On the Structure of Moving Cyclones." *Geofys. Publ.*, Vol. 1, No. 2 (1919); Bjerknes, V., "On the Dynamics of the Circular Vortex with Applications to the Atmosphere and Atmospheric Vortex and Wave Motions." *Geofys. Publ.*, Vol. 2, No. 4 (1921); Bjerknes, V., "On Quasi Static Wavemotion in Barotropic Fluid Strata." *Geofys. Publ.*, Vol. 3, No. 3 (1923).
29. Bjerknes, J., and Solberg, H., "Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation." *Geofys. Publ.*, Vol. 3, No. 1 (1922).
30. Bjerknes, V., and others, *Physikalische Hydrodynamik.* Berlin, J. Springer, 1933.
31. Defant, A., "Die Schwankungen der atmosphärischen Zirkulation über dem Nordatlantischen Ozean im 25-jährigen Zeitraum 1881–1905." *Geogr. Ann., Stockh.*, 6:13–41 (1924).
32. Dorsey, H. G., Jr., and Brier, G. W., "An Investigation of a Trajectory Method for Forecasting Flow Patterns at the 10,000-Foot Level." *Res. Pap.* No. 8, pp. 50–58 in *A Collection of Reports on Extended Forecasting Research.* U. S. Weather Bureau, Washington, D. C., 1944.
33. Exner, F. M., *Dynamische Meteorologie*, 2. Aufl. Wien, J. Springer, 1925.
34. Humphreys, W. J., "Volcanic Dust and Other Factors in the Production of Climatic Changes, and Their Possible Relation to Ice Ages." *J. Franklin Inst.*, 176:131–172 (1913).
35. Kiepenheuer, K. O., "Neuere Ergebnisse über die Sonnenkorona." *Neue Phys. Blätter*, 1946. (See pp. 225–231)
36. Lexis, W., *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft.* Freiburg, F. Wagner, 1877; also Rietz, H. L., ed., *Handbook of Mathematical Statistics.* Boston, Houghton, 1924. (See pp. 82–88)
37. Maris, H. B., and Hulburt, E. O., "A Theory of Auroras and Magnetic Storms." *Phys. Rev.*, (II) 33:412–431 (1929).
38. Meinardus, W., "Über einige meteorologische Beziehungen zwischen dem Nordatlantischen Ozean und Europa im Winterhalbjahr." *Meteor. Z.*, 15:85–105 (1898); "Der mitteleuropäische Winter und seine Beziehungen zum Golfstrom." *Das Wetter*, 16:8–14 (1899).
39. Namias, J., "Physical Nature of Some Fluctuations in the Speed of the Zonal Circulation." *J. Meteor.*, 4:125–133 (1947).
40. —— *Extended Forecasting by Mean Circulation Methods.* U. S. Weather Bureau, Washington, D. C., 1947.
41. —— and Clapp, P. F., "Studies on the Motion and Development of Long Waves in the Westerlies." *J. Meteor.*, 1:57–77 (1944).
42. Norton, H. W., "Estimating the Correlation Coefficient." *Bull. Amer. meteor. Soc.*, 27:589–590 (1946).
43. Pettersson, O., "Ueber die Beziehungen zwischen hydrographischen und meteorologischen Phänomenen." *Meteor. Z.*, 23:285 (1896).
44. Rossby, C.-G., "On the Distribution of Angular Velocity in Gaseous Envelopes under the Influence of Large-Scale Horizontal Mixing Processes." *Bull. Amer. meteor. Soc.*, 28:55–68 (1947).
45. —— and Collaborators, "Relation between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action." *J. mar. Res.*, 2:38–55 (1939).
46. Schell, I. I., "Dynamic Persistence and Its Applications to Long-Range Foreshadowing." *Harv. meteor. Studies*, No. 8, 80 pp. (1947).
47. Scherhag, R., *Neue Methoden der Wetteranalyse und Wetterprognose.* Berlin, Springer, 1948.
48. —— "Synoptische Untersuchungen über die Entstehung der atlantischen Sturmwirbel." *Meteor. Z.*, 54:466–469 (1937).
49. —— "Die Verwendung der Höhenkarten im Wetterdienst." *Forsch. ErfahrBer. Reichs. Wetterd.*, Reihe B, Nr. 11 (1943).
50. Schmauss, A., "Singularitäten im jährlichen Witterungsverlauf von München." *Dtsch. meteor. Jb. Bayern*, B (1928); "Kalendermässige Verankerungen des Wetters."

Meteor. Z., 53:72–74 (1936); "Synoptische Singularitäten." *Meteor. Z.*, 55:385–403 (1938).

51. Schott, G., *Geographie des Atlantischen Ozeans*, 3. Aufl. Hamburg, C. Boysen, 1942.
52. Seilkopf, H., "Speziélle Grosszirkulatiou und Witterung." *Ann. Meteor.*, 1:312–325 (1948).
53. University of Chicago, Dept. Meteor., Staff Members, "On the General Circulation of the Atmosphere in Middle Latitudes." *Bull. Amer. meteor. Soc.*, 28:255–280 (1947).
54. Wagner, A., *Klimaänderungen und Klimaschwankungen*. Braunschweig, F. Vieweg & Sohn, 1940. (See pp. 192–199)
55. Waldmeier, M., *Ergebnisse und Probleme der Sonnenforschung*. Leipzig, Akad. Verlagsges., 1941. (See pp. 118–120)
56. Wiese, W., "Die Bedeutung der Eisverhältnisse im Frühling im Grönländischen Meere und des Ost-Isländischen Polarstromes für die Temperaturverhältnisse des nachfolgenden Winters in Europa." *Bull. Inst. hydrol. Russ.*, 14:52–59 (1925); "Eis im Barents-Meer und Lufttemperatur in Europa." *Nachr. ZentBur. Hydrometeor., UdSSR*, 3:1–30 (1924); "Ice in the Polar Seas and the General Circulation of the Atmosphere." *J. Geophys. Meteor., Moscow*, 1:78–84 (1924).
57. Willett, H. C., and others, *Final Report of the Weather Bureau—M.I.T. Extended Forecasting Project for the Fiscal Year 1948–1949*. Cambridge, Mass., 1949.
58. Wünsche, W., "Über die Existenz langsamer Luftdruckschwingungen auf der rotierenden Erde." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 11:153–199 (1938).

EXTENDED-RANGE FORECASTING BY WEATHER TYPES

By ROBERT D. ELLIOTT

North American Weather Consultants

INTRODUCTION

During the last two decades there has been a considerable development in the field of extended-range forecasting, that is, forecasts for up to a week in advance. While it is known that a modest success in forecasting general weather trends for as much as seven days in advance has been achieved by several groups using techniques of their own development, no particular method has gained the widespread acceptance in meteorological circles enjoyed by the frontal and air-mass analysis techniques of short-range forecasting. However, most of the extended-range forecasting methods have a number of points in common and these will be discussed briefly.

Meteorologists engaged in effective extended-range forecasting must become accustomed to dealing with certain forms of representation which summarize the circulation characteristics of a given region over a finite period of time. This may take the form of fixed-time-interval mean charts such as the overlapping five-day mean maps used by the Extended Forecast Section of the U. S. Weather Bureau [20], or more variable and "natural" period representations such as composite maps [21], weather types [3], or steering patterns [26]. In each case the representation involves, basically, the concept of mean circulation patterns at the surface and/or aloft, covering a fixed or variable time interval of several days or more. Even a pure analogue selection technique must involve considerations of this nature in order to be successful. One-map "thumbprint" selection schemes have not proved successful in extended-period forecasting, and exhaustive statistical tests by Darling [6] show that persistence is the principal factor taken into account by this method.

Ordinarily, the time interval is chosen in such a manner that the local circulation patterns of simple migratory and transitory cyclones and anticyclones are smoothed out, thus delineating the more basic quasi-stationary and broad-scale features of the circulation pattern which are responsible for the longer-period weather characteristics in a given region. The mean patterns thus revealed are characterized by large-scale cyclones and anticyclones at the surface and by an undulatory pattern aloft having longer wave lengths than those observed on daily upper-level charts. The movements and changes in form of these patterns are on the whole slower than those of the individual migratory systems making up the mean, which at least partly compensates for the difficulty of having to forecast so far in advance. In this country, methods for forecasting these movements and changes in form are based upon statistical weather-type prediction schemes [8], the application of the conservation of vorticity principle [20], and a number of other techniques which will be considered later.

Once the extended-period mean or type pattern is forecast, the anomalies[1] for the weather elements for the time interval in question can be predicted. Considerable statistical work has been done relating temperature and precipitation anomalies for an extended period with the weather types [3] and with the mean isobaric pattern at upper levels [17, 20]. In general, extended-period surface temperature anomalies are negative behind mean troughs aloft and positive ahead, with precipitation and cloudiness at a maximum just ahead of the mean trough.

The problem of preparing daily prognostic charts for up to seven days in advance is considerably more difficult. It is of course obvious that the sum of the daily prognostic charts should correspond to the extended-period prognostic representation; however, there are a great variety of ways in which this pattern may be made up. A prognostic extended-period upper-level chart may be used as a "steering" pattern to fix in a general way the paths of migratory cyclones and anticyclones, the individual systems being extrapolated forward from forecast day with due consideration being given to the predicted strength of the flow aloft. In one technique this is supplemented by various prognostic zonal and meridional indices [20]. Another technique [4] resorts to the use of analogues selected from the archives of weather charts in which the mean flow pattern matches the predicted pattern. Still another method is to use "ideal" weather types [3] giving the characteristic daily surface charts for each day of the weather types as determined through statistical study of past cases.

The foregoing paragraphs very briefly outline the features which most extended-range forecasting techniques have in common. These may be summarized as follows:

1. A method for concise representation of circulation characteristics during an extended period.
2. A method for forecasting movement and changes in form of circulation features delineated in 1, thus permitting the preparation of a prognostic representation.
3. A method for forecasting extended-period anomalies of the weather elements from information provided by 2.
4. A method for preparing daily prognostic charts and forecasts for daily weather elements from 2.

1. The term "anomaly," as used in this article, refers to departure from a long-term average over time and has nothing to do with the departures from means computed for latitude circles.

Although some investigators who are inclined toward statistical analysis advocate the elimination of one or more of the intermediate steps by going directly from the representation scheme to the forecast of the weather elements themselves, in the opinion of the author our present knowledge concerning weather processes is so limited that the use of the intermediate steps is required in order that the forecaster can grasp more readily the physical significance of the processes involved.

A general discussion of the California Institute of Technology weather-type technique is presented in the following section.

THE WEATHER-TYPE APPROACH

Various sorts of weather types had been developed for different areas in the past [2, 14, 23]. Often the earlier attempts at typing were based upon single typical synoptic charts and therefore did not truly take into account weather processes occurring during an extended period. After some preliminary investigation during the late thirties [10, 18] a six-day North American weather-type scheme was developed wherein the type itself embraced a sequence of days characterized by certain locations and orientations of the semipermanent elements of circulation such as the Pacific anticyclone, the Aleutian cyclone, the trajectories of polar outbreaks, and cyclone paths. Empirical evidence was discovered which indicated that the mean period between the passages of cyclone families past a given point was about three days and that about two-thirds of the time the type characteristics associated with the passage of a given family occurred also in either the preceding or the following cyclone family. Because of this, the types were arranged so as to have a lifetime in any given sector or region of six days plus or minus one day. The instances where the type characteristic was not maintained during the passage of two successive cyclone families were handled by the introduction of the so-called "complex" types in which the first three days had the characteristics of one type and the last three those of another. Considerable work was done on these types over a period of several years, resulting in the development of a weather-type catalogue covering the period 1921–42 [22] and a set of "ideal" types [3] showing the most probable location of frontal systems and centers on successive days of each type, along with other statistical information on the weather element anomalies associated with these "ideal" types.

Although these types proved of value, continued study indicated a number of shortcomings, the most important of which seemed to be that forcing the types into a six-day lifetime seemed somewhat arbitrary. Therefore, during World War II, in response to a request from the Army Air Force, which was interested in extended-range forecasting techniques and had contracted with the California Institute of Technology for continued research on the subject, new three-day types [4] were developed wherein the life of the type corresponded to the passage of a single cyclone family across a region. This averaged three days, with considerable variation. The three-day weather types included all the essential and well-tested features of the older six-day types plus a number of refinements found necessary as a result of experience in their application. They were developed for each of four synoptic regions consisting of sectors of forty-five degrees of longitude extending along the belt of westerlies from 135°E through 180° to 45°W and a fifth region extending from 45°W to 45°E. Some twenty to thirty different types were determined for each region and a catalogue of weather types for forty-six years was prepared [4, 5, 15].

It was early recognized that the large-scale and persistent distortions of the upper-level pattern controlled and determined the weather types in each region. With an increasing file of adequate upper-level charts it has been possible to establish in a reliable fashion the upper-air patterns associated with each weather type [19]. Because this file does not extend far back historically, it has not been possible to employ upper-air type characteristics in cataloguing, except in recent years.

It had early been noted that the weather types for each region could be divided into two distinctive categories and that types in the same category often occurred simultaneously in neighboring regions. These two categories were the meridional and zonal flow types. Meridional flow types are characterized in the mean by upper-level crests and troughs of large amplitude which tend to steer the surface migratory systems along paths far to the north or south of normal, depending upon their location. Most of the large polar outbreaks which penetrate to southerly latitudes occur in this category. Elsewhere, unseasonably warm weather is experienced. Meridional flow patterns are particularly interesting because of the conspicuous anomalies of temperature associated with them. An interesting example may be cited in this connection: during the entire month of January, 1949, in North America there existed an extremely strong and persistent trough aloft in western United States with a persistent crest in eastern United States. As a result, temperatures were far below normal in the West and above normal in the East.

Zonal flow types are characterized by zonal flow aloft. Variations in the strength and latitude of the belt of westerlies lead to variations within this category.

It is beyond the scope of this article to present the details of weather types as delineated by daily or "ideal" charts. However, a set of schematic diagrams representing the major features of the mean flow pattern occurring with a number of the more important wintertime weather types of the North American area is shown in Fig. 1. The types are arranged in two columns; in the left-hand column are meridional flow types, in the right-hand column zonal flow types. In addition, the position of each type in its column is determined by the degree to which it displays the characteristics of the category. In the case of the meridional flow types, the more the principal mean trough over North America is shifted westward, the stronger the meridional flow character of the type. In the extreme of this category the trough is off the west

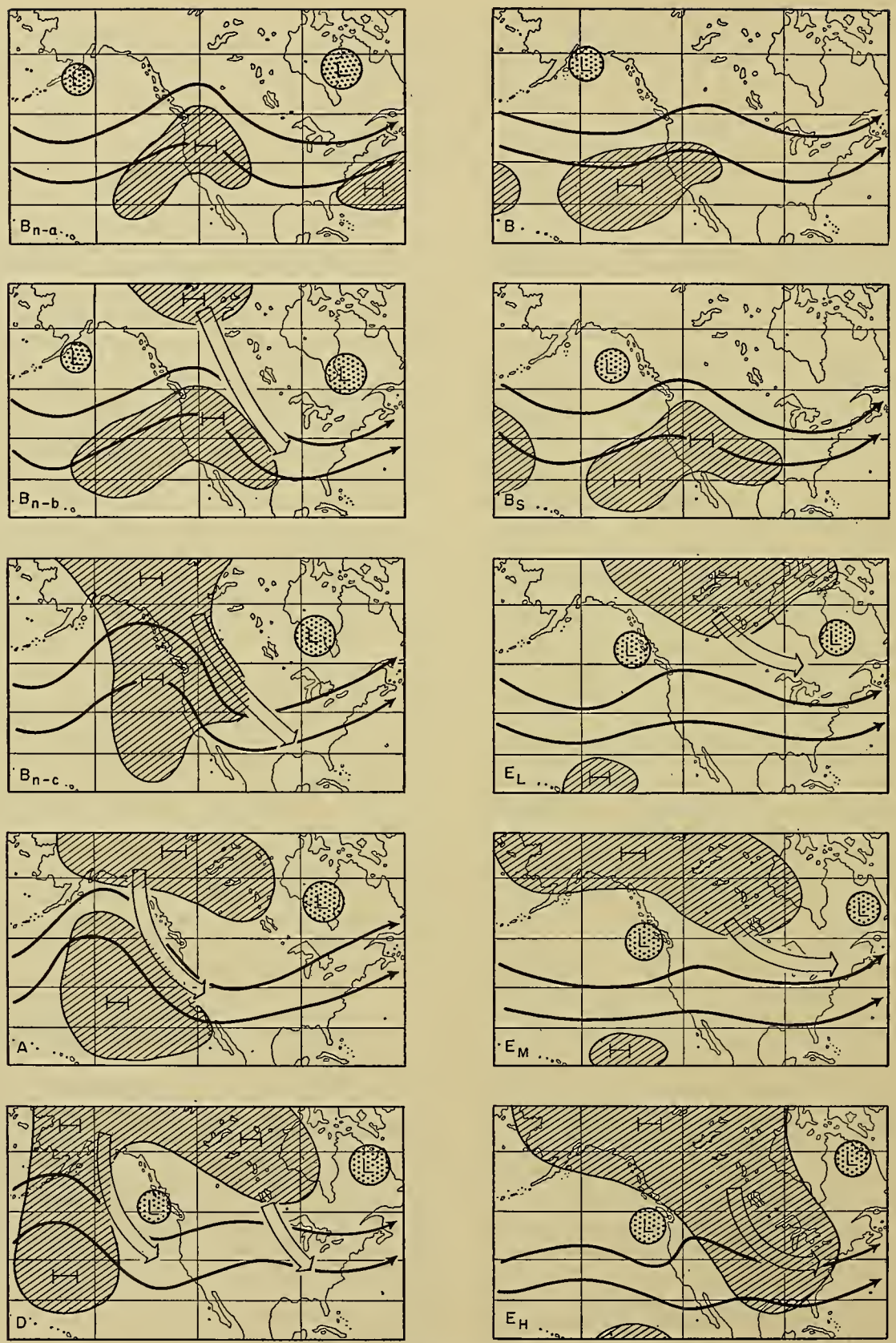


FIG. 1.—Schematic diagrams of weather types. Heavy lines with arrows at ends indicate upper-level mean flow. Hatched areas are regions of persistent subtropical or polar high pressure at the surface. Stippled areas are quasi-stationary low-pressure centers. Open arrows indicate the paths of polar outbreaks.

coast of North America. The depth of the trough as a rule increases with the westward shift. With the zonal flow types, the farther southward the westerlies are shifted over a broad sector, the stronger the zonal flow character of the type. In the extremes of this last category the southward shift occurs over a sector 90° (or more) wide and is accompanied by the formation of extensive polar anticyclones to the north of the principal storm path. This storm path is unusually far south.

In the meridional category, the *Bn-a* type has a definite upper-level trough at about 90°W and a crest over the Great Basin (eastern Washington, Idaho, eastern Oregon, Utah, Nevada, and northern Arizona). Because only minor polar outbreaks occur in this type and because of the presence of a large dynamic anticyclone over the Great Basin which deflects Pacific storms northward, the United States and southern Canada experience relatively warm dry weather. As one proceeds to the *Bn-b* type, it is found that whereas conditions to the west of the Rockies are nearly identical with those for *Bn-a*, in the East a polar outbreak occurs as indicated on the diagram. This tends to make the eastern trough aloft deeper than in the case of *Bn-a* and produces below-normal temperatures east of the Rocky Mountains. In the case of *Bn-c* type, the meridional character of the flow becomes quite marked with a well-developed trough aloft at about 100°W. Frequently a strong dynamic cyclone remains trapped for several days in the southern Great Basin. The polar outbreak, instead of being confined to the area east of the Rocky Mountains as in the case of *Bn-b*, flows westward through mountain passes as well as southward along the eastern front of the Rockies and produces much-below-normal temperatures throughout most of western United States. Ahead of the mean upper-level trough, surface temperatures are considerably above normal, and precipitation is likely to be excessive near the mean trough. It is interesting to note that the Great Basin anticyclone, present in the two preceding types, is now weaker and appears to be merged to the west with the Pacific anticyclone, which has shifted northward. A further westward shift of the mean upper-level trough position is exhibited in type *A*. Here the Basin anticyclone has disappeared from the scene as a semipermanent feature of the circulation at the surface. The eastern lobe of the Pacific anticyclone has moved unusually far north and the polar outbreak associated with this type consists of a mass outflow of cold air moving southward along the Pacific Coast, then southeastward into the Great Basin. In contrast to polar outbreaks over the land areas of North America, this outbreak is characterized by cyclonically curved isobars at the surface. Surface temperatures are below normal along the Pacific Coast and in the Great Basin but are above normal in the East. Because of the interaction of tropical and polar air masses in an area north of normal in the East, precipitation is excessive in the Mississippi and Ohio Valleys. This type and the closely related *Bn-c* type were responsible for the abnormally cold weather in the West and warm weather in the East during the winter of 1948–49 and also during January, 1937. Type *D* represents the most extreme of the meridional flow types. In this case the principal trough lies entirely off the West Coast with an unusually intense, northward-displaced Pacific anticyclone, which is merged with the cold anticyclone over Alaska. The principal outbreak of cold air is so far west that a second minor outbreak occurs downstream in eastern North America.

There exists a somewhat infrequent, yet important additional meridional flow type not shown in Fig. 1. This is the *C* type which resembles the *Bn-a* type except that there is a low-latitude trough aloft just off the southern California coastline. In this way the westerlies are split into two branches, the northernmost branch moving along a crest at about 125°W and the southern branch moving along a trough at this same longitude. The two streams converge downstream. For this reason there are essentially two storm tracks, a northern and a southern track. Very heavy precipitation occurs along the southern track in southern California, Arizona, and New Mexico.

The first of the zonal flow types, *B*, is characterized by a northerly storm track with a belt of well-developed subtropical anticyclones to the south. As a result, relatively dry and warm weather prevails over most of the United States and the southern Prairie Provinces of Canada. In the second zonal flow type, *Bs*, the upper-level westerlies are farther south in the East but in the West they are as far north as in type *B*, thus tending to accentuate the western crest. Although the Great Basin anticyclone is actually more persistent in the case of type *Bs*, it is considered to be more zonal than *B* because the westerlies are farther south on the average over a broad sector and are definitely stronger as reflected by the greater rapidity of movement of migratory systems along the principal storm path. Types E_L, E_M, and E_H are each in turn characterized by a more extensively developed polar anticyclone and a more southerly shift in the westerlies aloft and consequently in the storm tracks. The great masses of cold polar air associated with these types seldom move southward in any form of organized polar outbreak such as occurs with the meridional flow types but merely linger north of the mean frontal zone accentuating the air-mass contrast across the front and leading to excessive precipitation along the mean frontal zone itself. In the extreme E_H type the cold air occupies most of the United States and entirely blocks the eastward progress of Pacific storms beyond the Rocky Mountains.

GENERAL CIRCULATION CONSIDERATIONS

A recent study [13] of broad-scale and long-period weather processes occurring in connection with certain abnormal meridional flow patterns has supplied information providing more justification for the weather-type treatment thus far discussed. Among other things,

this study was aimed at ascertaining the significance of mean flow representations not only with respect to the mean pattern itself but also with respect to the smaller-scale and more transitory disturbances going to make up the mean flow. This involved the application of the turbulence principles of fluid mechanics to the circulation of the westerlies. Similar investigations [9, 24] have been undertaken in the past, but no attempt had been made to apply this concept through a more-or-less statistical treatment of a long record of weather data such as the forty-year Northern Hemisphere series [28]. In this treatment the smoothed-out fluctuations, the migratory and transient cyclonic and anticyclonic systems, were treated as turbulent eddies having the property of being able to transport heat and momentum across large horizontal sheets of the atmosphere. This analysis indicated that the very large amounts of heat picked up from the oceans off the east coasts of the continents [16] are removed from the westerlies downstream through lateral turbulent diffusion so that equilibrium is maintained. In addition it showed that in years when the north-south thermal gradient is in excess of normal, the lateral mixing in the westerlies is excessive, and the cross-westerly heat transport is also excessive. In connection with excessive lateral mixing there occurred an excess of meridional flow types.

The detailed analysis of the eddy motion in the westerlies indicated the existence of a spectrum of eddy periods with a mean of about six days in the heart of the westerlies and slightly higher values on either side. The most frequent value was three days, corresponding to the mean period in the passage of cyclone families as used in weather typing. The longer-period eddies even showed a tendency towards durations which were multiples of the three-day period.

Eddies of very long duration were associated with larger-than-average distortions from the normal flow pattern, with respect to both wave length and amplitude. In many cases these large eddies appeared on the surface charts as quasi-stationary anticyclones in the middle latitudes, with trapped cyclones to the south of them. In the study these were referred to as "blocking highs" because of their action in blocking the normal eastward progress of migratory systems. In cases where such large blocking eddies exist it becomes hardly proper to treat them as eddies in the time and space scales generally involved in the extended-range circulation representation schemes. The smaller and more frequent short-period eddies characterizing true migratory cyclone and anticyclone activity are steered around or completely blocked by these eddies. Therefore the large eddies merely serve to introduce large-scale distortions of the normal westerly flow on extended-range charts. In effect, these eddies tend to lengthen the closed path of the westerlies around the hemisphere and there is thus provided a longer perimeter along which the smaller eddies can act to transport heat and momentum across the westerlies. This situation corresponds to that existing in the meridional flow category discussed above. These types must then be those for which the cross-westerly heat transport is at a maximum, and the zonal category corresponds to the case where the girdle of westerly winds has a minimum length, at least in the regions where zonal types are in existence, and therefore represents a state of minimum cross-westerly heat transport.

It is often apparent that there is a certain time lag between the development of meridional flow types in various regions of the Northern Hemisphere. This observation coupled with the known quasi-periodic alternation between zonal and meridional flow types within one type region suggests that under ordinary circumstances the zonal flow types are incapable of providing the required cross-westerly heat transport needed to balance global radiational effects, whereas the meridional flow patterns are more than adequate for this purpose. Since there is no dynamically stable intermediate stage, an alternation between the two circulation types results. At times the alternation seems to be in phase throughout the Northern Hemisphere but at other times this is not true. In view of the rapid dispersion of energy indicated by recent theoretical investigations [25], one might expect the in-phase condition to predominate but there is some evidence of slower modes of energy propagation in the atmosphere. Regardless of this, in any given region the action suggested is analogous to that of a relaxation oscillator, the potential, as represented by the north-south thermal gradient, building up to critical levels from time to time and then relaxing as the flow of the westerlies breaks down into meridional flow patterns. A study by Schwerdtfeger [27] indicated that in a source region for polar outbreaks the polar air may be built up to critical strength from a stage of complete exhaustion through radiation processes within a period of some three weeks. Observation indicates, however, that the pulse period of outbreaks in North America is more on the order of a week (corresponding to rather large-sized eddies) with a longer-period fluctuation in the strength of the pulses on the order of a month. The latter would correspond to the oscillation in question but such pulsations are far from systematic and hence it seems that other events, such as perhaps variations in extraterrestrial radiation, may play an important role.

The effect of the irregular geographic distribution of oceans and continents, although complicating matters, should not in itself account for the lack of regular periodicity.

INTERACTION ON A HEMISPHERIC BASIS

As was pointed out above, the most difficult step in extended-period forecasting is that involving the prediction of the movement and changes in form of the systems appearing on the extended-period representations. In the weather-type method this amounts to the forecasting of the next weather type or types. This usually involves directly or indirectly the study of large-scale and long-period weather processes on a hemisphere-wide basis. Some of these general processes which can be used in connection with extended-period forecasting by the use of weather types were set down

by the author during World War II [11]. Since then, additional work has been done upon them [1, 13]. In line with the character of this article, a few very general remarks will be included covering this approach.

At times, particularly during meridional flow periods, certain systematic movements are apparent. Curiously enough, such motions frequently exhibit a westward displacement or retrogression, not so much of the individual broad-scale crests and troughs aloft but of the amplitude or energy of the disturbance. Thus, following the development of a trough and crest pair, the next major trough or crest-trough pair upstream will increase in amplitude. The rate of the propagation is some seven degrees of longitude per day in winter [1]. Sometimes this upstream effect appears to be propagated mainly through the action of the subtropical easterlies; at other times the polar easterlies appear to be the controlling factor. Other evidence [13] indicates that systematic distortions of the westerlies develop slowly over a period of two or more weeks both upstream and downstream from a major disturbance such as a blocking high. Recent theoretical studies [25, 29] of energy propagation indicate the possibility of a number of modes of energy dispersion and clearly indicate the possibility of interaction on a hemisphere-wide basis. This confirms much recently obtained synoptic evidence of the dependence of North American weather developments on events occurring elsewhere in the Northern Hemisphere.

Considerations of the type just mentioned have been handled statistically through the use of the data appearing in the forty-six-year weather-type catalogue. Naturally the first and simplest statistical aid to be developed from the catalogue was merely the determination of the seasonal expectancies of the various types. Next came the expectancies for various types in one region following a given type. Later, the North American types were arranged in a certain order depending upon such items as the latitude of the Pacific anticyclone, strength and location of polar outbreaks, and natural statistical relationship, and numbers were assigned in accordance with this order. Then type-prediction regression formulas were developed [8] whereby future types could be predicted on the basis of the past sequence of types in one region. Success was limited. It is believed that the greatest shortcoming of these earlier regression formulas lay in their disregard of present and past types in other zones. More recently the type data for the five zones have been punched on International Business Machine sorting cards [1], and prediction schemes based upon present and past types in each of the five zones will be developed.

The use of the type catalogue in analogue selection is of course obvious. It is highly efficient in this respect but selections must be made with extreme care in order to avoid the above-mentioned shortcomings of "thumbprint" analogues. In a theoretical study [8] of "thumbprint" analogue prediction schemes it has been shown that the weakness of this system compared to a generalized prediction scheme lies in the fact that it is not possible to consider direct and cross matches between analogue elements with different time lags. The analogue scheme thus assumes that weather is a one-dimensional function of time.

In statistical investigations using the type catalogue it has been found necessary to consider processes of much greater duration than those covered in the usual extended-period considerations. Some rather interesting results have been found in studies of seasonal characteristics. An unpublished study shows that in a normal season the extreme meridional flow types occur frequently in the autumn but that, as winter approaches, the frequency of their occurrence drops off

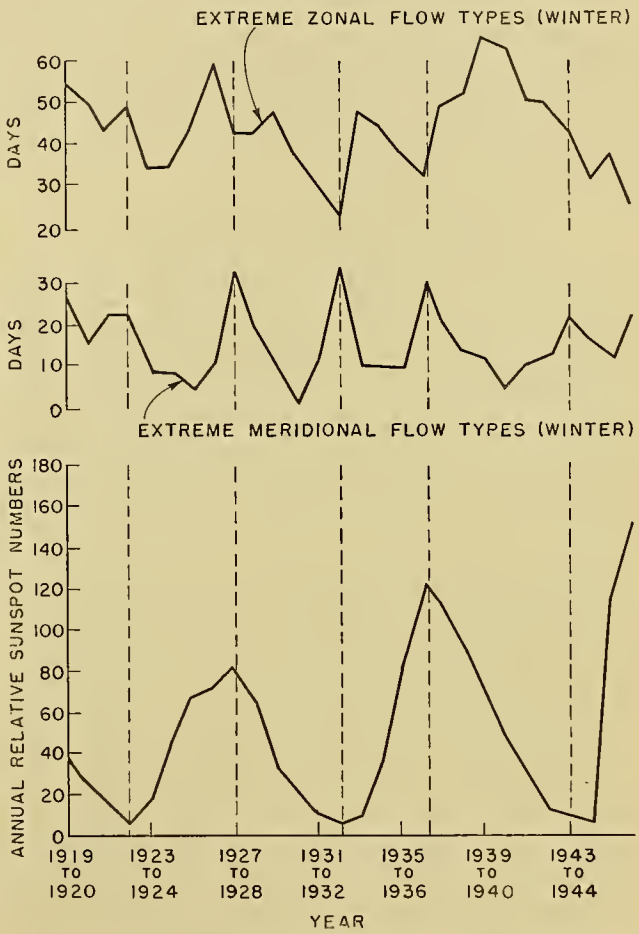


FIG. 2.—Curves showing the frequency of meridional (middle curve) and zonal (upper curve) flow types during the winter seasons 1919–20 through 1943–44. The lower curve gives the annual relative sunspot numbers during the same period.

rapidly and the extreme zonal flow types (E) increase in frequency. Again, in the spring the frequency of the meridional flow types increases and the zonal flow type decreases. Apparently, cross-westerly heat transport and general lateral mixing are greatest in the spring and the autumn. In certain years there is a distinct tendency for the extreme meridional flow types to continue with high frequencies on into the winter season. Since winter meridional flow types exhibit stronger abnormalities than those of autumn and spring, this leads to quite exceptional weather (*viz.*, January, 1937 and 1949) and for this reason the phenomenon was studied in detail. The effect can best be

demonstrated by graphing the number of days of occurrence of extreme meridional flow types (*D*, *A*, *Bn-c*) in the winter season (December, January, and February) for each year. This appears as the middle curve in Fig. 2. The curve for the extreme zonal flow types (*E*L, *E*M, *E*H) is the top curve in the figure. These curves show quasi-periodic variations with a period of five or six years and an appreciable amplitude. The two curves are out of phase.

The period is suggestive of half the sunspot cycle, and the curve of sunspot annual relative numbers is shown at the bottom for comparison. The sunspot curve is based upon the annual value corresponding to the year of the January and February portion of the plotted winter-season type characteristics and hence is in effect somewhat lagged in phase from the top two curves. The correspondence between the peaks in the extreme meridional flow types and the sunspot maxima and minima is striking.

PROSPECTS FOR THE FUTURE

With regard to prospects for the improvement of techniques for extended-range forecasting, it is the author's opinion that two fields of investigation offer the most attraction. One of these is the study of the long-period interaction between broad-scale circulation features in different parts of the Northern Hemisphere and, for that matter, over the whole world. This, it would appear, involves the modes of energy dispersion. A joint theoretical and synoptic attack is in order.

The other field involves a thorough study of the radiational properties of the earth and its atmosphere and the study of the effects of variations in solar output upon atmospheric motions. The example of a general relationship between solar activity and weather-type frequencies presented above is just one of many such relationships developed from studies of weather types. The works of Abbot, Clayton, and many others provide evidence favoring a real relationship. Meteorological events, seemingly inexplicable from the viewpoint of a closed system consisting of the atmosphere and the earth, are continually occurring and require a systematic investigation of extra-terrestrial effects.

REFERENCES

1. American Institute of Aerological Research, *Air Force—American Institute Aerological Research.* Res. Rep. No. 3, Part 2, pp. 31–44, 1949.
2. Bowie, E. H., and Weightman, R. H., "Types of Storms of the United States and Their Average Movements." *Mon. Wea. Rev. Wash.*, Supp. No. 1 (1914).
3. California Institute of Technology Meteor. Dept., *Synoptic Weather Types of North America.* Pasadena, Calif., 1943.
4. —— *Preparation of a Classification Graph of Synoptic Weather Sequences for North America, January 1899 through June 1939.* Rep. No. 943, Weather Div., Army Air Force Headquarters, 1945.
5. —— *Preparation of a Classification Graph for East Asia-West Pacific Synoptic Region.* Pasadena, Calif., 1944.
6. Darling, D. A., Rep. Nos. 5, 7, and 7a, Calif. Inst. Tech.–Army Air Force Res. Unit, 1944.
7. —— *Relationship between the Analogue Forecast and Other Methods of Forecasting.* Calif. Inst. Tech.–Army Air Force Res. Unit, 1944.
8. —— *Report on the Accomplishments of the Statistical Project.* Calif. Inst. Tech.–Army Air Force Res. Unit, 1942.
9. Defant, A., "Die Zirkulation der Atmosphäre in den gemässigten Breiten der Erde." *Geogr. Ann., Stockh.*, 3:209–266 (1921).
10. Elliott, R. D., *Studies of Persistent Regularities in Weather Phenomena.* Meteor. Dept., Calif. Inst. Tech., 1942.
11. —— *Extended Weather Forecasting by Weather Type Methods,* 55 pp. U. S. Navy Dept., Long Range Weather Forecasting Unit, Washington, D. C., 1944.
12. —— "Forecasting the Weather." *Weatherwise,* 2:15–18, 40–43, 64–67, 86–88, 110–113, 136–138 (1949).
13. —— and Smith, T. B., "A Study of the Effects of Large Blocking Highs on the General Circulation in the Northern-Hemisphere Westerlies." *J. Meteor.*, 6:67–85 (1949).
14. Gold, E., "Aids to Forecasting: Types of Pressure Distribution with Notes and Tables for the Fourteen Years 1905–1918." *Geophys. Mem.*, 2:149–174 (1920).
15. International Meteorological Consultants, *Catalogue of Atlantic-European Weather Types.* (Unpublished report)
16. Jacobs, W. C., "On the Energy Exchange between Sea and Atmosphere." *J. mar. Res.*, 5:37–66 (1942).
17. Klein, W. H., "Winter Precipitation as Related to the 700-Mb Circulation." *Bull. Amer. meteor. Soc.*, 29: 439–453 (1948).
18. Krick, I. P., *A Dynamical Theory of the Atmospheric Circulation and Its Use in Weather Forecasting.* Meteor. Dept., Calif. Inst. Tech., 1942.
19. Lewis, C. H., *3000 Dynamic Meter Ideal Charts Corresponding to the Ideal Surface Types for Zone 3 and 4.* Thesis, Meteor. Dept., Calif. Inst. Tech., 1947.
20. Namias, J., *Extended Forecasting by Mean Circulation Methods,* 57 pp. U. S. Weather Bureau, Washington, D. C., 1947.
21. Pagava, S. T., and others, *Basic Principles of the Synoptic Method of Long-Range Weather Forecasting,* 3 Vols. Leningrad, U. S. S. R. Hydrometeor. Publ. House, 1940. (Translation prepared by Weather Information Service, Headquarters, Army Air Forces.)
22. Paulhus, J., and Blewitt, S., *Six Day Weather Types of North America.* Meteor. Dept., Calif. Inst. Tech., 1943.
23. Reed, T. R., "Weather Types of the Northeast Pacific Ocean as Related to the Weather on the North Pacific Coast." *Mon. Wea. Rev. Wash.*, 60:246–252 (1932).
24. Rossby, C.-G., "Fluid Mechanics Applied to the Study of Atmospheric Circulations." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole occan. Instn.*, Vol. 7, No. 1, pp. 5–17 (1938).
25. —— "On the Propagation of Frequencies and Energy in Certain Types of Oceanic and Atmospheric Waves." *J. Meteor.*, 2:187–204 (1945).
26. Schell, I. I., "Baur's Contribution to Long-Range Weather Forecasting, Part II." *Mon. Wea. Rev. Wash.*, Supp. No. 39, pp. 79–87 (1940).
27. Schwerdtfeger, W., "Zur Theorie polarer Temperatur- und Luftdruckwellen." *Veröff. geophys. Inst. Univ. Lpz.*, (2), 4:255–317 (1931).
28. U. S. Weather Bureau, *Historical Weather Maps, Daily Synoptic Series, Northern Hemisphere, Sea Level,* Washington, D. C.
29. Yeh, T.-c., "On Energy Dispersion in the Atmosphere." *J. Meteor.*, 6:1–16 (1949).

VERIFICATION OF WEATHER FORECASTS

By GLENN W. BRIER and ROGER A. ALLEN

U. S. Weather Bureau, Washington, D. C.

INTRODUCTION

Verification of weather forecasts has been a controversial subject for more than sixty years and has affected nearly the entire field of meteorology. This paper will discuss some of the important reasons for this controversy and attempt to show that much of the existing confusion disappears when a careful analysis is made of the objectives of forecasting and verification. A number of verification systems that have been used will be described, but it is beyond the scope of this paper to make a complete survey of verification practices or of the literature on the subject. For the latter purpose the reader is referred to articles by Bleeker [1], Muller [6], and the U. S. Weather Bureau [10], which contain extensive bibliographies.

Definition of the Problem. Verification is usually understood to mean the entire process of comparing the predicted weather with the actual weather, utilizing the data so obtained to produce one or more indices or scores and then interpreting these scores by comparing them with some standard depending upon the purpose to be served by the verification. In the discussion which follows, it will be assumed that both forecasts and observations have been expressed objectively so that no element of judgment enters into the comparison of forecast with observation, and it will be assumed further that errors of observation are unimportant. These assumptions are discussed in a subsequent section. The selection and interpretation of an index or score which will meet the objectives of the verification study usually constitute the most difficult part of the problem since, as will be shown, practical considerations often require that the score fulfill a number of requirements and furnish information on a number of different characteristics of the forecasts. Selection of arbitrary scores intended to measure a number of parameters usually leads to difficulty in the last step, interpreting the score.

PURPOSES OF VERIFICATION

One of the earliest purposes of forecast verification was to justify the existence of the newly organized national weather services, and thus the question of verification immediately became a football to be kicked around by the supporters and opponents of the national weather services. Some, such as Klein [5], claimed that the official synoptic forecasts had little or no practical value and produced verification figures intended to prove their point. Others, for instance Schmauss [7], claimed that the weather forecasts could not be subjected to rigorous statistical tests or that the tests that had been performed had no meaning. Today the value of the national weather services is so widely accepted that this particular purpose of verification no longer is of much importance. However, the effects of this controversy still show their influence by often preventing a realistic attempt to solve some of the problems of forecast verification since some meteorologists object a priori to any scoring system that does not produce high enough figures to make the forecasts appear favorable in the eyes of the public or government appropriating bodies.

Economic Purposes of Verification. Since the sole purpose of practical forecasting is economy of life, labor, and dollars, it would seem that one of the chief purposes of verification is to determine the economic value of the forecasts which have been made. But such an evaluation, especially for the whole economy, is difficult or impossible since the uses and users of forecasts are so diverse and ramified. Forecasts that may have considerable value for one user may have little or no utility for another user or may actually have a negative value if their accuracy is low. The measurement of the economic value of the forecast is thus a separate project for each user, and this is usually impracticable or impossible because the individual has neither the essential economic data nor the facilities to make such an evaluation. Thus a farmer may feel that an accurate weather forecast enabled him to make some saving in seed cost and planting labor, but he may be unable to estimate the cost of his own labor, and the total saving might depend upon the value of the harvested crop which is, of course, unknown until harvest time. Sometimes such economic evaluation of the forecasts is possible, the necessary data being found in the cost accounting and financial structure of the particular business [2]. However, since evaluation in economic terms is usually impossible, it is most often desirable to determine the reliability of weather forecasts by measuring their approach to the truth and expressing the result in terms of some arbitrary scale such as errors in degrees Fahrenheit or per cent of hits. The user of the forecast can then interpret the information in terms of his operations. Thus an electric power company may want to know when thunderstorms are forecast for a given area and what the likelihood is that the forecast will be correct, or, on the other hand, what per cent of observed thunderstorms in an area will be correctly forecast.

Administrative Purposes of Verification. One of the most useful purposes of verification is to determine the relative ability of different forecasters. In a weather organization the relative ranks of forecasters may be desired for a number of reasons, such as selecting promising personnel for further training in forecasting, selecting able men for difficult or specialized forecasting

assignments, selecting research personnel, or rating the efficiency of the forecaster. The verification scheme adopted may apply only to the regular official forecasts made by the individual or to a special practice-forecast program. Such a program may involve people not assigned to regular forecasting and thus provide a basis for discovering hidden talent. In the classroom a practice-forecast verification program is used to measure the progress of student forecasters and can assist the professor in vocational guidance. A student showing low forecasting ability but high mathematical ability might be encouraged to go into theoretical meteorology rather than into practical forecasting. Numerous other examples of administrative purposes of verification could be given.

Verification procedures, when applied to the official forecasts of a weather station, can make a valuable contribution to the control of the quality of the output of the station. In industry it has been found that scientific sampling procedures are necessary to maintain uniformity in the manufacturing process. If the forecasts made at a particular station over a period of time appear to fall below the standards of accuracy that have previously been attained, administrative action to investigate the reasons for the falling standards might be desirable. There is also a further advantage that the mere existence of a checking scheme, even if imperfect, tends to keep the forecasters more alert and interested in maintaining the accuracy of forecasts.

Scientific Purposes of Verification. One of the goals of scientific meteorology is to be able to predict precisely the state of the atmosphere at any time in the future. This goal is a difficult one to attain, but considerable progress has been made in the past several decades in understanding the physics of the atmosphere. Sometimes the question is raised as to whether there has been any increase in the accuracy of weather forecasts over a period of time. The question asked may be quite general, such as whether temperature or precipitation forecasts made by meteorologists in general are more accurate than they were fifty years ago. Sometimes the question may be directed at a specific group of forecasters using special methods on an experimental basis. In both these cases verification statistics can be used to provide information on the trend in forecast accuracy, although the technical difficulties in obtaining accurate statistics may be great. When some new advance in meteorological theory which bears on forecasting is proposed, it may be possible to compare the verification scores of experimental forecasts made using the supposed advance with forecasts made without the new theory. In this case verification is equivalent to testing the validity of a scientific hypothesis.

Another scientific purpose of forecast verification is the analysis of the forecast errors to determine their nature and possible cause. This is, in the opinion of some, the most important and most fruitful objective of verification since it is more susceptible of scientific treatment than some of the other purposes. A search may be made for indicators of forecasting difficulty which will help to locate the synoptic situations under which forecasts are most likely to be wrong. It is commonly thought, for example, that situations where weather changes are taking place rapidly are more difficult to forecast than other situations, but it should be noted in passing that the literature does not reveal any precise studies to support this view. Likewise, although this is a widespread belief, it remains to be shown by verification figures that the forecasting accuracy for some element such as precipitation occurrence is lower for cases in which deepening or filling of a low is poorly forecast than for cases in which the deepening or filling is accurately forecast. Such verification as this can be used to discover the weaknesses of forecasting systems in order to decide where research emphasis is needed.

FUNDAMENTAL CRITERIA TO BE SATISFIED

Terminology of Forecasts. One essential for satisfactory verification is objectivity, which requires that the forecasts be explicitly stated, either numerically or categorically, thus permitting no element of judgment to enter the *comparison* of forecast with subsequent observation. But the relation between the forecaster and the public, which depends upon the forecaster's terminology, is usually subjective in nature, since even objective terms and the actual weather are interpreted subjectively by many individuals. Bleeker [1] discusses this point clearly. If every forecast is accompanied by a definition of the terms used, the forecaster sometimes objects on the basis that it hinders his freedom to express himself adequately. At this point it is easy to raise questions about the psychology of forecasters and of the public and many arguments in forecast verification revolve around this point. Although these are practical and very important problems to those attempting to serve the public with weather information, they are in essence outside the field of forecast verification and the goals of verification would be reached much sooner if this were more generally recognized. Only those forecasts which are expressed in objective terms can be satisfactorily verified *as forecasts*. The extent to which public forecasts satisfy the public is a different question which can be answered, it appears, only by public opinion polls; and the answer generally will contain little information about the agreement of forecast and actual weather.

Meteorologists themselves sometimes advocate subjective verification, particularly in the case of prognostic charts which attempt to portray the pattern of some weather element (such as pressure) rather than to specify the weather at individual points. In some cases this has led to the use of boards of experts who compare the prognostic charts with observed charts. The difficulty in objective comparison arises because of the unsatisfactory state of knowledge as to what are the important parameters of the prognostic pattern; or in other words, the forecaster is unable to specify objectively just what he is trying to represent with a prognostic chart. In effect, the forecaster is trying to verify something which cannot be observed, hence this situation does not meet the definition of verification.

An extreme view denying the need for objectivity is that expressed by Hazen [3], who says,

> ... to make a proper verification of a weather forecast, and one that shall be rigidly applicable to every case ..., it should be done by ... one thoroughly acquainted with the average conditions and he should verify from the map on which the prediction was based and not from subsequent maps. If the verification is from a subsequent map, care should be taken to consider what abnormal conditions have occurred which could not have been anticipated.

Selection of Purposes of Verification. Before any satisfactory verification scheme is adopted it is necessary to determine the primary purpose or purposes to be served by the verification This may appear to be obvious, but the history of the forecast verification controversy during the past half-century makes it clear that this fundamental point has been forgotten again and again. A scheme that is adequate for one purpose may be unsatisfactory for another, just as an automobile is suited for travel over the highway but is quite inefficient for flying through the air. Unfortunately, much time has been wasted and much confusion has resulted from attempts to devise a verification method or single score that will serve all purposes. The very nature of weather forecasts and verifications and the way they are used make one single or absolute standard of evaluation impossible. A set of forecasts has many different characteristics. Some of the forecasts are correct, and knowledge of the percentage which are correct may serve some purposes. Usually the weather occurrences in one part of the range will seem to be more important than those in other parts of the range, and this must be considered in specifying the purpose of the verification. Thus if the set of forecasts are forecasts of temperature in a citrus grove, the characteristic which is most important may be the percentage of forecasts of temperature below freezing which were actually followed by temperature below freezing, and the percentage of forecasts which were correct may be unimportant. The situation is in important respects analogous to measuring an object, for instance a table. The table has length, breadth, height, smoothness of the top, hardness of the top, number of legs, etc. A measure of any such characteristic is of no value unless a way of using the measurement has first been established.

In general, if each purpose of verification is exactly specified in advance, in the form of a hypothesis, not only will it be much easier to select verification scores to satisfy each purpose, but there will be no doubt as to what action is indicated by any numerical value which the verification score may have. It will often be desirable to select the purpose and the score in such a way that the result will either support or reject an a priori hypothesis.

Specifications of a Scale of Goodness. Another essential criterion for satisfactory verification is that the verification scheme should influence the forecaster in no undesirable way. One of the greatest arguments raised against forecast verification is that forecasts which may be the "best" according to the accepted system of arbitrary scores may not be the most useful forecasts. Resolving this difficulty becomes a question of how to define "best." That there is no unique answer to this question can be seen by considering the following example in which it is desired to verify some forecasts of minimum temperature. Suppose that on some particular occasion the probability of occurrence during the subsequent night of the various integral degrees of minimum temperature is actually known by a forecaster to be as follows:

Minimum temperature (°F)	31	32	33	34	35	36	37
Probability of occurrence	.05	.10	.15	.25	.30	.10	.05

If the forecaster is required to state a single temperature figure in his forecast for verification purposes, what figure should he state on this particular occasion so as to maximize his score? There are a number of answers. If the verification scheme counts as hits only the occasions when the forecasts are exactly right, he should forecast 35F, since this value has the greatest probability of occurrence. If he is being verified on the basis of mean absolute error, he should say 34F, the median of the frequency distribution, since this will minimize the sum of the absolute deviations. If he is to be verified on the basis of the square or root-mean-square of the error, he should forecast 34.15F, which is the mean value of the frequency distribution. Any number of such arbitrary scoring systems could be devised and they will all influence the forecaster, at least to some extent, or in effect actually do part of the forecasting. The verification scheme may lead the forecaster to forecast something other than what he thinks will occur, for it is often easier to analyze the effect of different possible forecasts on the verification score than it is to analyze the weather situation. Some schemes are worse than others in this respect, but it is impossible to devise a set of scores that is free of this defect. There is one possible exception to this which will be discussed in the section on verification of probability statements.

VERIFICATION METHODS AND SCORES

It has been stated above that, in general, different verification statistics will be required for each different purpose. The discussion of various verification statistics in the following paragraphs can therefore only hint at the use which might be made of each of the scores, since so few types of scores appear ever to have been of real value.

Contingency-Table Summaries. When forecasts are made in categorical classes a useful summary of the forecast and observed weather can be presented in the form of a contingency table. Such a table does not constitute a verification method in itself, but provides the basis from which a number of useful pertinent scores or indices can easily be obtained An example is given in Table I where precipitation was forecast in three classes, heavy, moderate, or light, for thirty

occasions. One of the greatest advances in forecast verification was made when the limits of the various classes were chosen in such a way that each category had an equal probability of occurrence, based on the past climatological record. Thus there is no incentive for a forecaster to choose one class in preference to another because of purely probability (climatological) considerations. This principle, with slight modification, has been used in the verification of the Extended Forecasts of the U. S. Weather Bureau [11], and during the last war the Army Air Forces [8] devised a scheme in which thirty classes (called *trentiles*) were used, each class representing $\frac{1}{30}$ of the frequency distribution based on past records of the particular element being verified.

Table I. Contingency Table for Precipitation Forecasts

Observed precipitation	Forecast precipitation			
	Heavy	Moderate	Light	Total
Heavy	5	2	1	8
Moderate	8	4	1	13
Light	2	2	5	9
Total	15	8	7	30

From this table a number of interesting verification statistics can be obtained. A comparison of the margins reveals whether the various categories were forecast with the same frequency as they occurred. Thus it is noted that, although heavy precipitation was forecast 15 times, it occurred only 8 times, the opposite tendency being shown for moderate precipitation. This may be only a sampling difference, or it may be great enough to cause the forecaster to reject the hypothesis that he is able to distinguish the relative frequencies of the various classes. The relative frequency of occurrences of the various classes can also be compared with that expected on the basis of climatology.

Percentage Correct. From the contingency table a frequency distribution of errors can easily be obtained. In the example given, 14 forecasts are exactly right, 13 forecasts are wrong by one class, and 3 forecasts are wrong by two classes. A commonly used score is the per cent right, in this case $\frac{14}{30} = 47$ per cent. More useful information is provided by constructing two other tables, Tables II and III. The extent to which subsequent observations confirm the prediction when a certain event is forecast is shown by Table II. The term *post agreement* has been suggested for this attribute of the forecasts [9] and the term *prefigurance* for the extent to which the forecasts give advance notice of the occurrence of a certain event (illustrated by Table III). Thus it is seen that forecasts of heavy precipitation were followed by the occurrence of heavy precipitation 33 per cent of the time while occurrences of heavy precipitation were correctly indicated in advance 62 per cent of the time.

Table II. Per Cent of Time Each Forecast Event Occurred for a Particular Category

Observed class	Forecast class		
	Heavy	Moderate	Light
Heavy	33	25	14
Moderate	53	50	14
Light	14	25	72
Total	100	100	100

Table III. Per Cent of Time Each Observed Category Was Correctly Forecast

Observed class	Forecast class			
	Heavy	Moderate	Light	Total
Heavy	62	25	13	100
Moderate	61	31	8	100
Light	22	22	56	100

For some economic uses these two scores are probably the most important verification figures. In planning an operation which is influenced in an important way by heavy rain, for example the operation of a series of flood-control dams, it is potentially of value to know both the percentage of heavy-rain forecasts which are likely to be correct, that is, the reliance to be placed on a heavy-rain forecast, and the percentage of heavy-rain occurrences which are likely to be correctly forecast.

Skill Score. The information contained in the contingency table is often combined into a single index (S), called a *skill score* and apparently first proposed by Heidke [4]. It is defined by

$$S = \frac{R - E}{T - E},$$

where R is the number of correct forecasts, T is the total number of forecasts, and E is the number expected to be correct based on some standard such as chance, persistence, or climatology. This score has a value of unity when all forecasts are correct and has a value of zero when the number correct is equal to the expected number correct. It will be discussed at greater length in the section on comparison with control forecasts.

Average Error, Root-Mean-Square-Error, Etc. When the values of the forecast element are expressed on a continuous numerical scale, as temperature usually is, it is often desirable to express a verification score in terms of an average absolute error or a root-mean-square-error (*RMSE*). If in a series of N forecasts F_i represents the i-th forecast and O_i the corresponding observation, the average absolute error $\bar{e}$ is given by the formula

$$\bar{e} = \frac{\sum |F_i - O_i|}{N}$$

and the *RMSE* by

$$RMSE = \sqrt{\frac{\sum (F_i - O_i)^2}{N}}.$$

One disadvantage of using either of these scores is that it gives the "cautious" forecaster an opportunity to

hedge by playing the middle of the range. Thus a forecaster may feel that the minimum temperature during the night will be near 40F if the cloud cover remains but will be near 30F if the skies clear. If he is being verified on the basis of the *RMSE*, he might tend to choose 35F in cases of doubt even though he may be quite sure that the temperature will not be close to 35F.

These scores can also be used to verify prognostic pressure-patterns by comparing the forecast and observed pressures or pressure gradients at a number of sample stations or points on the map.

Correlations. The correlation coefficient is a statistic which measures association between two sets of values, such as forecast and observed temperatures. One simple formula for the coefficient is

$$r = \frac{N\sum F_i O_i - (\sum F_i)(\sum O_i)}{\sqrt{N\sum F_i^2 - (\sum F_i)^2}\cdot\sqrt{N\sum O_i^2 - (\sum O_i)^2}}.$$

When used as a verification score it has the advantage of being relatively free from the fault of influencing the forecaster in an undesirable way. However, it is insensitive to any bias or error in scale that a forecaster may have. Thus the centigrade and Fahrenheit scales are perfectly correlated, but a person would not use one scale to verify forecasts made using the other scale. The correlation coefficient is also difficult to interpret for the purpose of making operating decisions and is subject to abuse of interpretation, such as attaching too much significance to the value of a coefficient obtained from a small sample. It is influenced by trends that exist in both series so it is usually desirable to compute it by using departures from normal.

Verification of Probability Statements. There appears to be one situation where it is possible to devise a verification scheme that cannot influence the forecaster in any undesirable way. This is the case when the forecasts are expressed in terms of probability statements. Suppose that on each of N occasions an event can occur in only one of r possible classes and on one such occasion i, $f_{i1}, f_{i2}, \ldots, f_{ir}$ represent the forecast probabilities that the event will occur in classes $1, 2, \ldots, r$, respectively. If the r classes are chosen to be mutually exclusive and exhaustive, $\sum_{j=1}^{r} f_{ij} = 1$ for each and every $i = 1, 2, \ldots, N$.

A score P can be defined by

$$P = \frac{1}{N}\sum_{j=1}^{r}\sum_{i=1}^{N}(f_{ij} - E_{ij})^2,$$

where E_{ij} takes the value 1 or 0 according to whether the event occurred in class j or not. For perfect forecasting this score will have a value of zero and for the worst possible forecasting a value of 2. Perfect forecasting is defined as "correctly forecasting the event to occur with a probability of unity or 100 per cent confidence." The worst possible forecast is defined as "stating a probability of unity or certainty for an event that did not materialize" (and also, of course, "stating a probability of zero for the event that did materialize"). It can be shown that if $p_1, p_2, p_3, \ldots, p_r$ are the respective climatological probabilities of classes 1, 2, $3, \ldots, r$ then in the absence of any forecasting skill the best values to choose for f_{ij} will be p_j for all N occasions. This will minimize the score P for constant values of $f_{1j} = f_{2j} = \ldots f_{Nj}$ and the expected value of the score will be

$$E(P) = 1 - \sum_{j=1}^{r} p_j^2.$$

To illustrate the procedure consider the following table of ten actual forecasts of "rain" or "no rain" in which a probability or confidence statement was made for each forecast.

TABLE IV. EXAMPLE OF FORECASTS STATED IN TERMS OF PROBABILITY

Occasion	Rain		No rain	
	Forecast probability	Observed	Forecast probability	Observed
1	0.7	0	0.3	1
2	0.9	1	0.1	0
3	0.8	1	0.2	0
4	0.4	1	0.6	0
5	0.2	0	0.8	1
6	0.0	0	1.0	1
7	0.0	0	1.0	1
8	0.0	0	1.0	1
9	0.0	0	1.0	1
10	0.1	0	0.9	1

In the table, unity is placed in the "rain" column if rain occurs. If the event is "no rain," unity is placed in the "no rain" column. The score P is therefore

$$P = \tfrac{1}{10}\{0.7^2 + 0.7^2 + 0.1^2 + 0.1^2 + 0.2^2 + 0.2^2 + \ldots 0.1^2 + 0.1^2\} = 0.19.$$

Since rain occurred $\frac{3}{10}$ of the time, the minimum (or best) score that could have been obtained by making the same forecast every day would be

$$P_{\min} = 1 - (0.3^2 + 0.7^2) = 0.42.$$

Thus the forecaster is encouraged to minimize his score by getting the forecasts exactly right and stating a probability of unity. If he cannot forecast perfectly, he is encouraged to state unbiased estimates of the probability of each possible event. On the other hand, with complete absence of knowledge or forecasting skill he is encouraged to predict the climatological probabilities and not just forecast the most frequent class.

TABLE V. VERIFICATION OF A SERIES OF 85 FORECASTS EXPRESSED IN TERMS OF THE PROBABILITY OF RAIN

Forecast probability of rain	Observed proportion of rain cases
0.00–0.19	0.07
0.20–0.39	0.10
0.40–0.59	0.29
0.60–0.79	0.40
0.80–1.00	0.50

When a series of forecasts has been made using probability statements, a study can also be made to determine whether the forecast probabilities are related to the relative frequency at which the events occur. An example of this type of comparison is shown in Table V (based on a more extended series of such forecasts) which suggests a relationship between the forecast and the observed probabilities, but which indicates that the forecaster should modify or adjust his scale to improve the forecasts.

Arbitrary Scores for Special Purposes. An infinite number of scores can be devised for verification purposes, but it is beyond the scope of this paper to discuss many of them in detail. In particular cases, such as verifying the forecasts of cloud heights or visibilities, it may seem desirable to express errors in terms of per cent and compute a score such as

$$e_p = \frac{1}{N}\sum_{i=1}^{N}\frac{|F_i - O_i|}{O_i},$$

where F_i is the forecast value and O_i is the observed value.

In other instances, verification of selected cases in which either forecast or observation was of particular importance may be required, or it may be desired to transform the forecasts into forecasts of change before verifying them. It is common, for example, to want to verify only cases when low visibility either is forecast or occurs, since high values are much less critical for airplane operations. It may be desired to verify only the cases when ceiling changes across the critical value, either in the forecast or in the occurrence. With more intensive study of the decisions which are made on the basis of such forecasts and of the consequences of right or wrong decisions, it would be possible to devise verification scores based on the relative values of different forecasts, but as it is, practically all such specialized verification systems are arbitrary. This is not intended to discourage the use of such scores, but it should be emphasized that conclusions as to the relative *values* of two sets of forecasts will always be uncertain so long as the value of any individual forecast cannot be estimated. Innumerable other scores to be used for some special purpose could be mentioned.

CONTROL FORECASTS FOR COMPARISON

Chance or Random Forecasts, Persistence, Climatology. Up to this point most of the discussion has been concerned with methods of comparing the forecast weather with the actual weather and obtaining some indices or scores. The final phase of the verification procedure is the interpretation of these scores, which is usually performed by comparing them with some standard. It has long been recognized that a figure such as the percentage of correct forecasts is often meaningless, since the same figure might be obtained by chance or by some scheme of forecasting requiring no meteorological knowledge.

It has been suggested that forecast scores be compared with scores obtained using various "blind" forecasts such as random forecasts, persistence forecasts, or climatological forecasts. Many arguments have arisen over the proper choice (if any) of the blind forecasts to use as a standard. There is no correct answer, of course, since the choice depends not only upon the use that is made of the forecasts but upon the purpose of the verification.

To illustrate these comparisons, let us return now to the skill score defined in the earlier section. The expected number of forecasts correct E_R, based on chance or random forecasts for the data in Table I, is computed by the following formula:

$$E_R = \frac{\sum R_i C_i}{T},$$

where R_i is the total of the i-th row and C_i is the total for the i-th column. In the example given, this is

$$\frac{8 \times 15 + 13 \times 8 + 9 \times 7}{30} = 9.6,$$

so the skill score based upon the margins of the contingency table is

$$S_R = \frac{14 - 9.6}{30 - 9.6} = 0.22.$$

If, in this same example, persistence forecasts had been made by predicting a continuation of the preceding precipitation, the expected number right would have been 12, so the skill score would be

$$S_p = \frac{14 - 12}{30 - 12} = 0.11.$$

If the climatological frequencies of heavy, moderate, and light precipitation were P_H, P_M, and P_L, respectively, the expected number right E_c is given by the formula

$$E_c = P_H(F_H) + P_M(F_M) + P_L(F_L),$$

and

$$E_c = \tfrac{1}{3}(30) = 10$$

in the example, since the three classes were defined as equally likely.

Generally, any score which may be devised can be computed for a set of control forecasts as well as for the actual forecasts, and a comparison can be made.

Adjustment for Difficulty. When the purpose of verification is to compare a number of forecasters, a practical difficulty often encountered is that the forecasts are not comparable because they were made at different times and under different conditions. It is known that different synoptic situations present varying degrees of difficulty and it is invalid to make comparisons between forecasters working at different times unless some attempt is made to take account of this source of variation. This can be partly accomplished by finding parameters dependent on the synoptic situation that are related to the errors in the forecasts.

For example, it might be found that there is a high correlation between the verification scores for visibility forecasts made by an individual and the scores obtained by a visibility forecast based simply on persistence. By

means of a regression analysis the scores S made by a number of forecasters can then be adjusted to equivalent "persistence" values before making comparisons by use of the formula

$$S_A = S + b\,(S_p - \bar{S}_p),$$

where S_A is the adjusted score, S the unadjusted score, S_p the persistence score corresponding to S, $\bar{S}_p$ the mean persistence score, and b the average within regression of forecast score on persistence score. This procedure can be generalized to take account of two or more parameters measuring difficulty.

Forecast-Reversal Test. A useful and simple procedure that can often be used to advantage might be called the "forecast-reversal" test. Suppose that a long-range forecaster claims he can pick out a year in advance the days of the month upon which rain will fall, provided a "tolerance" in timing of one day in either direction is allowed. On this basis some arbitrary verification is proposed which appears to give a relatively good score when applied to some actual forecasts. In the forecast-reversal test the same arbitrary verification procedure would be used, but the forecasts would be reversed with "rain" days being called "no-rain" days and vice versa. If approximately the same verification score were obtained on the reversed forecasts, one would be led to suspect either that the forecasts had no merit or that the verification scheme had no merit, or both. If this procedure were more generally used, especially for long-range forecasts, probably fewer claims would be made by long-range forecasters. It would also probably result in the selection of more sound verification schemes for use in determining the accuracy of such forecasts.

PITFALLS OF VERIFICATION

There are many pitfalls of verification to entrap both the novice and the expert. One of the greatest dangers lies in attempts to compare the relative abilities of forecasters on the basis of forecasts which are not comparable because of differences of location, season, time of day, length of forecast interval, etc. The reason for this is that the degree of forecasting difficulty varies so much from one circumstance to another that a very large sample of forecasts is needed to assure that the average weather has been approximately the same in the two sets of forecasts being compared. Even if the forecasts being compared are for the same event, there may be other factors to be considered, such as whether equal map facilities were available to each forecaster.

Interpretation is made even more difficult when scores on two or more forecast weather elements are arbitrarily combined to form a single index to be used for comparison purposes. It *may* be true, for example, that there is some particular use of a forecast for which an error of 2° in temperature is equivalent to an error of 0.25 in. in a precipitation forecast, but it is *certain* that this (or any other) arbitrary weighting is not true in general Those who must make a selection between forecasters based on verification scores may demand a single index to represent the verification of all forecasts made by each man, but since a logical combination of scores is in general impossible, it would be far better to consider each score separately. As pointed out in the section on Selection of Purposes of Verification, if it is decided ahead of time just what measures of accuracy are needed and what will be done with each, the need for combining scores could be largely eliminated.

Another practice that may lead to difficulty is the use of overlapping classes for verification purposes. Thus rules may be set up stating that a ceiling forecast of 200 ft will verify if the observation is between 100 and 400 ft, and that a forecast of 400 ft will verify if the observation is between 200 and 800 ft. Although such rules may seem reasonable they tend to encourage forecasters to hedge by choosing the classes having the widest range. Also, the choice of rather wide tolerances in verification limits tends to give rather high and uniform percentage scores to all forecasters, thus failing to discriminate between the better and poorer forecasters. Since the use of a contingency table for comparing forecasts and observations presents the data in one of the most useful forms, the practice of using overlapping classes is to be discouraged because such verification cannot be displayed in this way.

FUTURE RESEARCH

The preceding discussion has emphasized that verification is less difficult than has been thought and that the lack of recognition in the past of the various and diverse objectives of verification has been the real obstacle to progress. One of the most promising fields of study in this connection is that of setting up realistic problems to be solved and selecting scores which would furnish the desired information. Such studies might profitably be conducted in collaboration with administrators who need to select or rank forecasters; with economists or business advisors who need to know the effect of weather on operations and who are in the best position to determine what characteristic of the forecasts is related to profit and loss factors, or with meteorologists who know what characteristic of the forecast is useful for verifying scientific hypotheses. This appears to be a field of study in which private meteorologists should play a very important role, for their knowledge of the real uses which can be made of weather forecasts should be invaluable in specifying the questions which need to be answered by verification. On the scientific side of verification, the relation of forecast error to measures of forecast "difficulty" needs further investigation. If measures relating the synoptic conditions to forecast errors can be found, it would not only assist in the comparison of forecasts made at different times, but would also provide information as to where applied meteorological research is most needed and might contribute to fundamental knowledge of atmospheric processes.

Investigations regarding the effect of verification systems on the forecaster are needed since very little concrete evidence on this point is available. In this connection the use of probability statements in forecasting needs to be explored in more detail, for if such

statements can be verified without influencing the forecaster in any undesirable way, an important objection to verification is removed. Furthermore, forecasts expressed in this manner might very well be more useful at the same time. As was stated in the section defining the verification process, it has been assumed that errors of observation are unimportant in practical verification. This is not always the case, and further investigation of the effect of such errors is needed. Most weather forecasting is so far from perfect that forecasting errors far overshadow observational errors, but exceptions may be found, for example in visibility forecasting and perhaps in other cases.

There are many other unsolved problems in verification, but on careful analysis it appears that most of them are meteorological in nature. As long as our knowledge of the physics of the atmosphere remains so incomplete this state of affairs will be reflected in the verification problem.

REFERENCES

1. Bleeker, W., "The Verification of Weather Forecasts." *Meded. ned. meteor Inst.*, (B) Deel 1, Nr. 2 (1946).
2. Brier, G. W., "Verification of a Forecaster's Confidence and the Use of Probability Statements in Weather Forecasting." *U. S. Wea. Bur. Res. Pap.* No. 16 (1944).
3. Hazen, H. A., "The Verification of Weather Forecasts." *Amer. meteor. J.*, 8: 392–396 (1891).
4. Heidke, P., "Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst." *Geogr. Ann., Stockh.*, 8: 310–349 (1926).
5. Klein, H. J., "Misserfolge des staatlichen Wetterprognosendienstes in drei Monaten seines Bestehens." *Gaea, Köln*, 42: 641–652 (1906).
6. Muller, R. H., "Verification of Short-Range Weather Forecasts (A Survey of the Literature)." *Bull. Amer. meteor. Soc.*, 25: 18–27, 47–53, 88–95 (1944).
7. Schmauss, A., "Die Treffsicherheit der Prognosen." *Wetter*, 28: 68–71, 167–168 (1911).
8. U. S. Army Air Forces Headquarters, Weather Information Branch, *Short Range Verification Program.* Report No. 602, Washington, D. C., 1943.
9. U. S. Army Air Forces, "Critique of Verification of Weather Forecasts." *Air Wea. Serv. Tech. Rep.* No. 105–6 (1944).
10. U. S. Weather Bureau, *Methods of Verifying Day-to-Day Weather Forecasts.* Report by the Special Committee on Forecast Verification. Washington, D. C., 1940 (unpublished).
11. Willett, H. C., Allen, R. A., and Namias, J., "Report of the Five-Day Forecasting Procedure, Verification and Research as Conducted between July 1940 and August 1941." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 9, No. 1, 88 pp. (1941).

APPLICATION OF STATISTICAL METHODS TO WEATHER FORECASTING

By GEORGE P. WADSWORTH

Massachusetts Institute of Technology

It must seem very odd to the layman and to the scientist who is not connected with the field of meteorology that so little practical progress has been made during the last decade in the all-important problem of weather forecasting. This defect is particularly conspicuous when such phenomenal advances have been made in the field of nuclear physics and thermodynamics. However, it is necessary to spend only a moderate amount of time examining observational data of the meteorological elements to understand that the problem is much more difficult in many of its aspects than those considered in most allied fields. If we regard the atmosphere as a dynamic model, it soon becomes apparent that the whole process behaves as a complicated mechanism in which past and present values do not determine the future as in most linear processes, but they themselves have an effect upon the system which in turn produces nonlinearity of a very peculiar type. It is with this phenomenon that the meteorologists must contend.

There seem to be only two ways which are available to solve the problem of the motions of the weather systems and therefore the problem of weather forecasting: one is the dynamical approach and the other is the statistical approach. Oftentimes one thinks of these two methods as conflicting programs but actually they are attempting to get at one and the same thing. In the dynamical approach the laws connecting various meteorological phenomena are investigated. These laws are considered to be precise in action even though the data are subject to fluctuations which are random and are thus necessarily incomplete. On the other hand, in the statistical approach the quantities are taken as they are found and the distributions examined both singly and in such combinations as one chooses for the purposes of investigation. In an ideal survey all possible statistical parameters would be considered and not merely a partial selection, but such an undertaking is impractical because of the sheer magnitude of the task. If, however, there exist among the statistical parameters a few which are connected by sharp dynamic laws, then some of the parameters would be determined by a knowledge of the remainder, and the dynamics would be obvious as a statistical fact. The failure to date of statistical methods to contribute markedly to the progress of meteorology may have been due to one or more of three facts:

1. The parameters in the analyses to date are not the important ones.

2. The parameters considered are insufficient in number to give us a true picture of the operation.

3. These parameters have been observed so inaccurately that they are not in fact the significant parameters of the dynamics.

It might however be said that it is impossible for dynamics to exist and be really significant and not be brought out by a proper statistical examination of the relevant quantities, and in fact, techniques available in the statistical category have the added advantage that it is unnecessary to make many simplifying assumptions in order to yield relevant information concerning the meteorological phenomenon.

At this point it cannot be too strongly emphasized that the use of statistical techniques for the solution of the behavior of the weather systems is equivalent to the setting up of equations of motion and obtaining the solution. The use of generalized harmonic analysis (discussed in a later paragraph) as a method of attack on the problem has the advantage that it automatically takes care of data which have superimposed a random component upon the dynamic motion. In other words, it is a technique which is especially designed to handle situations where, because of ignorance, there has to be omitted a certain number of factors which can be considered in combination as a random phenomenon superimposed on the dynamic movement. There is no doubt that the ideal way to solve any dynamical problem is to set up a physical model in terms of mathematical symbols and, after having solved the resulting mathematical model, to use statistical methods to solve for the basic parameters. However, if the mathematical model does not contain all of the variables which are actually important, the attempt to solve for the parameters by statistical methods will give an unreliable result, and it is for this reason that a technique which takes account of this unknown group of variables is very appropriate for this problem. An example of such a situation would be one in which the phenomenon in which we were interested had a displacement y given by $y = A \sin kt$. Superimposed upon this oscillation might be a group of frequencies which had periods entirely different from those of the phenomenon. This group of frequencies would have been reflected in our observations since they each enter the picture for short periods of time and more or less at random. It would be extremely difficult by observing a small section of the data to fit the parameters A and k, the constants in our equation. However, a statistical analysis would be much more successful in separating the basic signal from the random noise.

It is perhaps during the last ten years that an added impetus has been given to this method of approach because of the growing conviction on the part of many investigators that the problems of the general circulation and of specific forecasting were not to be solved in the near future by the usual dynamic methods. This is reminiscent of other physical sciences wherein, at the beginning, a conceptual foundation was borrowed

from physics and mathematics, but it was only after extended programs of experimental work had been carried out that the subjects progressed independently and produced the astounding developments which have occurred in the last century. A continual succession of experimentation, followed by new theory to explain the results obtained, and then a new period of experimentation, has characterized the development of most of the physical sciences. Unfortunately, meteorology is not in the same position because it is impossible to conduct experiments with controlled variables since no formal method for the inclusion of all the pertinent variables into a single mathematical system has yet been found as a substitute for experimental control. Thus the most powerful techniques which have been used in the development of other physical sciences are not available to the meteorologist at the present time and he must search elsewhere for his method of attack upon this formidable problem.

According to a prevalent misconception, statistical methods mean the neglect of root causes and the substitution of shadow for substance. To be sure, statistical methods often lead to useful rules of thumb even when qualitative understanding is absent, but these rules should never be regarded as anything but makeshift. Doubtless the facility with which empirical recipes can be invented, together with a natural human tendency to let well enough alone, has engendered a willingness on the part of some statistical practitioners to content themselves with ersatz. But we do not hold medicine in contempt because of quackery, and it would be fallacious to condemn statistics on account of superficiality. It is perfectly true that at the present time too many meteorologists and statisticians attempt to indulge in an orgy of correlation analysis. Because the processes which they are examining lack the property of being stationary, results are obtained which are inconsistent from one period of time to the next. This difficulty becomes less and less important if a certain amount of physical reasoning is the basis for the correlation; and as time goes on and more is understood about the phenomenon, this operational approach of examining the physics of the situation will correct much of the condemnation which is now directed at statistical studies. Thus the future of meteorology depends decisively upon an enlightened and energetic use of this important tool.

The meteorologist is primarily interested in the behavior of his phenomenon as a function of time, and all his observations are either taken continuously in terms of this variable or are taken at discrete points along the time axis. Each one of the series, therefore, which are considered in the problem of prediction of single meteorological elements or combinations of them can be looked at from the time-series point of view. Since the beginning of World War II, considerable impetus has been given to the development of generalized harmonic analysis because of its importance in the analysis of time series which occur not only in the weather problem but also in all fields where prediction is important. Unfortunately, most of the work has been done in the field of what might be termed *stationary time series.* In this type of series the actual dynamics which motivate the series itself is assumed constant from one period of time to the next.

Considering the characteristics of these time series by themselves, we note that the sequence of observations has certain inherent dynamic properties as well as a superimposed random component. This random component may well be due to other variables not considered. The actual statistical problem is to separate the dynamics from the random element so that the dynamics can be studied and classified. This situation is entirely analogous to a similar one encountered when information is transmitted by mechanical or electrical methods. In this case, the signal frequently becomes distorted and when this distortion phenomenon is of a purely random statistical nature it is called "noise." Hence, as also in weather phenomena, the thing that we observe, or the message, contains the original signal plus a random noise, and the problem is to isolate the signal from the message.

In the case of pressure, if we consider the sequence of values as a continuous function, the behavior is analogous to the response of a sluggish oscillator to a series of impulses, where the dynamics are represented by the parameters of the oscillator. Because the oscillator is sluggish, the effect of the impulse at any one time is not immediately worn off but is carried over to affect the pressure at a future time. Even the random components, introduced probably by the variables not considered, affect the future value of the pressure variable since they are equivalent to an additional impulse on the oscillator.

The problem of weather forecasting can therefore be thought of in the following terms: In Fig. 1 the solid

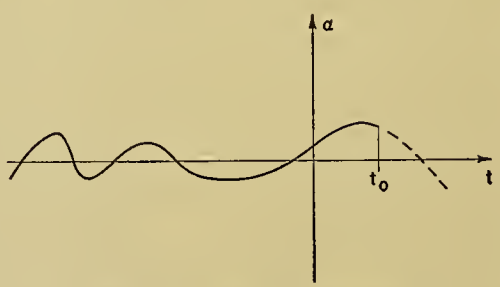


Fig. 1.—Schematic time series.

line represents the past performance of some meteorological phenomenon α. For example, it might be past values of the pressure at a single station or even a characteristic of an entire synoptic map. By analyzing this sequence from some period of time in the distant past up to the present time t_0, the predictable part of this sequence is to be determined. This predictable part represents the true dynamical component, and it is this component which permits one to extrapolate the solid line into the future to give a prediction (dotted line). Naturally, the further one extends this prediction

the worse the agreement is between the predicted and actual values if there exists a random component. In other words, even after the entire past of all the variables has been exploited, there may exist an upper limit to the attainable accuracy of prediction; this limit decreases as the forecast period increases. The complete problem, then, is the consideration of an ensemble of such sequences, all of which are known in the past up to a specific present time t_0, and, regarding them as a group, the evaluation of that component which represents the true mechanism through which the phenomenon is operating.

In the theory of generalized harmonic analysis as it affects the prediction of time series, we are approaching the problem from a more general point of view than either that of harmonic analysis, where only certain frequencies or periodicities were assumed to be present, or that of difference equations where a differential relationship is assumed connecting present and past values of the time series. In generalized harmonic analysis, advantage is taken of the entire spectrum of frequencies, and this spectrum in most geophysical phenomena does not consist of a series of discrete lines (exact frequencies) but rather of a continuous distribution of frequencies over the entire range of frequency, usually from zero to infinity. This technique permits us to develop the best linear operator in a very general sense in order to predict the future. Actually, it is this ability to predict which expresses how much dynamics there exists in the series. The linear operator is a weighting function which is applied to past values of a time series in order to obtain estimates of future values, and this weighting function is determined mathematically in such a way that the mean square error of prediction is minimized. Actually, of course, this technique can be extended to more than one variable, and thus many time series can be considered simultaneously.

Unfortunately, the basic arguments and ideas are true only if the process or processes are stationary. This, of course, cannot be true in meteorological phenomena since the basic movements of the weather systems certainly are different, at least from season to season. To a certain extent this lack of stationary property can be overcome by dividing the data into what might be termed more homogeneous periods, but, nevertheless, this difficulty is inherent in a simple analysis since the seasons themselves do not arrive and leave at any fixed time. Another factor which markedly affects the analysis of time series of this type is the fact that most of the elements have definite trends which are functions of the time of year. It is always possible to compute what might be termed a normal by averaging over a large number of years, but it is extremely doubtful whether this normal has any particular significance for a specific year. The dynamics are determined by the movement above and below a state of equilibrium which might be called the normal for a specific year, although it unquestionably is different from the over-all normal for all years. The shift of this trend line from one year to the next has a very marked effect upon the dynamics. Since it is necessary to deal with relatively short periods of time in order to obtain compatible data, it is very hard to separate that part of the spectrum which deals with long-period phenomena from the short-period oscillations which tend to dominate a rather minute portion of the entire record.

Actually, of course, the solution of a nonstationary time series has no meaning in the strict sense. The basic problem is to make the time series approximately stationary through appropriate changes in variables as a function of time. This approach will be increasingly successful as a clearer understanding of the mechanics of the phenomena is attained. In order to appreciate the immensity of the task of solving this problem, let us review briefly what is known about the behavior of some of the meteorological elements, utilizing the simple hypothesis of a stationary time series, and consider what can be gleaned from this information as to possible future courses of fruitful action.

One often hears the statement that abnormally warm winters and cold winters recur in a definite cycle. On this basis one might expect to find long-term effects in the analysis of the frequency-density function of average monthly temperatures taken, of course, as deviations from normal. If one considers either the monthly temperature or the daily temperature at any one particular locality, it may be regarded as an individual time series in itself. Actually there exists on the average a correlation of only about 0.35 between the predicted and observed values when one predicts one month in advance. This accounts for only around 10 per cent of the variability, implying merely that there is slightly more than a 50-50 chance that if the present month is above normal the next one will also be above normal. No matter how far one goes back in the past to obtain information (which means no matter how high an order of a linear differential equation one considers as representing this phenomenon) no additional information seems to be available, so that the past of the sequence has practically nothing to do with the future. Naturally other variables exist which might improve this situation, but we are dealing with a phenomenon which in itself shows no statistical regularity as far as long cycles are concerned. On the other hand one may be interested in a daily 24-hr forecast of temperature. If one restricts the mechanism to a given month during the year, a developmental sample based on four or five years will give a good estimate of the linear operator. In a 24-hr forecast, a correlation of about 0.75 between the predicted and actual values can easily be realized. Furthermore, interestingly enough, this same correlation will be obtained if one picks any other year at random for the same month and applies the operator to this other year. This shows that there exists a certain amount of high-frequency oscillations which are similar from year to year and actually represent dynamics in the time series. In other words there exists a true signal which can be separated from a disturbance which we know nothing about but designate as random noise. The properties of the air mass can be introduced by considering a network of

stations and using the entire group as predictors for an individual station. Correlations as high as 0.97 for 24 hr and remaining between 0.80 and 0.90 for 48- and 72-hr intervals show that a certain degree of statistical regularity is demonstrably present. Despite these high correlations, the forecasts leave much to be desired, for the linear operator seems incapable of detecting sudden changes in weather regime. Unfortunately, nonlinear techniques, in the elementary sense, have not yet succeeded so that it is impossible to improve the forecasts without some form of subjective analysis, such as a synoptic classification of the regime. Unquestionably there exist additional variables which, if clearly understood, would change the linear dynamics slightly from year to year and permit better predictions.

Equivalent studies have been made on pressure, including daily mean pressures, nonoverlapping five-day means, and running averages taken over five, fifteen, and thirty-one days. No long-term predictable components in any one of these time series taken at many localities over the Northern Hemisphere ever produced spectra which would indicate any long-term periods. Once again, networks involving a group of stations surrounding the one to be predicted give the best daily predictions. Correlations in the neighborhood of 0.70, which remain stable from year to year, indicate the order of the linear component, while four or five stations surrounding a given station at a distance of about 500 mi seem to produce all of the dynamics of a linear type which is in the network. Thus, increasing the network of stations either in number or in radius seems in no way to increase this predictability. Studies of this nature seem to afford ample proof that the synoptic situation expressed in terms of present temperature, pressure, humidity, etc., does not uniquely determine the future distribution and that the behavior of the pressure distribution outside this ring of 500 mi does not have much influence upon the future of the pressure at the center of the ring. Furthermore, if a forecast is made for 24 or 48 hr in advance by predicting a group of stations individually (considering each one as made up of an individual network), the correlation of the actual predicted map with the observed conditions is not much better than the correlation that one would obtain between predicted and observed pressures at an individual station. This indicates that the errors of forecast are correlated over the entire country and that we are not making a series of individual random predictions. The fact that the errors prove to be highly correlated means, of course, that an outside influence has moved into the picture and that all of the mechanisms fail to predict simultaneously. This brings out rather clearly the concept of changes in regimes in the entire weather pattern, followed again by a stabilization of whatever linear component exists. Once the regime has established itself, pressure changes of considerable magnitude can be forecast quite accurately, but the physical bases for the changes must already exist within the network. Six principal conclusions can be arrived at through a close analysis of pressure forecasting utilizing straightforward statistical techniques: (1) The performance of the linear forecast from year to year is stable in that the mean errors of prediction show no significant difference between years and the root-mean-square errors are homogeneous, (2) a small network is adequate, (3) errors cannot be ascribed to improper weighting of the predictors, (4) it is extremely doubtful that linear prediction can be improved upon by using fixed functions of higher degree, and therefore additional variables not considered at the present time are important, (5) the errors in the forecast are not caused primarily by a failure to indicate pressure changes as such, but rather by an inability to predict extreme values whether they represent changes from the initial situation or not, and (6) statistical forecasts fail under the same circumstances that synoptic forecasts fail and apparently for the same reasons.

In order to give a little more background to the direction that future research should follow, let us consider for a moment the combined motion of the four major centers of action for the North American continent, namely, the semipermanent highs and lows—the Atlantic high, the Pacific high, the Icelandic low, and the Aleutian low. It is in the motion of these centers of action that one can first observe the fact that considerable dynamics exists in the atmosphere and that it is not a random phenomenon nor anywhere near random. If we consider any one of these cells as the one to be predicted and utilize its past and the present and past of the other three cells as predictors, prediction for a considerable time in advance is possible for any particular year. One can think of this prediction process, of course, merely as a reproduction of the data after the fact, but nevertheless the reproduction of the data by means of the operator is so startling that it deserves notice. Both the spectra obtained from the autocorrelation functions and those obtained from the cross-correlation functions are extremely active, showing that the density of frequency is located in rather narrow bands. The graphs of these spectra are, however, quite different from year to year, indicating that we are dealing with a phenomenon which has certain specific characteristics at any one time but that these characteristics vary from year to year because of some outside influence. The sharp spectral characteristics which occur in these control cells unquestionably lead oftentimes to the wrong conclusions, namely, that periodicities exist in the atmosphere in the sense of being true line spectra. For periods of time, therefore, one can observe certain fluctuations of these meteorological elements and erroneously conclude that there is a law, perfectly determined, which is influencing their behavior. This is the clearest indication of why it is possible to tie up cycle theory with weather phenomena of various types, particularly for specific short periods of time.

Although a large amount of work has been done by many investigators in computing correlations of pressure at one point in the atmosphere with that at another point, correlations of precipitation and tem-

perature with the forward movement of ice in the North Pole region, etc., the illustrations set forth in the preceding paragraphs give a picture of about the order of magnitude of the results which one can expect to find in setting up linear hypotheses involving pressure, temperature, humidity, etc., either singly or in combination and at one station or in the network. With this background, I should like to discuss possible future courses which statistical analysis can follow toward the solution of this complicated situation involving dynamics and thermodynamics.

Perhaps one of the greatest contributions which statistical methods can, at the moment, make to the field of meteorology is in the classification of climatological information. It is well known, for example, that the forecasters in the short-range verification program during World War II greatly improved their accuracy by being given the probability of the occurrence of the events for which they were forecasting. This climatological information, however, is only part of that which is available to the forecasters to improve their forecasting techniques. There exist certain statistical laws which hold individually for all climatological elements and for all possible locations. These relationships have very profound effects upon the distribution of these elements for a given week or a given month and are entirely different from the over-all distributions that are at present given as the climatological expression of the variability. It is the peculiar behavior of these daily, weekly, monthly, and seasonal distributions at various localities which are the real characteristics of the climatology and could be extremely useful in understanding better the weather processes themselves and also in improving the forecasts at a particular location. Thus a re-evaluation of the whole method of presenting the climatology of various localities and the individual behavior of climate during reasonably short periods of time at a particular locality might do much to increase our knowledge concerning the general behavior of the circulation. In fact, it might be stated that if the basic characteristics of the climatology were plotted on a Northern Hemisphere map, obvious simultaneous relationships could be studied and the whole forecasting problem could be reduced to a group of much more specific questions.

A great deal of thought has been given by many meteorologists to the use of analogues as a method of forecasting. The basic argument against the use of analogues is, of course, that the weather never follows exactly the same pattern. However, in the light of past experience in forecasting by statistical methods, it is obvious that in order to progress much further it is going to be necessary to find a way of classifying the dynamics. Dynamic classification might be achieved through the use of analogues, and in that event it should be possible to utilize the correct statistical operator for the prediction mechanism. Many attempts have been made by various people to express the dynamics of a present system in terms of some linear combination of the dynamics of past experience or similar situations. This seems to be extremely logical.

One can set up a practical analogy. An engineer wishes to determine the dynamics of a particular machine which, for some reason, he is unable to analyze completely. It might, however, be possible for him to analyze three or four machines which are similar to the original machine but differ in some minor details. It is then quite possible that by observing the dynamics of the similar machines he may infer how the machine in question will behave and operate. Actually if the dynamics are changing, in the sense of the series' being nonstationary, about the only way that one can judge what these dynamics might be in the future is through the use of such a technique. Our failure up to this time to extrapolate the dynamics properly from the known situations is probably due to the fact that the parameters used to express the synoptic picture are not dynamic in themselves. It is therefore necessary for us to solve more completely the following problems:

1. Parameters must be found that will adequately classify the static picture of the weather situation over a large area. They should represent the general features, particularly the flow patterns, over a region even though they might not bring out every individual detail. The analogue technique calls for a classification in terms of a general flow pattern, and such a classification would tend to eliminate the influence of extremely random fluctuations which hinder progress in fathoming the dynamics. Whatever parameters are used should maximize the discrimination between different synoptic situations.

2. Parameters should be found which actually classify the dynamic features of large areas. Having real physical significance, they would have some chance of being associated with both the static picture and the dynamic processes of the weather elements over an area.

3. The relationship between the static and dynamic properties of the parameters should be distinguishable so that it would be possible to go from one to the other and realize what is important from both points of view.

Because the entire Northern Hemisphere affects the weather patterns at one single locality, it would seem advisable to attempt to make these parameters of a Northern Hemisphere character. If the characteristics of the general flow could be determined by examining the dynamics of past situations and extrapolating them in order to obtain the dynamics of the present one in the correct manner, then it is entirely possible that the network theory of simple linear hypothesis would be sufficient to predict the individual weather at various localities by the use of an operator fitted to a correct dynamic model.

Although none of the problems mentioned above have to my knowledge been solved as yet, there are certain specific conclusions about analogues in general which seem to indicate that the method holds some hope for future development: (1) When the correlations between the sequence of past and current maps main-

tain themselves with a reasonable value for at least three days, the analogues are, on the average, synoptically and kinematically sound, (2) for the purpose of making long-range forecasts up to ten days' duration there usually exists one analogue which gives a precipitation forecast as good as, if not better than, the present 24-hr forecast (this, of course, is after the fact, and at the present time no method is available for detecting such an analogue in advance), (3) there are indications that certain previous temperature distributions aid materially in picking that analogue, and (4) maps which show the same general movement of the essential features are usually good analogues.

In order to determine why it is that the major centers of action have such variable dynamics from year to year, it may be desirable to examine more carefully variations in sea-surface temperature. Preliminary studies have indicated that in certain regions the sea surface has undergone rather marked changes in temperatures and that the variation of the difference between the air temperature and the water temperature shows considerable change from year to year. Since the water temperature is a very persistent phenomenon, these long swings of above and below normal over a period of many months may have considerable influence on the behavior of the weather. Also, since about three-quarters of the world is ocean, perhaps too little thought has been given to these variations as potential sources of the fluctuations in the dynamics of the weather system.

There are, of course, many variables which should be considered in attempting to get a complete statistical model, particularly in light of the fact that we are dealing with a servomechanism. The low-pressure area produces the clouds, the clouds change the energy input, the change of the energy input causes a change in the energy of the system, and that in turn affects the movement of the low in its future course. Sunspot activity, variations in total pressure over the poles and over the equatorial regions—all these must contribute their part. It therefore seems advisable that investigations utilizing the statistical approach be greatly increased so that more and more information will be available, thus clarifying the nature of the phenomenon with which we are dealing. As this statistical information establishes the variables which have dynamic properties, it should be possible to set up mathematical models on the bases of reasonable hypotheses inferred from the observational facts, and in this way we may hope that the final solution of the weather forecasting problem will be reached. On the other hand, it might develop that meteorology is faced with a situation analogous to that treated in the kinetic theory of gases, wherein the behavior of an individual molecule is impossible to predict, and only certain average features of all the molecules can be computed. Nevertheless, if the problem has to be reduced to this type of conclusion, it would at least be possible to estimate the respective probabilities for the occurrence of each possible state.

Thus it appears that if the general forecasting problem of meteorology is to progress as rapidly as possible toward its solution, meteorologists and statisticians should combine their efforts in a more united fashion. It is clear from the type of problem with which we are dealing that it is only through the statistician's understanding the meteorological point of view and his being able to understand and utilize the experience of the men who have been watching the progress of weather situations for a long period of time that he can be effective. Reciprocally, the meteorologist cannot take advantage of statistical techniques which would be of great material aid until he himself understands the basic principles upon which the techniques are based and also understands what the statistician wishes to accomplish.

REFERENCES

In the following publications the reader will find an account of applications of statistics to weather forecasting.

1. Baur, F., *Einführung in die Grosswetterkunde.* Wiesbaden, Dieterich, 1948.
2. —— "Der gegenwärtige Stand der meteorologischen Korrelationsforschung." *Meteor. Z.*, 47:42–52 (1930).
3. —— "Die Störungen der allgemeinen atmosphärischen Zirkulation in der gemässigten Zone." *Meteor. Z.*, 54:437–444 (1937).
4. Darling, D. A., *Report on the Accomplishments of the Statistical Project.* Calif. Inst. Tech.–Army Air Force Res. Unit, 1942.
5. —— *Relationship between the Analogue Forecast and Other Methods of Forecasting.* Calif. Inst. Tech.–Army Air Force Res. Unit, 1944.
6. Elliott, R. D., *Extended Weather Forecasting by Weather Type Methods.* U. S. Navy Dept., Long Range Weather Forecasting Unit, Washington, D. C., 1944.
7. Haurwitz, B., "Final Report on the Use of Symmetry Points in the Pressure Curves for Long-Range Forecasts." U. S. Air Force, *Air Wea. Serv. Tech. Rep.* No. 105–7 (1944).
8. Namias, J., *Extended Forecasting by Mean Circulation Methods.* U. S. Weather Bureau, Washington, D. C., 1947.
9. Schell, I. I., "Dynamic Persistence and Its Applications to Long-Range Foreshadowing." *Harv. meteor. Stud.*, No. 8 (1947).
10. Schmauss, A., "Synoptische Singularitäten." *Meteor. Z.*, 55:385–403 (1938).
11. Schumann, T. E. W., *Statistical Weather Forecasting.* Meteor. Res. Bur., Pretoria, Un. of South Africa, 1947.
12. Starr, V. P., *A Physical Characterization of the General Circulation.* Dept. Meteor., Mass. Inst. Tech., Rep. No. 1, General Circulation Project No. AF 19-122-153, Geophys. Res. Lab., Cambridge, Mass., 1949.
13. Wadsworth, G. P., "Short Range and Extended Forecasting by Statistical Methods." U. S. Air Force, *Air Wea. Serv. Tech. Rep.* No. 105–38, 1948.
14. —— *Analogue Techniques in the Forecasting of Rainfall.* Geophysical Research Directorate Reps. No. 1, 2, and 3, 1948; *Dynamics of the Semi-permanent Pressure Cells*, Rep. No. 4, 1949; *Further Analysis of Dynamics of Major Pressure Cells*, Rep. No. 5, 1949; *Position of Major Pres-*

sure Cells in Relation to Rainfall, Rep. No. 6, 1949; *A Search for Pertinent Parameters to Classify Pressure Distributions*, Rep. No. 7, 1949; *Statistical Prediction of Migratory North American Anticyclones. An Introductory Attack upon the Non-Stationary Time Series in Weather*, Rep. No. 8, 1950; *Preliminary Studies on Variations of Sea Surface Temperatures*, Rep. No. 9, 1950. Contract No. W28-099-ac-398, Mass. Inst. Tech., Div. of Industrial Cooperation, Cambridge, Mass.

15. Walker, Sir Gilbert T., "Symmetry Points in Pressure as Aids to Forecasting." *Quart. J. R. meteor. Soc.*, 72:265–283 (1946).
16. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, Mass., Technology Press of Mass. Inst. Tech.; New York, Wiley, 1950.
17. Willett, H. C., and others, *Final Report of the Weather Bureau-Massachusetts Institute of Technology Extended Forecasting Project for the Fiscal Year 1948–1949*, 109 pp. Cambridge, Mass., 1949.

TROPICAL METEOROLOGY

Tropical Meteorology *by C. E. Palmer* 859

Equatorial Meteorology *by A. Grimes* 881

Tropical Cyclones *by Gordon E. Dunn* 887

Aerology of Tropical Storms *by Herbert Riehl* 902

TROPICAL METEOROLOGY

By C. E. PALMER

Institute of Geophysics, University of California

INTRODUCTION: THREE METHODS OF ANALYSIS

In 1947, when the University of California established its Institute of Geophysics at Los Angeles, one of the research projects initiated with the Institute was an investigation of the synoptic meteorology of the tropical Pacific Ocean. Although the literature contained only scanty references to the topic, it was known that many meteorological data had been collected during World War II and that new ideas had been fermenting in the area throughout the period of conflict. In order to obtain some insight into the general condition of synoptic meteorology in the Pacific as a preliminary to the investigation the Director of the Institute at that time (Dr. J. Kaplan) issued a questionnaire to meteorologists who were known to have been associated with Pacific observation and research during the war, and answers from about half of those addressed were finally received in the Institute. It would serve no good purpose to enter here into a detailed analysis of the replies; however, one striking feature revealed by the survey is relevant, since it turns out to be characteristic of the present state of tropical meteorology in general. The replies fell into three well-marked classes, distinguished by the method of approach to tropical weather problems displayed, in some cases consciously, in others unconsciously, by the authors.

The meteorologists whose replies fell into the first class tended to put the greatest emphasis on climatological concepts: even in their approaches to synoptic problems they were dominated by the results of statistical analysis, by mean maps, by statistical notions like the *trades*, the *monsoons*, the *doldrums*, by the conviction that the day-to-day weather in the tropical Pacific differs very little from that revealed by monthly and annual means. They extended this approach to the forecasting problem. Some felt, for example, that the best guide to forecasting the tracks of individual typhoons and tropical storms might lie in the study of mean storm tracks. Their dynamic theories, when they had any, were expressed in terms of a hypothetical general circulation of the tropics, and this was the entity whose characteristics were to be explained by physical reasoning from first principles, not the vagaries of the daily weather map. For the purposes we have in mind here, we may term the method of approach which is based upon these conceptions, the *climatological method.*

In the second group we must place those meteorologists whose thinking was much under the influence of concepts derived originally from the study of high-latitude weather, the followers of the *air-mass method.* The more extreme replies of this class maintained more or less explicitly that the tropical atmosphere differed from that in higher latitudes only in having a higher temperature and an easterly instead of a westerly mean wind direction. The dynamics, and by this they usually meant the frontal dynamics, remained the same. In the tropics there were fronts, air masses with source regions and modifications as in higher latitudes, an equatorial front similar in principle to the polar front, and above all an occlusion process that resulted in the formation of typhoons and tropical storms. Small temperature differences became important, and the "slope" of systems a matter of earnest discussion.

Finally we have the third group of replies, sent in by workers originating in or influenced by the Institute of Tropical Meteorology at San Juan, Puerto Rico, a school set up by the University of Chicago during World War II but now, so far as I know, no longer active. This group takes its origin from Gordon E. Dunn's paper on easterly waves [25] and its members have at some time been closely associated with the University of Chicago, drawing their theoretical stimulus from C.-G. Rossby. It is natural to term the method followed by this group the *perturbation method.* The concepts of the climatological group, referring to a hypothetical general circulation of trades, monsoons, doldrums, etc., identifiable as such on the mean maps, were taken over by these meteorologists. However, the basic currents were now considered to be for the most part zonal but subject to perturbations, and the perturbations to be accompanied by characteristic weather and pressure patterns, identifiable on daily weather maps. This group was much preoccupied with *models* of various kinds of perturbations. In dynamics, one gets the impression that they considered all dynamical problems connected with the perturbations as already solved by Rossby's work and that their great task was to explain the dynamics of the general and zonal circulation.

The significance of the three broad classes of Pacific meteorologists revealed by this survey lies in the fact that, once one knows what to look for, the same divisions can be discerned in the work of all tropical meteorologists. It is not that they may be divided into those whose interests are primarily in climatology, or in synoptic meteorology, or in dynamic meteorology. On the contrary, the division rests upon distinct methods of approach to all branches of tropical meteorology, statistical, dynamic, and synoptic alike. It is therefore convenient to regard writings on tropical meteorological topics as emanating from three distinct schools of thought, the climatological, the air-mass, and the perturbation schools, even though it is only in the last class that the opinions have been common to a locally concentrated group of workers. And, of course, it must be remembered that there will be many borderline cases,

writings that are hard to place unequivocally in one of the three groups. These things understood, however, we will find it easy enough to survey the present state of tropical meteorology. It will consist of the conflicts of these schools as revealed in the literature. The scope and detail of the conflicts is likely to astonish the meteorologist working on high-latitude problems, accustomed as he is to a large measure of agreement on the fundamental descriptions of temperate and high-latitude weather. In the tropics there is not even agreement on the common forms of tropical clouds or on the meteorological conditions accompanying precipitation. It has become the custom to generalize in haste and to reject inconvenient observations at leisure. Large areas of the tropical atmosphere have not been explored by observation and even the better-known areas have not yet yielded statistics of sufficient scope or reliability to justify the tropical parts of the "models" which are incorporated into the textbook descriptions of the general circulation of the atmosphere. In synoptic meteorology, conditions are little better than in climatology. The role of water substance in the genesis of tropical depressions is a matter of dispute, the existence or nonexistence of fronts the subject of irreconcilable speculations. A dynamical meteorology of the tropics can hardly be said to exist. The assumptions of such work as has been done are usually overlooked, and the results applied in a manner quite beyond the expectation of the original authors. Worse than this, however, is the tendency to apply dynamic theories depending for their validity on assumptions that hold in high latitudes but not in low latitudes. For example, the extent to which reasoning resting implicitly on the geostrophic-wind equation has been applied in equatorial meteorology is disquieting. This state of affairs, however, should not give occasion for pessimism; it is a sign that tropical meteorology, the oldest branch of scientific weather study, is undergoing a new development and that that development is a vigorous one. Each school of thought has had its achievements, has clarified some previously obscure field of knowledge, and has had its defects subject to sharp criticism. It is our task now to estimate the extent of these achievements and to mitigate the evil of those defects.

THE CLIMATOLOGICAL METHOD

There is no doubt that, at least over the oceans, the observed values of the meteorological elements deviate from their monthly, seasonal, and annual means to a far lesser degree in the tropics than in higher latitudes. The mean values therefore can be said, in a special sense, to be more representative of the synoptic values. The steadiness of the wind, for example, has long been known, and the representativeness, in the above sense, of the mean trades and mean monsoons is perhaps the most ancient fact of modern meteorology, antedating in its origin the very notion of a synoptic map.

The word "trade" itself is a contraction of "to blow trade," that is, to blow constantly in the same direction, along the same track. The empirical discovery implicit in the expression has, of course, been well confirmed by scientific investigation. Gallé,[1] for example, has given an analysis of the *persistence* of the wind in various parts of the Indian Ocean. The *persistence* may be defined as the ratio of the mean vectorial wind speed to the mean speed without regard to direction. If, over the period covered by the means, the wind has the same direction, the persistence would be 100 per cent; on the other hand, if the directions were distributed at random, the persistence would be zero. According to Gallé, the persistence of the trade in the square 10°S–20°S, 80°E–90°E does not fall below 69 per cent and for most months lies between 80 per cent and 90 per cent. This can be contrasted with the corresponding figures for the square 35°S–45°S, 70°E–80°E, where the persistence over any month does not rise above 57 per cent and, in the mean, amounts to 45 per cent for the year. But Gallé's figures merely give quantitative expression to the conclusion which anyone can draw from the surface-wind maps that are published from time to time by national meteorological organizations [69]. Scrutiny of the wind data on these maps confirms not only the great persistence of the trades but also a similar large persistence, within the seasons of their greatest development, of the monsoons. To a lesser degree the maps also give one the impression of *uniformity*: over a very large proportion of the tropical oceans, the predominant component of the wind is the east component and this, combined with the known persistence of the trades, leads even careful workers to fall into the habit of regarding all tropical west winds occurring outside the monsoon regions as anomalous.

At first sight the representativeness of the means of meteorological elements other than the wind appears not to be so striking. Temperature and pressure, for example, show synoptic variations that may depart widely from the monthly, seasonal, or yearly means. However, the periodic nature of these variations is evident, the pressure showing semidiurnal oscillations whose amplitudes are largest at or near the equator, while the temperature has a clear diurnal periodicity whose amplitude becomes less the more closely the conditions of observation approach those typical of the open sea. When allowance is made for the periodicity and for calculable orographical effects by correction of the synoptic observations, the mean values for a month or a season turn out to be as highly representative of the corrected synoptic values for temperature and pressure as for wind. Further, in the case of temperature, remarkable uniformity is disclosed. Over vast stretches of the western Pacific, for instance, the temperature of the surface air during the whole "wet season" varies little from 85F [61]. Even in monsoon regions, where we might expect to find large seasonal variations of temperature, remarkable instances of uniformity are known.

> At Colombo the mean difference [sic] in surface temperature between the North-East and South-West monsoon seasons is less than 2°F. and just about as much rain falls

1. Quoted in [36, p. 45]. The original paper is not available to me.

during one monsoon as the other, notwithstanding that the North-East monsoon starts out as a cold, dry continental air mass and the South-West monsoon as a warm, maritime air mass [46, p. 27].

Connected with the uniformity of surface temperature over the oceans is the well-known uniformity in the height of the base of tropical cumulus. This rarely is less than 1500 ft or greater than 2000 ft, and departures from these values are so clearly connected with the occurrence of precipitation that the problem of forecasting cloud bases reduces to that of forecasting the presence or absence of precipitation.

The persistent and uniform values of the meteorological elements are correlated. Both the trades and the monsoons are more than winds blowing with great regularity. Other elements, such as the precipitation and the cloud amount, the temperature, pressure, and humidity regimes, and the vertical variation of the elements all have structures or values characteristic of the trades or of the monsoons. Bergeron [5] has emphasized these correlated characteristics and has held up the trade and monsoon of the tropical climatologist as paradigms of concepts that should be used in the air-mass climatology of higher latitudes. As an example he pointed out that "good monsoon" in India means a complex of heavy rains, strong steady southwest wind, and a characteristic vertical temperature and humidity distribution. The term monsoon, therefore, is to be applied to a persistent kinematic and thermodynamic complex. Similarly numerous observations, culminating in the *Meteor* results [18], have shown that the trade is a persistent complex. Here again we have not only a characteristic and persistent wind, but also a typical associated cloud form, precipitation regime, and lapse rate of temperature and humidity. The "trade-wind inversion" is indeed well known and the equally well-known trade cumulus owes its peculiarities to the association of the inversion with a typical vertical wind-shear.

So far we have mentioned the representativeness of the statistical parameters that are derived from observations over the great tropical oceans. At first sight, conditions are different over the land, particularly over the continents. The highly developed land and sea breezes of tropical islands, and the thunderstorms that break out with great regularity in the archipelagos of the East and West Indies are departures from mean conditions that are, through travelogue and fiction, known even to the layman of high latitudes. But these departures are clearly periodic, and synoptic observations may be corrected for them, either quantitatively or qualitatively. The *seasonal* means for the tropical islands are then representative of these corrected synoptic values. Moreover, the diurnal departures from the mean can be so clearly related to astronomical and orographical causes that the short-period forecast seems to depend only upon an adequate knowledge of the statistics and of geography. Upon close examination the same principle seems to apply to forecasting in continental regions. Here we have the added advantage that the seasonal variations themselves appear to be due to the same causes as the diurnal changes, though operating on a larger scale and over longer periods of time. On this view, we may take the slowly varying oceanic conditions as the ground state and regard both diurnal and seasonal variations over the land as perturbations whose causes, orographic and astronomical, are known. The perturbations are thus thermal in their immediate origin and there appears to be no problem of explanation so far as synoptic variations are concerned. Similarly there ought to be no forecasting problem for periods shorter than a season that cannot be solved with an adequate knowledge of statistics and of orography.

In the foregoing account I have outlined an extreme climatological view of tropical meteorology. On this view, there are no synoptic problems in the tropics. Synoptic problems become climatological problems; the future advance of tropical meteorology therefore depends on the collection and reduction of longer and longer series of observations from more stations in order to obtain maps of stable statistical parameters and their seasonal variations which are to be combined with a more detailed knowledge of orographical peculiarities. The remaining problems are then dynamical and relate almost entirely to the "ground state" that was mentioned above. If corrections for diurnal and seasonal variations may be made by assigning them to known astronomical and orographic causes, it is clear that we are left with the statistical picture of the "trades" considered as kinematic-thermodynamic complexes to be explained by the methods of mathematical physics. Since the complexes are regarded as the most persistent features of the atmosphere, they represent a steady state in the dynamic sense and the statistics give us a direct insight into what is known as the general circulation, at least for latitudes below 20°. What has to be explained, therefore, is the tropical part of the "planetary circulation" or "winds of the globe" and this explanation turns out to be one of the easiest ever attempted in meteorology. If it should appear that I have represented a very extreme view and one which would not be defended in this form by any modern meteorologist, I suggest that almost all textbooks that refer to the tropics at all (*e.g.* [49]), and certainly some recent authors of research papers dealing with the general circulation [45, 56], imply this view in their treatment of the dynamics of the tropical atmosphere. This is true not only of those we may rightly classify as belonging to the climatological school but also of those in the frontal and perturbation schools, insofar as they treat the general circulation. Those authors will be dealt with in their proper place.

What kind of picture, then, is obtained by pursuing these principles to their final conclusion? First, let us look at the now-modified empirical laws governing the general circulation in the tropics. Abstracting the data in such a way as to omit the diurnal and seasonal perturbations of thermal and mechanical origin, we are left with a system of trade winds, blowing steadily from the east and toward the equator in the lower layers of the atmosphere. The trades of the two hemi-

spheres approach one another at the equator, as is clear from the statistics. There, they become light and variable and this weakening is obviously part of the observed horizontal velocity convergence in the two currents, with a correlated increased vertical component in the motion. The lapse rates of temperature and humidity confirm this conception. At some distance from the equator, the lower atmosphere is characterized by the presence of an inversion of temperature, the trade-wind inversion, above which the humidity drops to very low values, both absolutely and relatively. But this inversion, under the influence of the predominantly vertical motions, becomes higher and weaker as the equator is approached, finally disappearing in the zone of intense upward convection at or near that boundary. At high levels, the motion of the air is predominantly and necessarily (on continuity considerations) away from the equator, and at relatively small latitudes it takes on a westerly component to form the *antitrades*. The antitrades are thus westerly winds with a poleward component and a downward vertical component that reaches its maximum values in the neighborhood of 30°. A circulation cell is thus completed and the composition of the air in it agrees well with the distribution of the vertical component of the steady-state motion. In particular the region of light variable winds near the equator, which we shall now call the *doldrums*, is, as the statistics show, also a region of large cumulonimbus clouds, with attendant heavy precipitation, the "squalls" of the early navigators. On the other hand the trade-wind zone is the region of trade cumulus, with only light showers or none. The *horse latitudes*, finally, are regions in which fair-weather cumulus or stratocumulus predominate; sometimes they are clear of cloud.

When we come to explain this abstracted picture of the tropical circulation we find the necessary concepts ready to our hands. The equatorward component of the trades was first explained by Halley in 1686; this constitutes, in fact, the first attempt to explain any feature of the general circulation. According to Halley, the equatorward component is to be attributed to the replacement of air that has risen by upward convection at the equator. Surface heating at the equator, supposedly the region receiving the greatest amount of heat from external sources, is sufficient on this view to drive the whole meridional circulation of trades and antitrades. Halley's explanation of the *zonal* components of the motion, however, was not accepted. In 1735 John Hadley gave the accepted theory, invoking the conservation of absolute angular momentum. Since that time the combined explanation has been handed on from text to text. It was, for example, embodied in the earlier dynamic theories of the Chicago group. Rossby [56] specifically has used it under the guise of the "direct cell" in the general circulation. Although little quantitative work has been done, the qualitative use of the explanation, it is assumed, solves the dynamic problem for the tropical atmosphere. On this assumption, any further work would be directed to improving the theory quantitatively, by referring to better and better climatological data.

Summing up the extreme climatological view, we have the following conceptual apparatus: the tropical atmosphere is composed of persistent wind systems, with their associated composition and thermodynamic properties, the *trades* and the *monsoons*. Diurnal variations of the meteorological elements are due to perturbations of these systems, attributable to thermal and orographic disturbances of known origin. The monsoons are seasonal perturbations of the trades due to the same causes. The trades themselves are thermally driven and all features of that circulation are explicable in terms of what happens in the *doldrums*, a region of intense convection and light variable winds coinciding with the equator in the ideal model. Astronomical variations, combined with the different absorptive properties of land and sea, therefore account for everything—trades, doldrums, monsoons, land and sea breezes, the regime of precipitation, temperature and humidity, cloudy and fine weather, and lastly (by dynamic functional relations embodied in the equations of motion) the pressure distribution. There are no problems left to solve. On this basis, a large part of the more successful short-range tropical forecasting in World War II was implicitly conducted; the little long-range forecasting that was done was based explicitly on this statistical method; on this basis, also, most present-day models of the general circulation incorporate the tropical atmosphere.

Two unfortunate circumstances mar the beauty of the picture. The first is a fact that has been known to western science for four hundred years or more. Oceanic tropical regions are not only the seat of steady trades and monsoons, they are also, in their season, notorious for the presence of the most violent storms dealt with by synoptic meteorologists—typhoons, hurricanes, and tropical cyclones. How are these storms to be accounted for in terms of the causal theories of the climatological school? The first impulse is to explain their origin in terms of the standard astronomical and orographical variables previously invoked. During the last century, the tendency was to regard tropical cyclones as being perturbations due to local intensifications of convection in the doldrums [43]. Those who wished to advance more explicit causes attributed the convection to local maxima of heating [27]. But this explanation was soon seen to be untenable, since the region of origin of at least some typhoons was discovered to be over the parts of the Pacific Ocean where the temperature gradients in the lower layers were very weak. It must be confessed that the climatological school never successfully dealt with the problem of the origin of tropical storms. The best that could be done was to chart the regions and seasons in which they occurred, to show that the mean tracks generally followed the march of the sun and that there was some justification in the statistics for regarding tropical storms as connected in some way with the only region, in the view of the school, that showed the necessary high variability of the meteorological elements:

the doldrums. But even the doldrum origin of tropical storms was not quite general.

At least some of the hurricanes that afflict the West Indies and the southern coast of the United States originate in the Caribbean or in the Gulf of Mexico, and when this happens they appear almost without warning in the trade current, not in the doldrums [25].

The second circumstance that is fatal to the extreme climatological view is that the really steady trades occupy only a small fraction of the total oceanic areas in the tropics. Further, the antitrades are similarly found only in comparatively restricted areas. The typical trades, as described in models of the general circulation, are found only in the eastern parts of the equatorial branches of the great subtropical anticyclones. The regions of high persistence are these same areas. In the western fractions of the tropical oceans the trade wind is variable, not as variable as the belt of the westerlies, it is true, but still subject to perturbations of the wind field, so clearly independent of thermal causes as to belie the classical description. In the western oceans also, the meridional component of the wind may not be directed toward the equator, even in the mean. At Pukapuka (11°S, 166°W), for example, the mean wind at the surface is directed during the summer away from the equator, though this station always lies south of the doldrums [24]. Similarly the antitrade at higher levels is best developed and most easily discovered above the eastern borders of the subtropical highs. To the west, the westerlies are found at higher and higher levels, and in the lower latitudes and near the western borders of the anticyclones they may not be present below the tropopause. The deathblow to the notion of universal antitrades in tropical latitudes was given by J. Bjerknes in his discussion of the subtropical anticyclones [6]. It is clear from that paper that the oceanic anticyclones are major perturbations of the atmosphere with features independent of any direct thermal cause, and that any explanation that leaves them out of account by dismissing the trades as part of a simple "direct cell" symmetrical on all longitudes is artificial. While such explanations may suffice in a discussion that is directed mainly toward clarifying the vagaries of the westerlies, they are too unrealistic to be useful in tropical meteorology.

A more serious difficulty arises when, having recognized this oversimplification of the general circulation, we attempt to discuss the details of the observed tropical circulations, daily, seasonal, or annual. While we have an excellent series of observations over the oceans (excellent, that is, from the climatologists' point of view), we have very few observations and those for comparatively short periods over most of the land masses. India, it is true, has long provided observations from a dense network, but we still lack long and reliable series from Southeast Asia. The data are even more unreliable and sparse from equatorial Africa and from Brazil. As we shall see later, the wartime synoptic data from the latter areas strongly suggest that the simple picture of the monsoons in those areas is probably a climatological abstraction as misleading as that of the trades. What good data we have, moreover, consists of surface observations. Upper-air observations are still so widely spaced, even at the present day, that most high-latitude analysts would refuse to commit themselves to a space analysis of either the synoptic data or the means. It must be remembered, also, that except for some Indian stations, and a few in the Caribbean and Panama, practically no radiosonde or meteorograph data antedates World War II. Further, the upper winds in India and the East Indies, upon which much emphasis has been placed in some papers [23], are derived from the observations of pilot balloons. Dr. Mintz of the Meteorology Department of the University of California at Los Angeles has pointed out to me how misleading mean winds derived from such observations in the tropics can be. In "The General Circulation of the Atmosphere over India and Its Neighbourhood" [51], for example, are published data giving the mean wind at 4 and 8 km according to the pilot balloons. In the same volume are also given the mean directions of the clouds at and about those levels. The mean circulations for January derived from these two independent sources are directly opposite to one another, that indicated by the balloons corresponding to the east end of an anticyclone situated at the appropriate level, that from the clouds corresponding to the west end. It is clear that the pilot balloons are selective and that mean winds derived from them are probably typical of clear weather; the balloons are very easily lost from view in tropical regions even with small amounts of convective cloud, since the cloud pillars are so much higher in low than in high latitudes. On the other hand, winds derived from middle and upper clouds are selective in the opposite direction, since such clouds are, in low latitudes, most commonly the result of convection, being formed from the middle and upper parts of cumulonimbus.

Such considerations as the foregoing ought to make us hesitate to generalize when discussing the upper parts of the tropical circulations, either in the trade or in the monsoon regions, especially when our point of view leads us to place the greatest emphasis on seasonal and annual means. There are literally not sufficient data to make a single reliable statistical generalization applicable to the upper levels. Our best hope is to allow the lapse of time and the gradual extension of the tropical observing networks to accumulate data that might later give us an insight into the upper circulation in low latitudes. In the meantime, the chief problems seem to be synoptic and dynamic, particularly the dynamic problems growing out of synoptic experience.

THE AIR-MASS METHOD

When, after World War I, meteorologists of the Bergen group began to extend their application of frontal and air-mass analysis to regions outside Scandinavia, they found the climatological concepts of tropical meteorology already adapted to their purpose. We have already referred to Bergeron's approval of the notions, *trade* and *monsoon*, as modified to include not

only the wind structure of those entities but also the associated temperature and humidity distributions and the typical cloud and precipitation regimes. For kinematic and thermodynamic complexes of this kind were precisely the new elements the Bergen group had introduced into the synoptic meteorology of high latitudes under the name, "air mass." Both the trade and monsoon of the climatological school of tropical meteorologists fitted the prescription of an air mass perfectly. The trade, for example, had a specific source region: the subtropical anticyclone. Its path from the source was well determined and the character of the surface over which it had to pass, the great tropical ocean, had already been investigated and its thermodynamic effect evaluated. These factors together gave to the trade its uniform, persistent features; the relatively steep lapse rate in the lower layers could clearly be attributed to the fact that it moved across the sea-surface isotherms toward the heat equator; it was clearly a cold mass of maritime subtropical origin ("tropical" in the new vocabulary) and the presence and variation of the convective cloud in it confirmed, through indirect aerology, the thermodynamic classification. The trade at its source was a stagnant slowly subsiding mass, almost in equilibrium at all times with the sea surface; here there were no clouds or perhaps small amounts of low stratiform cloud. As it moved toward the equator, however, the mass became heated by the warmer waters in lower latitudes so that a relatively steep lapse rate was characteristic of the lower layers, while the air above continued to be warmed by subsidence in the antitrades, thus stabilizing the trade-wind inversion. The closer the mass approached the equator, the more intense became the heating from below and the less the subsidence aloft; the result was that a deeper lower layer was overturned by convection, the trade-wind inversion was found at a higher level, and became weaker as it approached the doldrum region corresponding to the heat equator. The clouds used in indirect aerology confirmed this history, for as the mass moved toward the equator the cumulus reached to higher and higher levels and showers associated with cumulonimbus became more frequent. This picture not only fitted the older climatological view of the trade remarkably well, it also removed some of the difficulties of the older theory by emphasizing the role of the subtropical highs as source regions. The most typical trade would be found at the eastern ends of the anticyclones, because there the trade had the largest component toward the equator, and moreover the reconstruction of the vertical motions around a subtropical cell, suggested by J. Bjerknes, required that the predominantly downward motion be confined to the eastern ends. Toward the west the circulation in the cell required an increasingly great upward component in the middle atmosphere—it was in those regions, empirical investigations showed, that the trade wind was more variable, the trade-wind inversion frequently weak or absent, and the weather worse than the older theorists could allow.

The monsoon, wherever it occurred, in Africa, Southeastern Asia, or Australia, was also an air mass. It was, in fact, a trade that had passed from one hemisphere to the other, and thus, under the influence of a Coriolis force of opposite sign from that in its place of origin, acquired a westerly, instead of an easterly, component. This mass was accelerated under the influence of a solenoidal field that could be attributed ultimately to the different absorptive properties of land and sea, as required by the older theory. It was considered to be unstable to great heights and to have, in the upper layer, a higher specific humidity than the trade because it had been subjected to the destabilizing influences of the equatorial passage for a long time; further, it was moving on to a continent during the warmest season, and hence might be expected to show cold-mass characteristics to a high degree.

What, then, could be said about fronts in the tropics? On the new view, the air masses were tremendous volumes of air, moving away from their source regions and preserving the marks of their origin in their homogeneous properties and of their history in the modifications of their lower layers; ultimately they came into juxtaposition with other air masses along more or less sharp boundaries, the fronts, where the meteorological elements would necessarily show rapid space variations, particularly in the horizontal planes. Were there such boundaries to the trades? This question had already been answered for the Pacific. In 1921 Brooks and Braby [11] had described in the western South Pacific what they called a zone of convergence between the trades of the two hemispheres. The title of their paper, "The Clash of the Trades in the Pacific," is suggestive. In the central Pacific the mean trades from the two hemispheres meet at a small angle along a line that corresponds very well with the mean position of the equatorial low-pressure trough. Farther west, however, the mean trades meet at a larger angle and, in the extreme west, near the Solomons and New Hebrides during the wet season, the boundary is well marked, separating westerly "monsoons" to the north from easterly trades to the south. Rainfall follows the intensity of the convergence[2] as judged from the angle of impact of the currents. In the east, the rain zone is narrow and corresponds in position with the boundary; in the west, the rain zone, while roughly corresponding with the boundary, is broader and, in the mean, the rainfall is heavier. Finally, the tropical cyclones of the South Pacific seem to originate near or at the boundary, to move along it for a time, but in most cases to curve away from it into higher latitudes. Brooks and Braby had done their work and prepared the paper, it seems, before they had heard of the new polar-front theory; before publication, however, they added a footnote, drawing attention to the similarities of the boundary between the trades and the polar front of higher latitudes of which they had just heard a description by V. Bjerknes at a seminar in England. They then pro-

2. In the paper by Brooks and Braby [11], it is clear that "convergence" means convergence of the streamlines, not horizontal velocity convergence.

posed to call the new entity they had introduced into tropical meteorology, *the equatorial front.*

The work of Brooks and Braby seems to have been overlooked by the Norwegians, for they proceeded more or less along the same path without referring to it. However, their researches, in which they used the same method of analyzing surface mean winds and rainfall, were more extensive, ultimately covering all the tropics. They describe the "intertropical front" in the Pacific [10] and Atlantic and Indian Oceans, and in addition attempt to trace its course over the land in Africa and southeastern Asia. Seasonal movements of the front are attributed to different mean temperature conditions in the trades of the winter and of the summer hemispheres.

It is clear that, up to 1933, the new tropical meteorology still followed, at least implicitly, the assumptions of the climatological school. It was implied that the equatorial or intertropical front, being found on maps of mean winds, pressure, and rainfall, would therefore be found also on synoptic maps. As far as I can discover, this assumption was not questioned at any time; indeed, many authors did not recognize that it was an assumption. After 1933, therefore, there was a concerted attempt to apply frontal and air-mass analysis to synoptic maps in the tropics. The movement started slowly, being most active in the Pacific and the Caribbean. By 1939 the standard high-latitude methods of frontal analysis had been adopted by the Philippine Weather Service, by the meteorologists of Pan American Airways [17], in New Zealand [3], Australia [50], and in Martinique and other parts of the Caribbean [31]. World War II brought an immense increase in the amount of forecasting required in the tropics and, with it, a large influx of meteorologists whose training and experience had been largely in high latitudes. The latter applied standard air-mass analysis to their new problems, almost without question; the literature at that time gave them every reason for doing this. The only paper that might have thrown some doubt on the accepted view [25] had at that time attracted little attention. Today, as the survey mentioned in the introduction shows, almost half the meteorologists who have had experience in the tropics still adhere to the tenets of the air-mass and frontal school. We are, therefore, entering the most controversial part of this review and it is well to pause and examine our fundamental conceptions closely. For it is often found that an irreconcilable conflict of scientific views concerns not the facts but the explanations; most of the difficulties may be semantic in origin.

It must be realized that the complex of ideas advanced by Norwegian meteorologists may be divided into two sections. The first section is concerned with certain readily verified facts. The term front, in empirics, is applied to a set of atmospheric phenomena that are frequently discovered by our instruments and by visual observation to be closely associated. The most striking feature is the occurrence of an organized system of clouds, extending over immense horizontal distances; one type of cloud system is called a warm front, another a cold front, and combinations of the two are termed occlusions. However, these terms are only applied if, accompanying the cloud system, there is a narrow zone in which the meteorological elements humidity, temperature, and wind show rapid changes whose sense is defined for each of the three different types of cloud systems. Further, in high latitudes, the pressure gradient should have an oppositely directed component on either side of the system. In this sense an observer on the ground or a pilot in the air may identify a front without ever seeing a synoptic map. It is an observable entity. The Norwegians, however, did not stop at the descriptive stage of meteorology. They had also an explanation, a *theory about fronts.* This theory accounts for the observed fronts as being boundaries, approximating atmospheric discontinuities, between air masses which possess characteristic and more or less homogeneous values of the meteorological elements. Further, the theory requires that the fronts can be maintained for any length of time only if there is a density contrast accompanied by cyclonic shear at the discontinuity. Now this theory is not a priori true. There is no general theory of the atmospheric circulation by which we can derive from the principles of physics the result that fronts must form in this manner. The confidence we have in the theory as an explanation applicable to high latitudes comes not from dynamic meteorology but from experience with synoptic maps, by means of which we check the contention that the fronts observed at a single station or by the airman are in fact boundaries between air masses. This experience leads us to give the name front to regions where certain meteorological elements show steep gradients even in the absence of the typical cloud and weather pattern. This is done because such systems preserve the two most essential features of the theoretical front, viz., the density contrast and the cyclonic wind-shear.

The density contrast and the wind shear are essential to the further development of the theory. For, according to the Bergen school, the frontal surface is not always in stable equilibrium. Waves develop in it and some of these waves are dynamically unstable. Such unstable waves give rise to vortical circulations which roll up the surface and form the cyclones of middle and high latitudes. The kinetic energy of these cyclones is derived from the potential energy of the mass distribution at the undisturbed fronts. It is now obvious why *both* the density contrast and the cyclonic shear are needed at the theoretical front. The first provides the necessary source of energy for the storm, and the second is the destabilizing influence that leads to the occlusion process. They, of course, play other roles in the theory, but this is their part in the wave theory of cyclones. To emphasize the independence of the empirical and theoretical fronts, I should like to remind the reader that line squalls were known long before the frontal theory was promulgated and that very accurate descriptions of cold fronts exist, dating back to the last century [42]. If the theory of extratropical cyclone formation should ever be upset and discredited, cold fronts, warm fronts, and occlusions would continue to be observed and to be identified on synoptic maps,

though the explanation of their relations to one another and to the field of motion would be different.

I have labored the distinction between the empirical and the theoretical front because it was undoubtedly not made by the early workers of the frontal school in tropical meteorology. They succeeded in establishing the following facts by observation and synoptic analysis:

1. In all tropical latitudes, organized lines of cumulonimbus cloud, closely resembling the cold-frontal cloud systems of high latitudes, are encountered by airmen and are observed to pass over fixed observing stations.

2. In the great majority of cases, these lines of cloud are accompanied by zones of rapid wind change. The total shift is not as pronounced as in high latitudes and is generally more gradual. The shear at the shift is often cyclonic. These facts have been emphasized by Deppermann [20], by Harmantas [35], by Frolow [31], and by Alpert [1]. From personal experience I can assure the reader that they are facts.

3. Differences of temperature can sometimes be found between the air on either side of the cloud line. This is reported by Sawyer [60] and by Solot [67]. However, in certain oceanic regions, though there is often an abrupt drop in temperature at the surface with the passage of the cloud line, it may recover its former value when the weather clears.

4. There may be very marked differences in specific humidity across the cloud line—not, it is true, at the surface, but in the middle atmosphere.

5. The lines are frequently but not always associated with a low-pressure center or trough at the surface, but owing to the weak pressure gradients and the great amplitude of the semidiurnal pressure oscillation, it is difficult to correlate any sharp change in the barogram with the passage of the cloud line, unless a violent thunderstorm occurs at the station.

6. One can sometimes find these cloud and wind-shift lines day after day in the region of lighter winds around and outside typhoons, hurricanes, and tropical cyclones [21]. Moreover, the storms sometimes appear to originate in the regions where such lines have previously been discovered.

To most tropical meteorologists, after 1933, these discoveries were sufficient to identify the lines as fronts, and since they made no distinction between the empirical and the theoretical front, they adopted the entire high-latitude apparatus of frontal and air-mass analysis and forecasting. The way had been prepared, as we have seen, by Brooks and Braby, by Bergeron and, most authoritatively, by the authors of *Physikalische Hydrodynamik* [10]. The frontal wave theory of storm formation was adopted without question; indeed it seemed to be confirmed by (5) and (6) above. A few meteorologists were concerned by the frequent lack of temperature differences across the fronts and were careful to distinguish "tropical fronts" where the differences were very small (in many cases they were, in fact, absent) from extratropical fronts where there was usually no difficulty in detecting the difference, at least by radiosonde. Deppermann in particular was always careful to make the distinction. Having decided that fronts and air masses existed in the tropics, the frontal analysts of the decade 1933–1943 were concerned to describe and classify the major discontinuities of the tropics. First, and most important, came the equatorial or intertropical front, whose synoptic existence was never doubted during this period. But there were other major systems, particularly in the western Pacific. Thus Deppermann distinguished a "tropical" front between the winter monsoon of eastern Asia and the trade of the North Pacific [20]. Other "major fronts" were described in different parts of the tropics [17] but these are now of only historical interest. In addition, much attention was given to the passage of polar fronts from the belt of the westerlies on into tropical latitudes.

Some extraordinarily detailed frontal analyses of hurricanes and tropical storms were produced during the thirties. In the South Pacific the full occlusion process leading to the development of tropical cyclones was illustrated by Kidson and Holmboe [38]. Similar analyses were published for storms in the Caribbean [37, 64] and in the Bay of Bengal [59]. An elaborate triple-point theory due originally to Scherhag was applied to North Pacific storms [20]. If one were to judge by the literature up to 1940, one could not doubt that the application of the frontal and air-mass theory to the tropics was completely justified. The chief problem seemed to be, not to account for the origin and development of tropical storms, but to explain why many more than were actually observed did not develop. The tentative theory advanced in *Physikalische Hydrodynamik*, accounting for the maintenance of the equatorial front in terms of temperature differences between the trades of the winter and the summer hemispheres, was extended to account for this also. The equatorial front, on this view, tended to become a sharp discontinuity, "to become active" in the words of the air-mass school, after the air in a great polar outbreak in the winter hemisphere reached the equatorial regions. This would communicate an impulse of an unspecified kind to the equatorial front which would thereupon become unstable and form a tropical cyclone. As early as 1935 De Monts had published a paper with the significant title "Rôle catalyseur de l'air polaire dans la génèse d'un cyclone tropical" [19] describing this process in the South Indian Ocean. Japanese meteorologists seemed particularly attached to this explanation [2]. Later Grimes [33] endeavored to specify the nature of the impulse that would be given to the equatorial front by the arrival of winter-hemisphere air after a polar outbreak. He considered that a "cyclone only forms when there is a surge across the equator of air which in the hemisphere of origin had sufficient anticyclonic vorticity, which is conserved as cyclonic vorticity in the other hemisphere."

Since shearing motion at the frontal surface would be the only destabilizing agent leading to waves of increasing amplitude, Grimes was evidently looking for some process which would bring about a sudden increase in cyclonic shear at the supposed frontal surface. The process was made to depend upon the conservation of the vertical component of absolute vorticity, which it is

now fashionable to suppose actually occurs in the atmosphere. In "The Movement of Air Across the Equator" [33], he enunciates this principle and applies it, under very restrictive special assumptions as to the nature of the motion. In brief, the motion must be horizontal, nonviscous, and autobarotropic, with a constant meridional component. Although as far as I know no investigation has been undertaken to find out whether vertical motion, frictional forces, solenoids, and variations in the meridional component of the wind can be neglected in the tropics, the results of this paper are frequently quoted without mention of the assumptions and applied as if they were perfectly general. Moreover, the efficacy of the deduction from the principle of the conservation of vorticity, as assigning a cause for cyclone formation in this context, depends upon the validity of the frontal theory in the tropics. If that theory is found to be inapplicable, the principle of the conservation of vorticity would have to be invoked in some entirely different context, with results that cannot at present be anticipated.

Although there are still a large number of meteorologists who adhere to the principles outlined above, with greater or lesser modification, the synoptic experiences of World War II, especially in the Pacific and the Caribbean, gave a death blow to the air-mass theory in its extreme form. This theory failed to pass the test to which all theoretical work in meteorology must finally come (though the evil day may be put off for almost a decade); it was found almost useless as a guide to short-period forecasting in low latitudes. The writer can vouch for this through personal experience, through conversation with many tropical meteorologists in the Pacific, the Caribbean, and the United States (where people with experience in all tropical areas are now to be found), and through the answers to the questionnaire issued by the Institute of Geophysics. But the reader need not depend on these sources of information; the recent literature reflects the disillusionment of a whole generation of tropical meteorologists. This is shown in attempts to doctor and patch the frontal theory in such a way as to take into account the many anomalous movements and the sudden appearances and disappearances of the fronts, especially of the equatorial front, or in frank abandonment of the air-mass theory altogether. The following modifications of the frontal theory, as previously outlined, have been proposed:

1. The polar front enters the tropics, not as a typical surface front but as a front aloft, in the westerlies. It persists in that situation but may come down to the surface in lower tropical latitudes and at a later time.[3] Forsdyke [29, p. 82] says,

> Waves in the circumpolar westerlies occur at the surface on the poleward sides of the subtropical high-pressure belts, and at higher levels they extend to low latitudes where the westerlies overlie the trades. It is tempting to identify these with the upper parts of cold fronts associated with the depressions of middle latitudes. At Mauritius they are frequently associated with nimbostratus cloud and rain which spread from the southwest above the lower current from the east-south-east. They should be indicated on the charts as upper fronts. It is possible, therefore, to have in the same part of the chart a lower front moving from an easterly, and an upper front moving from a westerly point, and crossed lower and upper fronts may occur.

2. On entering the tropics, the polar front separates into two portions: (a) the density discontinuity or shear line, and (b) the polar trough (see [16, 53, 65]).

3. The polar front is destroyed soon after entering the tropics. The very rational Memo 131/44 of the British Naval Meteorological Branch [46, p. 28], says,

> So far we have considered only discontinuities which *originate* in the tropics, notably the I.T.F. What happens to fronts which move into the tropics from higher latitudes? Generally speaking, they lose their well-defined slope and become diffuse: their movement becomes slow and indefinite and by the time they get near the equator, if they get that far, they bear little resemblance to their text-book prototypes. Most of them ultimately lose their identity in the equatorial low pressure trough which is a great leveller of air mass disparities. A good many of them get no further than the subsidence zone of the sub-tropical highs where they "dry-out."

4. There is little or no temperature contrast at the equatorial front which is, indeed, simply a more or less narrow zone at which the trades of the two hemispheres converge and rise. This represents a return to the original position of Brooks and Braby, and the climatological school. Brooks and Braby were careful to point out that no temperature contrast could be found across the equatorial front in the regions where, judged by the mean winds and precipitation, it was strongest. Numerous quotations are available on this point [15, Chap. 10; 46, 65]. Garbell [32, p. 70] says,

> The discontinuity between the two tropical air masses coming from the two hemispheres usually consists not so much in a difference in surface temperature as in a difference mainly in the vertical distributions of temperature, moisture, stability and wind velocity between the two easterly air streams.

5. The equatorial front is single and continuous [14].

6. The equatorial front is double [28, 72].

7. The equatorial front is part single, part double [32].

8. The equatorial front is discontinuous. There are many equatorial fronts [54].

9. The equatorial front may slope toward the summer hemisphere [13].

10. The equatorial front may slope one way in the lower atmosphere and the opposite way in the upper atmosphere [34].

11. The equatorial front does not slope at all [46].

12. The equatorial front moves discontinuously. It may jump from one position to another without passing through intermediate points [29].

13. The equatorial front may move by disappearing in one position while another forms in a different position [17].

The list of suggested modifications to the frontal

3. See Rep. No. 600-50, Publications of the Weather Division, USAAF, p. 15.

theory in the tropics could be extended almost indefinitely. To anyone, like the writer, who has had to review the literature on tropical fronts published in the last ten years, the total effect is one of intolerable confusion. And this confusion exists at the present time, unresolved. Where possible, I have purposely chosen the latest quotations for inclusion in the above list, which is, of course, very incomplete. Thus Garbell's textbook was published in 1947; Forsdyke's summary of tropical meteorology in 1949. Both these works will give the reader some idea of the complexities introduced by members of the air-mass school in an endeavor to remove the difficulties associated with the attempt to forecast tropical weather by high-latitude techniques. Forsdyke, indeed, frankly admits in places the impossibility of using standard methods. On page 37 of his paper he says,

> The pressure distribution is of little or no assistance in forecasting the movement of the intertropical convergence zone. The streamline chart is the best aid, the frontal zone usually being located at the line of separation between winds having northerly and southerly components respectively. Over wide areas, however, the surface winds are often the only winds available; hence it is usually the surface front which is so placed, whereas observation shows that the worst weather often occurs up to 200 miles away from the surface front. It may be assumed that the front will move toward the side on which the winds are weaker, but it is difficult to estimate the speed of motion from the observed winds. Often the front appears to jump.

It is easy to be wise after the event, but one may legitimately speculate on what the history of tropical meteorology would have been, during the last fifteen years, if the distinction between the observed wind-shift line accompanied by cumulonimbus, the old "line squall," and the theoretical front of the Bergen group had been kept in mind. Would not the early workers, realizing that the empirical front of the tropics was not necessarily a discontinuity in the density field, have also realized that frontal dynamics would not be applicable to it, that new synoptic entities were being investigated and that therefore new explanations were required? That is the position in which we now find ourselves. At a time when the tropical observing networks are beginning to deteriorate, when we have fewer and less accurate observations than at any time since 1939, we are realizing that tropical meteorology, in investigating these entities, has a contribution to make to general meteorology. Nowhere in the atmosphere are we better able to study the direct effect of horizontal velocity divergence in the lower atmosphere, free from other effects, than in the tropics. The tropical "fronts," which have been dismissed as particular, well-understood examples of a common atmospheric discontinuity, are now seen to be completely without dynamical explanation. When the explanation is forthcoming, it cannot but react on the theory of high-latitude fronts.

THE PERTURBATION METHOD

One of the most serious objections to either the doldrum or the frontal explanation of the origin of tropical storms is that a proportion of the hurricanes that affect the West Indies and the southern United States originate within a comparatively strong east current of great depth prevailing during the wet season over the Caribbean. The storms deepen rapidly and are often of great violence. They are usually of smaller diameter than the hurricanes that move in from the east and which clearly originate in the doldrum region of the eastern Atlantic. In 1940, Gordon E. Dunn published a paper dealing with the origin of these storms [25] and the ideas expressed in it have led to the development of a new outlook on tropical problems.

Dunn, using 24-hr isallobars to eliminate the great semidiurnal oscillations of pressure, showed that:

1. The deep easterly belt of the Caribbean in the wet season is subject to pressure waves of small amplitude, which move from east to west.
2. Ahead of the wave the trade-wind inversion is relatively low, and the weather is fine, with small clouds and little precipitation.
3. Behind the trough the trade inversion is high or absent and the region is the seat of many shower-clouds. He speaks of this as the "rolling away" of the inversion.
4. Some of the waves, and a proportion that increases as the season advances, grow in amplitude and give rise to small but violent hurricanes.
5. The isallobaric highs and lows associated with the waves, whether hurricanes are formed or not, follow tracks that correspond with the mean hurricane tracks of the appropriate month in the season.

Subsequent writers elaborated this picture of the easterly wave. The most active worker was Riehl [53] who investigated the vertical structure of the wave more closely and showed that it was easier to detect in the wind field than in the pressure field and aloft than near the surface. Later the easterly wave analysis was applied to the North Pacific and the North Atlantic. More recently easterly waves have been identified in the Indian Ocean [29] and in West Africa [63].

In opposition to earlier writers on the Caribbean, such as Kidson [37] and Scofield [64], Dunn insisted that no temperature discontinuity could be found either in the easterly wave or in the developing hurricane. Later research has confirmed his contention and there seems now to be no dispute on this point as far as the West Indian hurricanes are concerned. Radar investigation of the storms as they pass over the southern seaboard of the United States, however, seems to show that there are well-marked cloud and precipitation discontinuities in the mature storm [73]. As this topic will be discussed in the next section, I wish only to point out here that Dunn was explicit only on the absence of *temperature* discontinuities. He admits, however, that the southwest wind at Panama in the wet season may be colder in the lower layers than the easterly on the other side of the equatorial convergence zone. One gets the impression that he is willing to admit a frontal origin

for those storms that clearly originate in the doldrums but that he denies such an origin in cases where the storm can be traced to an easterly wave. Temperature differences have been reported at the equatorial front in other parts of the world, for example in India [60] and in equatorial Africa [67]; this probably predisposed Dunn to concede that a frontal origin of doldrum hurricanes was possible.

This concession, either in the form of an outright admission of a frontal origin of *some* tropical storms, or the lesser concession that storms that originate in the neighborhood of the equatorial low-pressure trough differ fundamentally from those that can be traced back to easterly waves, has been a characteristic of workers of the perturbation school. It has, in fact, done much to prevent their presenting a unified theory of the origin of tropical storms. Apart from this admission, however, the workers as a group have been rather extreme in denying the existence of fronts in the tropics. They have, rightly I think, insisted on the absence or irrelevance of temperature contrasts in the Caribbean and the western Pacific; but their zeal has also led them to minimize the well-authenticated accounts of cloud systems and wind-shift lines near the equator, lines that look like the cold fronts of high latitudes. As I have said, they may admit that such lines can exist in the equatorial low-pressure trough but deny their presence in the easterly wave in the Caribbean. Yet nothing is easier, with good aircraft reports, than to demonstrate the presence of such lines in certain Caribbean easterly waves. To admit this, perhaps, would seem to them to destroy the value of the contributions they have made; we would be back in the confusion that prevailed during the ascendancy of the frontal theory. The resolution of their difficulty must be left to the next section.

The detailed structure of the easterly wave was worked out by Riehl [53]. He assumes that the basic undisturbed current in the Caribbean is a deep easterly. If this assumption is conceded, it is a simple matter to define and to detect the chief axes of the easterly wave, both at the surface and aloft, for the axes, horizontal and vertical, along which the wind is easterly will be the "trough" and "wedge" axes. In the Northern Hemisphere we shall find northeast winds ahead of the trough and southeast winds behind it. By the analysis of upper-level wind maps, supplemented by time cross-sections at individual stations, it is then possible to discuss the slope of the systems. Riehl showed that the perturbations of the wind field were of greater amplitude than those in the pressure field, and that the amplitude of the former also seemed greater in the middle troposphere than at low levels. Generally, the trough axis slopes toward the east with height. Further, the lower wind fields are positively divergent[4] ahead of the trough and negatively divergent behind it; with this distribution of the horizontal velocity divergence, and (to a high degree of approximation) of the mass divergence, the westward passage of the wave is explained, as is also the "rolling up" of the trade inversion behind the wave axis and the consequent distribution of cloud and precipitation in the perturbation. The virtue of Riehl's work is that he establishes the correlations of the fields of motion, pressure, composition, and temperature to form a simple integrated picture of the perturbation; this has hardly been achieved for any perturbation in high latitudes except, perhaps, the frontal wave.

Riehl also discusses another type of perturbation observed in the Caribbean area, one we have mentioned before. During the winter in that region, the east wind near and at the surface is frequently accompanied aloft by winds with a westerly component; in other words there may be a typical trade-antitrade regime. In the lower easterlies, perturbations that look superficially like easterly waves on the low-level synoptic map appear; however, these disturbances move, not toward the west, but toward the east, against the lower easterly current. It is always possible to show that the trough in the easterlies is continuous with a surface trough in the westerlies of the higher latitudes and with an upper-level trough in the westerlies of the tropics. This trough Riehl called the *polar trough*; he accounted for the eastward movement of the system in the lower tropical atmosphere as being a reflection of the movement of the trough aloft in the circumpolar westerlies. This explanation had been advanced before, of course, but in a different form. The older explanations were in terms of an upper front in the westerlies—Riehl simply omitted the front, thus effecting a great simplification in this part of the model. However, it is a fact that true surface fronts move southward over the United States in wintertime and that they may move off the continent into the Gulf of Mexico and the Caribbean. The surface front becomes weak, but can often be traced far to the south into Central America. It is then found that it is dissociated from the polar trough (especially in the pressure field) and that the chief cloudiness and precipitation accompany the eastern part of that trough rather than the surface front. This picture has been elaborated by Cressman [16], Simpson, and others. As time goes on more and more complexities are introduced into it; one suspects that all is not well with the polar-trough model and that several different synoptic phenomena are being confused under one heading. At present, however, little research is being done on winter disturbances in the tropics and the time seems to be ripe for new investigations. In the meantime, if current synoptic maps from weather stations in the Pacific are a guide, there seem to be two rival explanations of winter disturbances in the higher tropical latitudes: the polar trough explanation and the older one which accounts for all winter precipitation in the trades in terms of classical fronts traveling from higher latitudes toward the equator.

In 1948 Freeman [30], on the basis of wartime experiences in the New Guinea area, introduced a new

4. Riehl is careful to show that the total horizontal velocity divergence, not merely the streamline divergence, varies in this manner. The distinction was never clear to workers of the air-mass school.

type of perturbation into equatorial meteorology. The easterly current near the equator was described by him as being on occasion the seat of a disturbance in the *speed* field in the lower atmosphere. The streamlines are not disturbed from their easterly direction, but maxima and minima of speed travel downstream in the current. Freeman places great emphasis on a "jump" in speed corresponding to the passage of the trough in the normal type of wave. However, it is probable that this is merely an extreme form of a longitudinal wave common in the New Guinea area and known to forecasters there since the elaboration of the pilot-balloon network in 1942 and 1943. For convenience in this discussion, I shall call all easterly waves that are detectable in the speed field but not in the streamline field, *Freeman waves*, without necessarily agreeing with the dynamical explanations advanced by Freeman. As far as I know this type of wave is found only in the New Guinea area, and there is reason to suspect that the peculiar orography and geographical position of that island play an important part in its genesis. However that may be, we may fairly add a new type of disturbance, independent of fronts or convergence zones, to those already known to affect the homogeneous air streams of low latitudes.

It was one of the great advantages of the air-mass theory that it seemed to give a complete or almost complete explanation of a multitude of different weather phenomena in both high and low latitudes. For low latitudes, the climatological school also appeared to have a complete system of explanation. This is not the case, at least explicitly, with the newer concepts advanced by the perturbation school. The ideas are new, the whole movement being less than ten years old. As a consequence there has been a tendency to emphasize methods of analysis, especially synoptic analysis, rather than to attempt complete theories. Further, the workers on perturbations of homogeneous streams have rightly adopted explanations from earlier meteorologists, where these have seemed to them to fit the facts, even though at the time it seemed difficult to reconcile those explanations with the newer outlook. We have already mentioned the fact that the equatorial front was accepted by the perturbation school; it was, of course, regarded as a zone of intense horizontal velocity convergence in the lower atmosphere, and its role as an atmospheric discontinuity was minimized. Nevertheless the newer analyses in the vicinity of the equator were almost indistinguishable in practice from those of the more enlightened members of the air-mass school, such as Deppermann. Similarly, in their treatment of the monsoons of Asia, Africa, and Australia, the group adopted almost without question the explanations of the climatological school. Thus, although no unified theory has yet been presented, we may reconstruct such a theory from these adaptations, from hints in the literature, and from the evident dependence of the group upon the dynamical conceptions of Rossby and his coworkers in Chicago.

In this reconstruction, the trade winds, considered as broad streams of air largely homogeneous in the lower horizontal planes, play a major role. Less emphasis than formerly is placed on the meridional components of these winds. They are, indeed, regarded as part of the circumpolar vortex, which happens to be retarded with respect to the earth's surface in low latitudes but which moves faster than the earth in higher latitudes and in the upper air over the tropical zone. The base currents of the general circulation, on this view, are primarily zonal. Meridional circulations arise as perturbations on these currents; the only *synoptic* problem in tropical regions, with the sole exception of the study of the equatorial front, is to describe and explain the perturbation of the zonal currents in the tropics. After description, the perturbations are classified as we have seen; there are (1) the major "dynamic" perturbations of the zonal motion, the subtropical anticyclones and the semipermanent lows of high latitudes; (2) the great seasonal perturbations of the trades, the monsoons, due to thermal causes, as understood by the climatological school; (3) the perturbations of the trades themselves—in winter the polar troughs, in summer the easterly waves and Freeman waves; and (4) the minor diurnal and orographic perturbations, for which an explanation has already been advanced. Hurricanes and tropical storms arise in two ways: (1) as a result of dynamic instability in the trade current, leading to the growth in amplitude of the easterly wave and its transformation into a vortex, and (2) as a result of some unknown process affecting the equatorial front (or, if one wishes to be free of all frontal taint, the intertropical convergence zone). The causes of the instability are unknown, though Riehl has advanced two accounts of the sufficient conditions [54, 55]; of conditions both necessary and sufficient there has been so far no hint in the literature. Whatever may be the difficulties of explaining the origin of hurricanes, typhoons, and other tropical storms, explanation of the major features of the perturbations (again, with the exception of the equatorial front) is achieved by the methods of dynamic meteorology as set forth by Rossby. The central notion of these methods is that the vertical component of vorticity is conserved during the isentropic motion of the air particles in the trade-wind zones. However, in the case of the easterly wave, the horizontal velocity divergence of the lower parts of the trade must be taken into account, and this makes the so-called vorticity equation so intractable mathematically that no quantitative evaluation of the wave characteristics can be carried out—the explanation must therefore follow the original qualitative discussion of Bjerknes [7], *i.e.*, in terms of a latitude and a curvature effect. Greater precision was given to this latter explanation by Bjerknes and Holmboe [8], but unfortunately the theory as handled by them leads to the result that all easterly waves of the type described by Riehl must deepen and give rise to vortices.[5] Freeman, moreover, has advanced an entirely

5. Professor J. Bjerknes has directed my attention to the possibility of effecting a reconciliation between the theoretical results and the Riehl model by taking into account convergence due to friction in the lowest levels of the divergent part of the wave. He thinks that this might bring about a balance in the integrated mass divergence over the region and thus stabilize the wave.

different theory for the origin of the waves he describes; so far no one has attempted to reconcile the two dynamic explanations.

The confusion that results from attempts to explain the dynamics of the perturbations in the tropics is probably temporary; at all events it is negligible compared with the greater confusion of ideas concerning the general circulation in low latitudes. Here the workers have followed Rossby with implicit confidence, and it is in his writings that we find the most complete expression of their views. We shall therefore examine in some detail those parts of his papers that relate to the tropics.

In 1941 Rossby published "The Scientific Basis of Modern Meteorology" [56], summarizing his views on the dynamics of the general circulation; presumably he held the same views in 1945, since the same paper was republished in the *Handbook of Meteorology* [57]. In this paper Rossby gives a masterly exposition of the climatological explanation of the tropical circulation. It is in fact Hadley's explanation in modern dress. First, we have the direct circulation cell, on a nonrotating earth. As the result of the surplus of incoming over outgoing radiation in the lowest layers of the atmosphere near the equator, the temperature in very low latitudes tends to increase, with consequent expansion of equatorial air columns; the contrasting radiative regime in higher latitudes produces a tendency to contraction of the vertical air columns. "Thus," says Rossby, "at a fixed level of, say, 5 km (3 miles) above sea level, a greater portion of the total atmospheric air column would be found overhead near the equator than near the poles." A direct circulation converting potential into kinetic energy would result; the air motion would have a component toward the equator in the lower layers and away from the equator aloft. On a rotating earth, zonal components would be introduced. The motion, the same on every longitude, would be subject to the conservation of angular momentum. "A ring of air extending around the earth at the equator at rest relative to the earth, spins around the polar axis with a speed equal to that of the earth itself at the equator. If somehow this ring is pushed northward over the surface of the earth, its radius is correspondingly reduced; and it follows from the principle set forth that the absolute speed of the ring from west to east increases." Here we have the origin of the antitrades. The trades, in like manner, acquire their easterly component as a result of the conservation of angular momentum. The final result is that though east winds prevail between 30°N and 30°S in the lower atmosphere, "Above 4 to 5 km (2½ to 3 miles), westerly winds prevail in all latitudes." Thus the observed features of the mean circulation in the tropics, after the abstraction of diurnal, seasonal, and orographical perturbations, is completely explained.

In 1947 Rossby completely rejected this model of the tropical circulation, for the following reasons [58]:

1. He quotes Vuorela [71], Kuhlbrodt [39], and the wartime data from the Marianas and Marshalls as showing that the easterlies near the equator extend farther upward and poleward than would be compatible with the direct-cell theory. He says, ". . . it is fairly clear that the simple thermal circulation outlined by Hadley cannot be the sole, or even the dominating, feature in the mechanism of the trade winds."

2. He then goes on to point out the great uniformity of the meteorological elements, especially temperature, in the tropics, and the difficulty of reconciling these observations with the simple thermal explanation.

3. He next mentions the "discovery" of Fletcher [28] that in some longitudes there is a mean west, not east, wind near the equator, and this in oceanic regions. Defant also is mentioned in this connection. Further, the well-known zone of divergence in the equatorial central Pacific is cited in support of the thesis that the direct-cell theory is inadequate.

4. "The remarkably small annual march of the doldrum belt" seems to Rossby to contradict the direct-cell theory.

Each of these points has long been known to tropical meteorologists. For example, both van Bemmelen [70] and Kuhlbrodt [39] have emphasized point (1) and it was thoroughly established by 1939 in the Pacific as a result of pilot-balloon observations; point (2) is as old as the earliest climatological maps (see also [61]); point (3) has been discussed by Deppermann [23]; and, further, the explanation of persistent west winds at the equator was first advanced by Piddington and Reid [52] one hundred years ago. The dry zone in the central Pacific has been discussed by Schott [62] and the New Zealand meteorologists, particularly Seelye. All these points, and others I have already mentioned, long ago led many tropical meteorologists to reject the extreme climatological view and some to embrace the frontal and air-mass theory as applied to the tropics.

After summing up these objections, Rossby generalizes from them, implying that because the direct-cell theory fails to apply over certain longitudes it fails over all. But this is to commit the typical fallacy of the climatological school: what I may call fallacious generalization from a given longitude. It arises when we become acquainted with new observations restricted to a narrow longitudinal band. Thus, when the older climatologists first became acquainted with the very persistent trade winds typical of the eastern border of the subtropical anticyclones, they generalized to form a model in which, along every longitude, these winds and the associated antitrades prevailed. We have recently become more aware of the mean state at the western ends of the subtropical highs, where "double equatorial fronts" and west winds on the equator are common; there is therefore a tendency to generalize from these longitudes to the whole tropical belt. The same remarks apply to the temperature distribution: at the eastern ends of the anticyclones, horizontal temperature gradients may be found both on the synoptic and on the mean maps, aloft and at the surface [18, 61]. The older climatologists could therefore justify their model of the trades by appealing to observations from these longitudes. However, if we concentrate our attention on the western parts of the tropical oceans, particularly of the Pacific, and generalize to all longi-

tudes, we become convinced that there are no horizontal temperature gradients in latitudes below 20°.

The new model, then, is open to the same empirical objections as the old. But there are also serious objections to the theoretical explanations which have been advanced, inasmuch as the principle of the conservation of the vertical component of the absolute vorticity is appealed to. Without entering into the details of the new theory, which would take us beyond the limits of this article, we may sum up these objections by saying that it has yet to be shown that the mean motion of the tropical atmosphere is autobarotropic, horizontal, frictionless, and nondivergent to a high degree of approximation. Until this has been shown, we must suspend judgement as to the applicability of Rossby's concepts in low latitudes. Even if we admit them, it will be only in certain longitudes, and not as a general explanation of the tropical part of the planetary circulation.

It is clear now that the great success of the perturbation school has been in the realm of descriptive synoptic meteorology; the models of the perturbations which they have promulgated have been applied in the field with far greater success than was ever achieved by the air-mass school. In dynamical meteorology and in their treatment of the general circulation, members of this school have been less successful. There have been complaints, moreover, that the model of the easterly wave published by Riehl is far too rigid. In particular, the slope of the actual systems is frequently different from what might be expected from the published accounts and, what is more important, the weather distribution about the wave axes departs widely from the model [26].

But these, as we shall see, are not serious objections; they can be removed by a more careful study of the perturbations in low latitudes and by avoiding the tendency to apply models to synoptic maps as if they were rubber stamps. I ought perhaps to point out that my sympathies lie with the perturbation school and that if the criticisms here seem to be harsh, it is because I believe that future advances in tropical meteorology will come through the application of the concepts and methods of this school, provided they are freed from fallacious generalizations.

THE PRESENT AND THE FUTURE

General Observations. Each of the methods of approach to the tropical problems which were described in the preceding sections has something to recommend it; each has contributed something to our empirical knowledge and our understanding. As far as we can see at present, progress will come by pursuing to further lengths, with new and more detailed data, the lines of investigation that have already proved profitable. It is unlikely, that is, that we shall entertain any radically new theory of tropical meteorology in the near future. As a guide to the assessment of the present state and possible future development of the science we ought therefore to extract from the confusion of present conflicts those concepts upon which the majority of tropical meteorologists would agree. Before we do this we ought, however, to make a few general observations on method, deriving our principles from the study, not only of the achievements of the three schools, but of the errors they have committed.

1. Before attempting causal analyses of tropical phenomena, or dynamic explanations, we ought to be sure that our empirical descriptions are correct. This principle was violated by the climatological school in describing the doldrums, particularly the doldrums in the Pacific; by the air-mass school in describing fronts, since they neglected to describe accurately the temperature and wind changes observed at the supposed discontinuities; and by the perturbation school in omitting all mention of the convergence lines of the frontal school in their description of the model of the easterly wave.

2. The fallacy of generalization from a given longitude should be avoided as far as possible. We ought always to be on guard against it, particularly when dealing with the mean circulation in the tropics. But faulty generalization also enters in other contexts, especially when we have theoretical preconceptions. Thus the frontal school, as soon as it found evidence of empirical "fronts" in the neighborhood of the equatorial low-pressure trough, assumed that the occlusion process would also occur. No evidence was, so far as I know, ever produced to back this up. We ought then to remember at all times that, owing to the fact that we never have as much data as we really need, the habit of hasty generalization is an occupational disease of meteorologists. All the geophysical sciences suffer from this handicap; we ought to be at least as careful as the pure physicists, and probably more careful.

3. We ought to guard against the statistical fallacy in meteorology. This consists in assuming that a mean motion of the atmosphere, arrived at by the standard methods of climatology, represents a physically possible steady-state motion.

Achievements and Future of the Climatological School. There is no doubt that the departures from the mean values of the meteorological elements that are attributable to the diurnal cycle of radiation and to orographical effects can be of great magnitude in the tropics. It is the virtue of the climatological school that they have always emphasized these variations. However, in recent years little research work that might lead to a better understanding of the phenomena has been carried out. It is encouraging to notice that Leopold [41] has entered this field with papers on the so-called sea-breeze fronts in the Hawaiian Islands. Wartime synoptic experience everywhere in the tropics has confirmed the existence of organized lines of cumulus or cumulonimbus, accompanied by wind shifts, which simulate the cold fronts of high latitudes but which are demonstrably due to diurnal and orographical causes. The dynamical problems raised by these "fronts" are among the most interesting problems of tropical meteorology. Their properties, however, can be investigated only on location; it is to be hoped that young meteorologists resident in tropical areas of high relief will, in the future, undertake the investigation. Apart from the theoretical interest of such "fronts,"

the greater precision that would be given to local forecasting in the tropics by a complete understanding and description of the diurnal variation of orographical organized cumulonimbus well merits military and commercial support of such investigations.

It is difficult to deny that, as far as the most general features of the monsoons are concerned, the climatological school seemed to have analyzed the mean data and to have correctly assigned astronomical and orographical causes to the variations. But it is probable that the monsoons are far more complex in their details than that school supposed. There are several reasons for this conjecture. First, the so-called monsoon lows over the land, whatever may be their properties in the mean, on synoptic maps show day-to-day variations which could be attributed to the westward movement of individual cyclonic circulations through the area. Moreover, where the conformations of the continents are favorable, it can be observed that these minor cyclones come in from the sea, fully formed. They cannot under any circumstances be regarded as "heat lows." These local circulations are known in India, in Australia, in Indo-China and China, and in East Africa. Second, and this is probably correlated with the facts first mentioned, the wartime synoptic observations confirmed what was after all known locally for many years: that the weather in the Asiatic monsoons is far more variable than popular and textbook accounts would lead one to suppose. It does not rain continuously for three or four months in the wet season. There are frequent spells, lasting sometimes for days, during which there is little precipitation and sometimes little cloudiness. Third, there are, on certain longitudes, quasi-permanent low-pressure areas which migrate to and away from the equator, lagging behind the sun in the same manner as the monsoon lows, and having, on their equatorward borders, belts of winds with a westerly component; they are situated, however, not over the land but over the ocean. There is a very persistent mean low of this type which migrates from eastern New Guinea to the eastern borders of the Philippines and back. There is another west of Mexico and Central America. If these lows had been situated over a land mass, we would have no hesitation in calling them monsoon lows, and in accounting for their movements and the distribution of the meteorological elements about them in the usual manner of the climatological school. We ought, perhaps, to follow up these climatological hints and, combining them with an intense synoptic and dynamic research into subequatorial semipermanent lows in general, improve our understanding and prediction of monsoon weather. Largely because of historical accidents, research has been concentrated in the trade-wind zones of the Pacific and Atlantic. The time is now ripe for a more intensive investigation of those parts of the tropics that show large seasonal variations in the meteorological elements.

We should, however, not neglect the oceanic regions. Particularly we should not completely reject the classical climatological explanation of the trades, as Rossby has done. It is possible that that explanation is still valid for the longitudes where the trade is really a trade, *i.e.*, where it shows, in the lower level, great persistence, a large meridional wind component, a comparatively strong meridional temperature gradient and, aloft, the typical antitrade, evidence of persistent and marked subsidence, with a low and strong trade-wind inversion. These conditions obtain, as has been pointed out, in the eastern and southeastern branches of the great quasi-permanent anticyclones. Downstream, toward the west, however, factors other than those invoked by the climatological school may dominate the motion. Here, Rossby's new explanations may hold.

Above all, our survey shows, the greatest need is for more complete data, especially from the upper air. At present the distribution of observing stations in all latitudes is largely determined not on scientific, but on economic and military grounds. This, it is to be presumed, will not be changed in the foreseeable future. Nevertheless, it seems to be accepted that improvement in practical forecasting for periods longer than forty-eight hours depends on improvement in our empirical and theoretical knowledge of the mean circulation of the earth's atmosphere. How this knowledge can be built upon an observing network that is half-way adequate only in North America and in Europe is something that, to my knowledge, has never been explained. In particular, if scientific meteorologists are content to advance models of the general circulation based upon the North American atmospheric section and theoretical explanations of this model based upon an obsession with the circular vortex, ignoring the tremendous gaps in our knowledge of the meteorology of the tropics (over half the troposphere), of the high latitudes in the Southern Hemisphere, of Asia, even of Central America—then they cannot blame those that hold the purse strings for supposing that all is well with the present setup. The fact that acquaintance with new data is enough to cause a complete *volte-face* in the theoretical work of an outstanding leader in dynamic meteorology, at least so far as the tropics is concerned, should show that the filling of the gaps mentioned above is an absolutely necessary (although not sufficient) condition of the solution of the long-range forecasting problem.

Achievements and Future of the Air-Mass School. The achievements of the air-mass school were two in number. First, they drew attention to the great horizontal homogeneity of the easterly and the westerly air streams found in tropical latitudes, described the sources and modifications of these masses, and accounted for them with varying degrees of plausibility. The work on the trades still goes on, at all events. There has recently been a study of the east wind in the Caribbean by Haurwitz and his collaborators that is a model for all future work of this type [12]. Similar work in the far eastern equatorial Pacific, based on the Panama Canal Zone, would be very welcome, particularly in giving us precise knowledge of the structure of the west winds in that area during the summer. Later, the researches could profitably be extended to the Far East.

The second achievement of the air-mass school is that the earlier frontal workers discovered in the tropics, and

described, the organized lines of cumulonimbus cloud, accompanied by appropriate wind changes, which they identified as fronts. The existence of the lines has been confirmed, but it now appears that the frontal explanation must be rejected. The following empirical facts may be regarded as established:

1. If an adequate streamline analysis of the lower wind field in which the cumulonimbus line is embedded is carried out, using the isogon method of V. Bjerknes and his collaborators [9], the line of cumulonimbus coincides in position with an asymptote of convergence *in the streamline field*. But not all asymptotes in the field are accompanied by a "front"; there is an organized cloud system only if, at the same time, the total horizontal velocity divergence is negative in the neighborhood of the line. Aloft, say at 20,000 ft, the cloud line may correspond to an asymptote of positive divergence in the horizontal wind field. We shall henceforth call organized lines of cumulonimbus (or tall cumulus, on occasion) *convergence lines*, not fronts. We ought to point out that regions in which there is negative horizontal velocity divergence in the lowest layers are not always the seat of lines of convergence; these occur only if there is also convergence in the streamline along an asymptote. Figures 1 and 2 illustrate some of the properties of convergence lines in the Marshall Islands.

2. Lines of convergence form readily as a result of diurnal and orographic disturbances of the wind field in the neighborhood of large mountainous islands. We may then get temperature contrasts across the line, but these contrasts are confined to the lowest layers of the atmosphere. Aloft there is usually horizontal homogeneity. Orographic lines are rarely more than 200 miles long.

3. Except on such mountainous islands, the lines rarely show temperature contrast in the lowest layers but may do so in the upper air, as a reflection of different lapse rates in the air on either side of the line. The contrast may intensify, vanish, or reverse its sign both in space and in time but without, as far as can be discovered, affecting the location, movement, or intensity of the cloud line.

4. By far the greatest number of the lines in the equatorial Pacific, and in the Caribbean in summer, show no temperature contrast whatever below 20,000 ft. Above this level, temperature gradients across the line may have either sign or be absent.

5. In low latitudes, the vertical component of the vorticity along the line may be positive, negative, or zero. Cyclonic vorticity is not a necessary condition for the existence of the line.

6. The convergence line may or may not be accompanied by a trough in the pressure field.

7. Convergence lines may develop within an air mass which is, within the limits of observational error, completely homogeneous.

8. Convergence lines do not move with the wind, nor are pressure changes useful in forecasting the movement.

We conclude: Density contrasts and cyclonic shear are conditions which are neither necessary nor sufficient for the formation of convergence lines and hence these lines are not fronts in the sense that that concept is employed in the frontal and air-mass theory. The

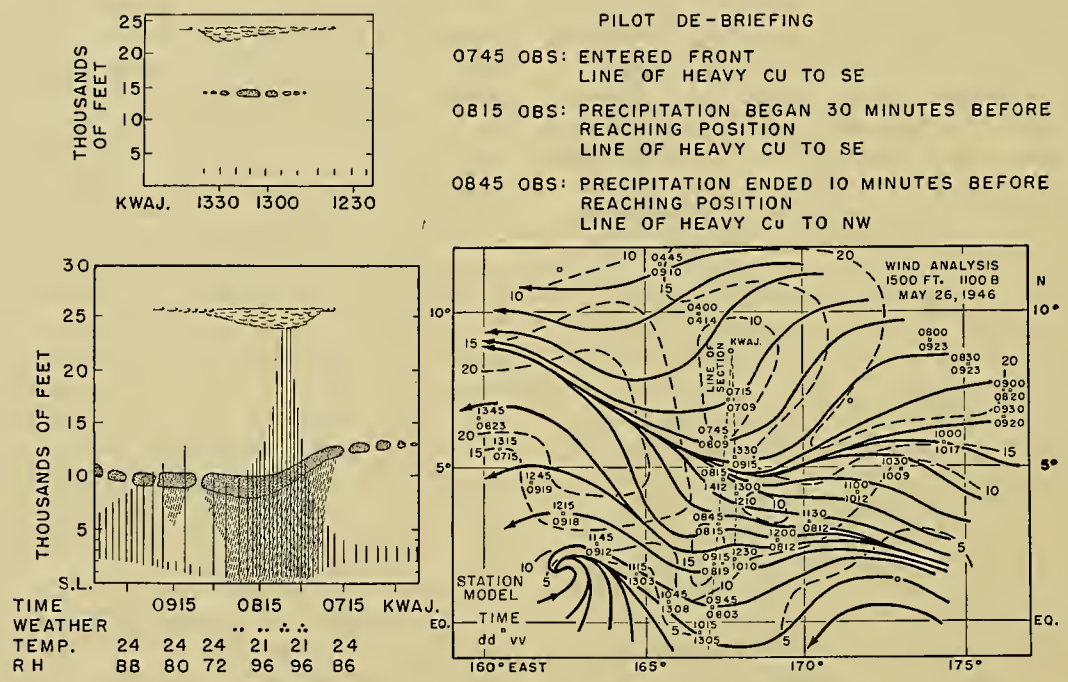


Fig. 1.—Streamline and isovel analysis at 1500 ft of an equatorial wave passing the Marshall Islands, 1100 Bikini Time, May 26, 1946, with two cloud sections observed by weather reconnaissance aircraft. Speed in knots. (Reprinted from the 7th quarterly report of the Tropical Pacific Project, U.C.L.A., a research supported by funds provided by the Geophysical Research Directorate, Air Materiel Command.)

only empirical conditions so far discovered which are necessary and sufficient are the existence of horizontal velocity convergence in the lowest layers of the atmosphere *plus* the existence of an asymptotic line in the streamline field. This statement, of course, in no sense gives a causal explanation of the lines; it merely gives descriptive correlations.

With these conclusions, we must dismiss the whole body of theoretical explanation developed by the air-mass school in the tropics, but at the same time we must incorporate their empirical findings with those of the perturbation school; this we shall attempt in the next section.

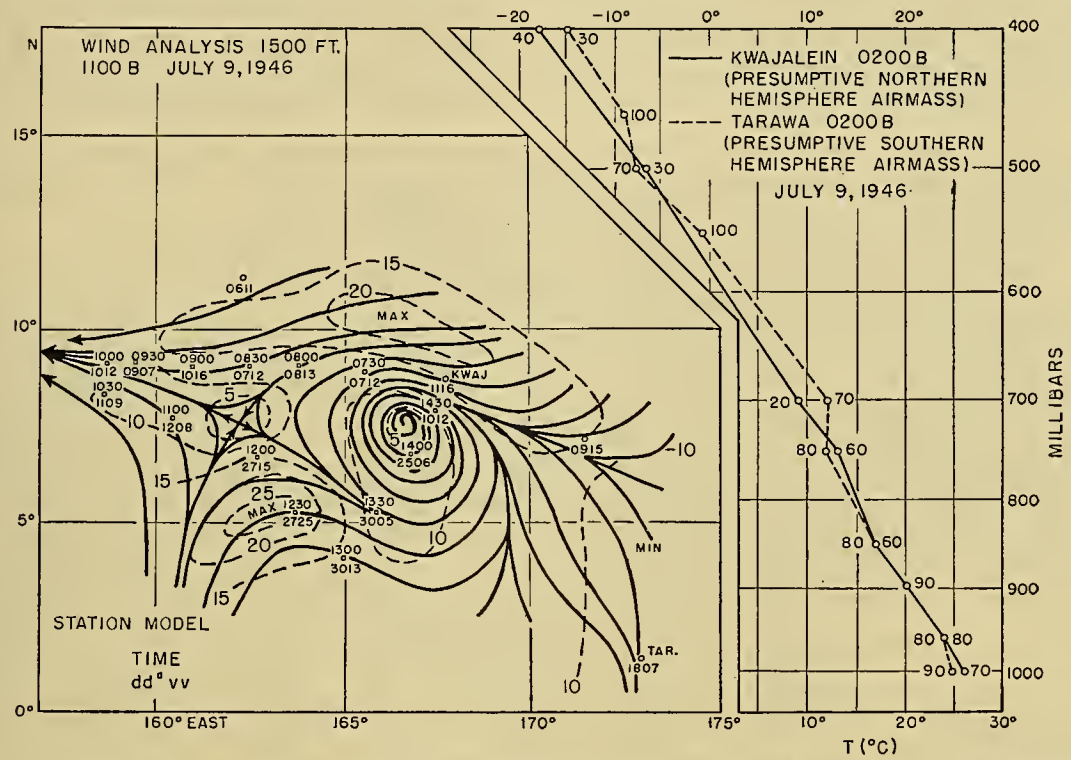


Fig. 2.—Streamline and isovel analysis at 1500 ft of an equatorial vortex passing the Marshall Islands, 1100 Bikini Time July 9, 1946, with soundings from Tarawa and Kwajalein. Speed in knots. (Reprinted from the 7th quarterly report of the Tropical Pacific Project, U.C.L.A., a research supported by funds provided by the Geophysical Research Directorate, Air Materiel Command.)

Achievements and Future of the Perturbation School. The outstanding achievement of this school is undoubtedly the discovery and description of the easterly wave and its relation to Caribbean hurricanes. Its fault lies in an almost complete neglect of the empirical findings of the air-mass school. This is to be attributed to the fact that members of the school have consistently used a streamline technique that is inadequate for the study of detailed features of the wind field. As a result, the school has placed great emphasis on the "trough" in the wind field and has marked it with a heavy line on the map, but at the same time has completely missed the asymptotes of the field which, indeed, are accurately located only by using the standard methods described by V. Bjerknes. This weakness in the synoptic methods of the school has recently been emphasized by Riehl's investigation [54] of the structure of the equatorial trough in the western Pacific. He comes to the conclusion, a correct one in my opinion, that the western part of the equatorial trough is the seat of a large number of eddies which for the most part move from east to west, either recurving ultimately as hurricanes and tropical storms or passing on to the continent of Asia. But his method of analysis is incapable of showing the relation of the "fronts" that develop *pari passu* with the vortices to the early stages of their development. From Figs. 16 and 17 in his paper, for example, and from the text, it is clear that Riehl envisages the lines of convergence as existing before the formation of the vortices. Further, the equatorial westerlies are conceived as being part of the general circulation, so that convergence lines between them and the trades, though not continuous throughout the equatorial low-pressure trough, may nevertheless exist independently of the vortex series. The result is a very complex picture of the normal state of equatorial circulation in the wet season.

When the standard methods of streamline analysis are applied to maps from this area, however, a much simpler picture is seen. Upstream, in the central Pacific, the trade winds of the two hemispheres lie side by side for the most part, forming an extensive easterly current that overlaps the equatorial low-pressure trough in both the Southern and the Northern Hemispheres. This easterly current is subject to perturbations of the same type as the easterly waves of the Caribbean save that they are at much lower latitudes and extend *across* the low-pressure trough. No lines of convergence are

found in or near the trough so long as the waves are of small amplitude. There is, of course, horizontal velocity convergence in the lower layers, associated with the wave motion, but it may be said to be "unorganized." As the waves move toward the west, some of them at least become unstable, growing in amplitude. The wave in the pressure field can now be detected with good observations, and an asymptote of convergence usually appears behind the "trough" of a wave and a corresponding asymptote of divergence ahead of it. The asymptote of convergence may be detected in flight as a line of cumulonimbus or tall cumulus. With further increase in amplitude of the wave a small vortex appears in the wave pattern, usually, but not always, in the equatorial trough itself. With the formation of this vortex, westerly winds near the equator *appear for the first time.* The vortex proceeds on its westward track; it may deepen very rapidly, forming a typhoon or storm, or it may remain weak, being accompanied in this case by one of the slight depressions that are characteristic of the doldrums in the western Pacific. When the vortex is well developed a complicated system of asymptotes accompanies it and the movement of these "equatorial fronts" is dependent not on the winds on either side of them but on the motion of the vortex itself. This brief summary of the equatorial waves and vortices is based upon recent study of the data collected during the Bikini and Eniwetok bomb tests and the whole subject has been more fully treated in recent publications [47, 48].

We are now in a position to reconcile the conflicting views that have been held concerning oceanic equatorial meteorology. We see first that the equatorial westerlies, as long ago realized by Piddington and Reid, are merely the mean expression of the fact that the vortices form in the central parts of the Pacific and move toward the west; there the passage of so many of them north and south of the equator and in both the wet and the dry season results in a mean west wind on the equator *in the western part of the ocean.* Since this west wind is the statistical result of the vortices it is clearly fallacious to import them into a hypothetical general circulation which is then perturbed to produce the vortices. Our task is to explain the mean maps as the integration of numerous synoptic maps, not to explain the synoptic map in terms of the mean and its perturbations. In other words synoptic models must always be checked by what is known of the mean circulation, but we cannot, except under very special circumstances, infer the synoptic models from the mean map. Figures 3 and 4 show schematically the correct explanation for the tropical Pacific. Neglect of this principle led to the conception of the continuous equatorial front; because it can easily be detected on the mean maps, it was assumed that it would be found on the synoptic maps. Recent research has shown that the synoptic convergence lines are quite different from the mean convergence line; the latter, however, can be derived as the statistical result of the properties and positions of the former.

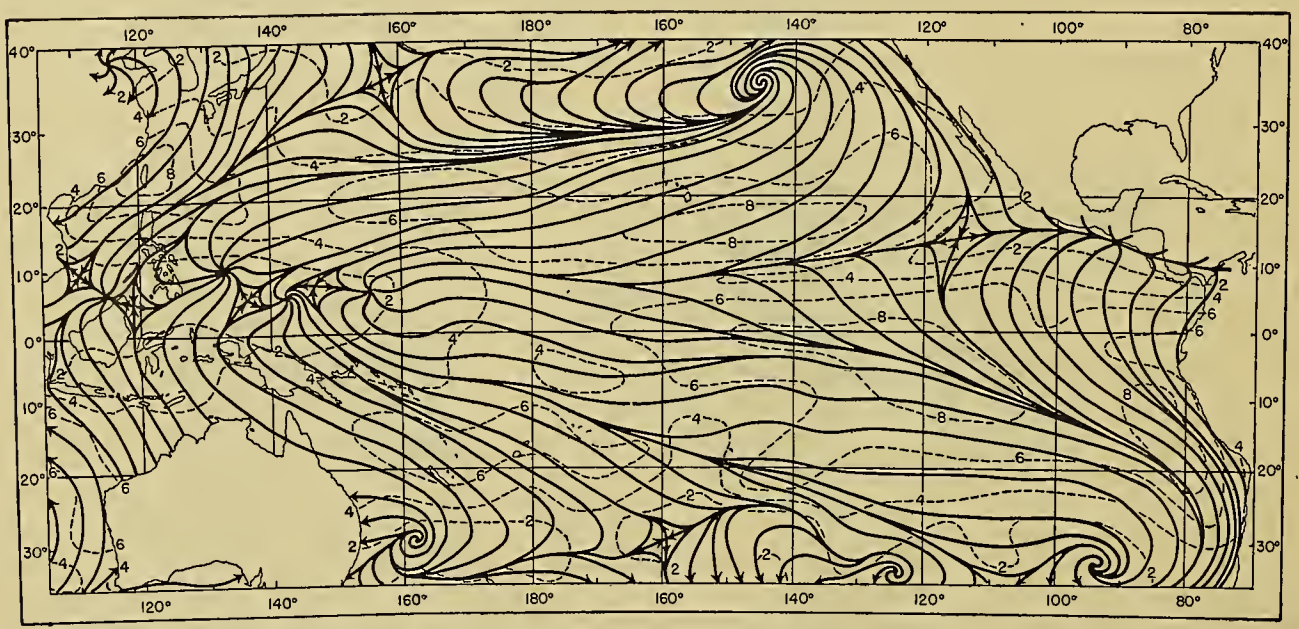


Fig. 3.—Streamline and isovel analysis of the mean resultant surface winds, tropical Pacific Ocean, Northern Hemisphere autumn. Speed in meters per second. Data derived from Deppermann and Werenskiold, and analyzed by G. A. Dean.

The special conditions under which we can pass from the mean map to the synoptic map obviously hold in the eastern parts of the oceans. Here we know that departures from the mean are very small compared with the mean. And it seems that in these regions the thermal effects long ago described by the climatological school must play at least the dominant part in driving the southeast and northeast trades. Here the direct-cell model is probably a good approximation to what actually occurs in the atmosphere. The result of the action of the direct cell, however, is not a mass ascent of the trades at the equator, and their return aloft as anti-

trades. A large part of the trade air, both at the ground and aloft, passes westward as a solid east current, quite analogous to a jet of fluid being directed into another fluid at rest. Like that jet, the east stream spreads downstream into its surroundings. It is also unstable, so that the slightest disturbance of the stream leads to waves which grow in amplitude and ultimately form a series of vortices which either move out of the stream into higher latitudes or end stagnating in the "monsoon low" over the land in the extreme west. Downstream the air is thoroughly mixed, there are only small temperature gradients, and it is here, if anywhere, that the dynamic mechanisms suggested by Rossby in his latest paper will be valid (see Fig. 4).

We have not, it will be noticed, done more than suggest a working hypothesis to apply to the large-scale features of the motion in a single ocean. It is premature to advance beyond this point while we have so little data over the land. But a great step forward would be taken if we could explain the main features of the circulation over one ocean. The gap in the data is clearly in the eastern oceanic sectors. The only extended series of observations in such a sector are those of the *Meteor* in the Atlantic [18]. We require similar data for the eastern sectors of the North and South Pacific and the South Indian Oceans, and clearly the first region to investigate should be the eastern sectors of the tropical Pacific, about which little is known. A full program of upper-air soundings in this region would enable us to decide whether the direct-cell model of the trades is to be taken seriously, for, with comparative isolation by the high mountain barriers to the east, it is here that we would expect the most favorable conditions for a direct thermal circulation to develop.

There remain other projects which we have suggested in the previous discussion of the perturbation school. The elucidation of the dynamics of easterly waves in very low latitudes is badly needed—an elucidation that, by including vertical motions as well as horizontal, would cover the main features of both the equatorial and the Freeman wave, since the latter is clearly only a limiting case of the former. It is probably too early to ask for a dynamic explanation of the kinematic features we now know attend the transformation of an equatorial wave into a vortex, since these will clearly involve perturbations of large and changing amplitude, with the consequent difficulty of solving nonlinear equations. But we may fairly ask the dynamic meteorologists for a treatment of the relation between the energy of the equatorial perturbations and that of the east current, particularly as the synoptic evidence suggests that kinetic energy is transferred from the east current to the perturbations as the latter move downstream. The approach of Kuo [40] here seems the most promising.

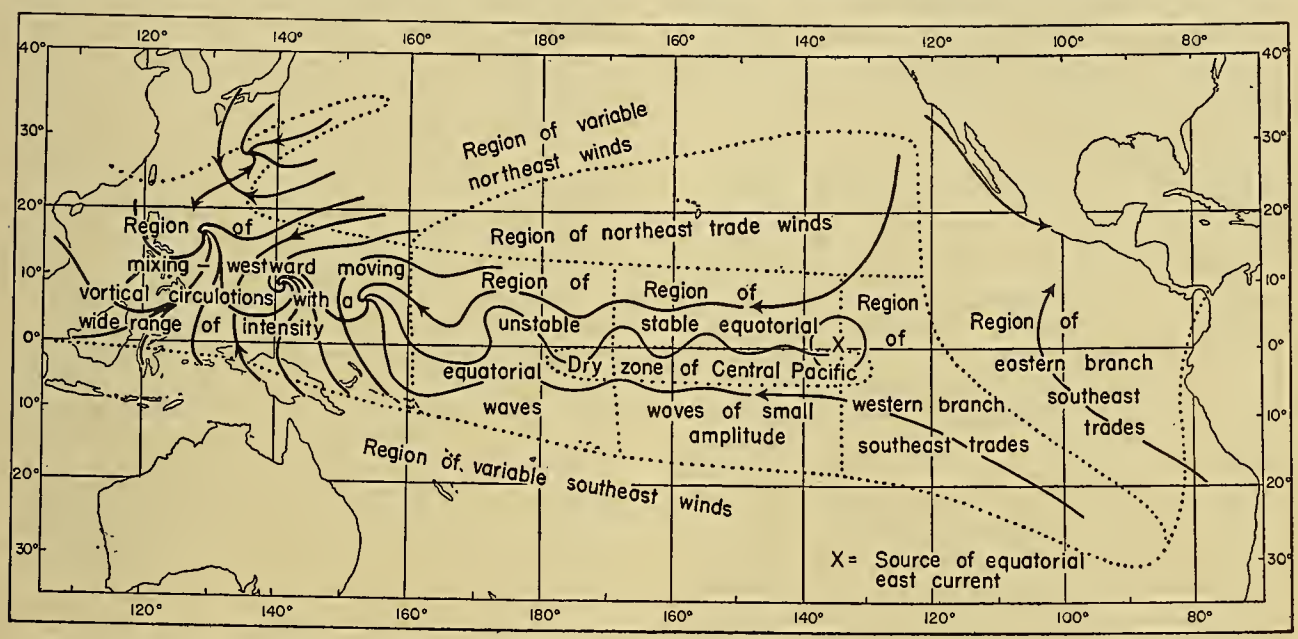


FIG. 4.—Schematic explanation of the mean circulation of Fig. 3 in terms of a "general circulation" (trades and equatorial easterliès) and its perturbations, stable and unstable.

The problem of tropical-cyclone formation is at present so different in aspect from what it was ten years ago that one is encouraged to think that the rate of progress will be maintained and will result in a complete solution in the near future. It now appears that the problem is that of deciding the sufficient and necessary conditions for the rapid deepening of a vortex already formed from an easterly wave, whether this be in high tropical latitudes, as in the Caribbean, or in low latitudes, as in the eastern Atlantic and the Pacific. We know that this deepening never occurs over the land, in spite of a plentiful supply of vortices; clearly a very moist atmosphere is necessary for the process to occur. But, though we have cause for optimism concerning the future of hurricane research, the problems connected with forecasting the movement, especially the recurvature, of tropical storms, remain as they were

twenty years ago, in spite of much work [44, 55, 66]. All forecasting methods so far suggested suffer from the defect that exceptions to the deduced rules occur too frequently in practice. Some methods suffer from the defect that they merely shift the forecasting problem from one field to another and often more difficult field. Thus all methods based upon isotherm steering suffer from the fact that it is frequently more difficult to forecast the isothermal field at upper levels than to forecast the track of the storm itself.

Above all we need an extensive investigation of the necessary and sufficient conditions for the formation of rain in the tropics. In spite of published accounts of the precipitation of cumulus clouds whose tops are well below the zero isotherm, and in spite of the use of this fact by all forecasters operating in oceanic tropical regions, one is constantly surprised by the scepticism concerning it among high-latitude meteorologists. The first worker to publish a complete atlas of precipitating tropical clouds will be performing a great service to meteorology, for though he may solve no problems he will at least draw the attention of the scientific world to the fact that the Bergeron-Findeisen theory of precipitation gives a very incomplete account of the causes of colloidal instability in clouds. In that atlas, I hope, will appear good pictures of the *complete* precipitation of a cumulus, *i.e.*, precipitation not only from the bottom but also from the sides and overhanging top of the cumulus—the almost explosive upsetting of the colloidal stability throughout the cloud mass. Deppermann has already published photographs of such clouds, which he calls "melting" cumulus—a very descriptive term [22]—but apparently he did not realize the great theoretical importance of the phenomenon he was depicting. The problem of precipitation has its synoptic aspects also. The present observations, as transmitted in the international code, give us very little clue to the type of rain that is falling, whether it be from an upper sheet of altostratus, unconnected with cumulonimbus, from altostratus connected with very distant cumulonimbus pillars, from cumulonimbus, from cumulus of great vertical extent, or from comparatively shallow cumulus. The result often enough is that the forecaster cannot tell whether the rain is connected with an extensive region of convergence in the lower atmosphere or whether it is to be attributed to the colloidal instability of cumulus clouds in a region which is in total divergent and subsiding.

Connected with this problem is that of the lapse rates of temperature and humidity within tropical clouds. It has been claimed that, level for level, tropical clouds are colder than their environment. A recent paper by Barrett and Riehl [4] has, however, pointed out the possible origin of such claims and has rightly emphasized the observational difficulties. More work, however, needs to be done. On the theoretical side, the recent work on entrainment [68] of air in tropical clouds seems to be capable of leading to the solution of an old problem of tropical forecasters—the problem of what to do with the radiosonde observations. Many soundings in the tropics seem to be completely uninformative. The lapse rate approaches very closely to the moist-adiabatic over a deep layer, and this can be clearly correlated with existing weather. However, without noticeable change in the weather one may next receive a sounding that differs from the first in the lapse rate both of temperature and of humidity, without apparent reason. It might be suggested that these vagaries are connected with the passage of the apparatus through different parts of cumulus and cumulonimbus clouds and that the soundings reflect the microstructure of the tropical air, not the broad-scale features we wish to study synoptically. Theoretical knowledge of the factors affecting the lapse rates in clouds derived from entrainment studies, together with information on the path of the sonde with respect to existing cloud masses, might do much to remove the perplexity that the present soundings sometimes cause.

Conclusion. One would like to conclude as a result of this survey that tropical meteorology, the most ancient branch of scientific meteorology, has shown more progress in the last ten years than in the previous fifty. However that may be, it is certain that it now offers a rich field for research to the climatologist, synoptician, and dynamic meteorologist alike. To the climatologist it offers the challenging task of collecting, reducing, and studying the unique wartime tropical data before they become lost forever; to the synoptician it offers the prospect of new empirical discoveries in a field from which the major errors have just been cleared away—discoveries, moreover, that cannot fail to react on the theory of fronts and air masses in high latitudes; to the dynamic meteorologists it presents a series of problems in perturbation theory, applicable to an atmospheric region which, more closely than any other on this earth, resembles his ideal: the homogeneous, horizontally moving, frictionless, and incompressible fluid.

REFERENCES

1. Alpert, L., "The Intertropical Convergence Zone of the Eastern Pacific Region." *Bull. Amer. meteor. Soc.*, 26: 426–432 (1945).
2. Arakawa, H., "The Formation of Hurricanes in the South Pacific and the Outbreaks of Cold Air from the North Polar Regions." *J. meteor. Soc. Japan*, Ser. 2, Vol. 18, No. 1, pp. 1–6 (1940).
3. Barnett, M. A. F., *The Cyclonic Storms in Northern New Zealand on the 2nd February and the 26th March, 1936.* New Zealand Meteor. Off. Note No. 22, 34 pp., Wellington, 1936.
4. Barrett, E. W., and Riehl, H., "Experimental Verification of Entrainment of Air into Cumulus." *J. Meteor.*, 5: 304–307 (1948).
5. Bergeron, T., "Richtlinien einer dynamischen Klimatologie." *Meteor. Z.*, 47: 246–262 (1930).
6. Bjerknes, J., "La circulation atmosphérique dans les latitudes sous-tropicales." *Scientia, Sér. phys.-math., Paris* (1935).
7. —— "Theorie der aussertropischen Zyklonenbildung." *Meteor. Z.*, 54: 462–466 (1937).
8. —— and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
9. Bjerknes, V., and others, *Dynamic Meteorology and Hy-*

drography, Part II—Kinematics. Washington, D. C., Carnegie Institution, 1911.

10. —— *Physikalische Hydrodynamik.* Berlin, J. Springer, 1933.
11. Brooks, C. E. P., and Braby, H. W., "The Clash of the Trades in the Pacific." *Quart. J. R. meteor. Soc.*, 47: 1–13 (1921).
12. Bunker, A. F., and others, "Vertical Distribution of Temperature and Humidity over the Caribbean Sea." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 11, No. 1, 82 pp. (1949).
13. Burns, M. C., *Frontal Conditions at the Line Islands and other Equatorial Islands.* Fleet Weather Central, Navy No. 128, Paper No. 2, p. 7, 1944.
14. Buxton, E. B., *Notes on Weather Analysis in the Tropics.* Coll. Papers prepared for Pan American Airways. U.S. A.A.F. Weather Research Center 4, No. 4, 1942.
15. Civilian Staff, Institute of Tropical Meteorology, "Tropical Synoptic Meteorology" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See pp. 763–803)
16. Cressman, G. P., "Relations Between High and Low Latitude Circulations." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 24, pp. 65–100 (1948).
17. Day, J. A., *Synoptic Analysis in the Tropical Pacific.* San Francisco, Transpacific Division, Pan American Airways, 1942.
18. Defant, A., Hsgbr., *Deutsche Atlantische Expedition auf dem Forschungs- und Vermessungsschiff* Meteor *Wiss. Ergebn.* Berlin und Leipzig, W. Grugter, 1933.
19. De Monts, "Rôle catalyseur de l'air polaire dans la génèse d'un cyclone tropical." *Ann. Phys. Globe France d'Outremer*, 2: 178–187 (1935).
20. Deppermann, C. E., *Outlines of Philippine Frontology.* Manila, Bureau of Printing, 1936.
21. —— *Are There Warm Sectors in Philippine Typhoons?* Manila, Bureau of Printing, 1937.
22. —— *The Weather and Clouds of Manila.* Manila, Bureau of Printing, 1937.
23. —— *Upper Air Circulation over the Philippines and Adjacent Regions.* Manila, Bureau of Printing, 1940.
24. Directorate of Meteorological Services, Royal New Zealand Air Force, *The Climate of Pukapuka—Danger Islands.* Climatological Notes, South Pacific Region, New Zealand Meteor. Off., Ser. C, No. 3, 1943.
25. Dunn, G. E., "Cyclogenesis in the Tropical Atlantic." *Bull. Amer. meteor. Soc.*, 21: 215–229 (1940).
26. Durham, C. O., and Collaborators, *A Study of Waves in the Easterlies.* Army Research Unit, U. S. Ninth Weather Region, Rio Piedras, P. R., 1945.
27. Espy, J. P., *Philosophy of Storms.* Boston, C. C. Little and J. Brown, 1841.
28. Fletcher, R. D., "The General Circulation of the Tropical and Equatorial Atmosphere." *J. Meteor.*, 2: 167–174 (1945).
29. Forsdyke, A. G., "Weather Forecasting in Tropical Regions." *Geophys. Mem.*, No. 82 (1949).
30. Freeman, J. C., Jr., "An Analogy Between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5: 138–146 (1948).
31. Frolow, S., "La frontologie aux Antilles." *Ann. Phys. Globe France d'Outre-mer*, 5: 118–124 (1938).
32. Garbell, M. A., *Tropical and Equatorial Meteorology.* New York, Pitman, 1947.
33. Grimes, A., "The Movement of Air Across the Equator." *Mem. Malay. meteor. Serv.*, No. 2 (1938).
34. —— *Notes Concerning the Inter-Tropical Front Between Longitudes 60 E and 160 E.* Malayan Meteorological Service. (Written in 1941.)
35. Harmantas, L., "Extracts from Notes Derived from Synoptic Maps and Flights in the Manila-Guam Sector." *Bull. Amer. meteor. Soc.*, 20: 301–303 (1939).
36. Haurwitz, B., and Austin, J. M., *Climatology.* New York, McGraw, 1944.
37. Kidson, E., "The Cyclone Series in the Caribbean Sea, October 17–24, 1935." *Quart. J. R. meteor. Soc.*, 63: 339–352 (1937).
38. —— and Holmboe, J., *Frontal Methods of Weather Analysis Applied to the Australia-New Zealand Area.* New Zealand Meteor. Off., Wellington, 1935.
39. Kuhlbrodt, E., "Das Strömungs-system der Luft über dem tropischen Atlantischen Ozean nach den Höhenwind-messungen der *Meteor*-Expedition." *Z. Geophys.*, 4: 385–386 (1928).
40. Kuo, H.-L., "Dynamic Instability of Two-Dimensional Nondivergent Flow in a Barotropic Atmosphere." *J. Meteor.*, 6: 105–122 (1949).
41. Leopold, L. B., "The Interaction of Trade Wind and Sea Breeze, Hawaii." *J. Meteor.*, 6: 312–320 (1949).
42. Loomis, E., "On the Storm Which Was Experienced Throughout the United States About the 20th December, 1836." *Trans. Amer. phil. Soc.*, (New Ser.) 7: 125–163 (1841).
43. Milham, W. I., *Meteorology.* New York, Macmillan, 1926. (See pp. 277–280)
44. Mintz, Y., "A Rule for Forecasting the Eccentricity and Direction of Motion of Tropical Cyclones." *Bull. Amer. meteor. Soc.*, 28: 121–125 (1947).
45. Namias, J., *Methods of Extended Forecasting.* Washington, D. C., U. S. Weather Bureau, 1943.
46. Naval Meteorological Branch, Admiralty, London, *Tropical Climatology and Forecasting in the Far East.* Memo 131/44, 1944.
47. Palmer, C. E., *The Equatorial Front.* Air Materiel Command, Base Directorate for Geophysical Research (Appendix to the 5th Quarterly Report of the Tropical Project, U.C.L.A.), Cambridge, Mass., 1949.
48. —— and Ellsaesser, H. W., "Notes on Tropical Meteorology." *2143rd Air Weather Wing Tech. Bull.*, 1: 1–49 (1949).
49. Petterssen, S., *Introduction to Meteorology.* New York, McGraw, 1941. (See pp. 110–117)
50. R.A.A.F. Meteorological Services, *Forecasting Weather in the Australian Equatorial Regions. Part III—Synoptic Meteorology.* Melbourne, R.A.A.F. Meteorological Services, March, 1944.
51. Ramanathan, K. R., and Ramakrishnan, K. P., "The General Circulation of the Atmosphere over India and Its Neighbourhood." *Mem. Indian meteor. Dept.*, 26: 190–195 (1939).
52. Reid, W., *The Law of Storms.* London, Weale, 1849.
53. Riehl, H., "Waves in the Easterlies and the Polar Front in the Tropics." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 17 (1943).
54. —— "On the Formation of Typhoons." *J. Meteor.*, 5: 247–264 (1948).
55. —— and Shafer, R. J., "The Recurvature of Tropical Storms." *J. Meteor.*, 1: 42–54 (1944).
56. Rossby, C.-G., "The Scientific Basis of Modern Meteorology" in *Climate and Man: Yearbook of Agriculture.* U. S. Govt. Printing Office, Washington, D. C., 1941. (See pp. 599–655)

57. —— "The Scientific Basis of Modern Meteorology" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See pp. 502–529)
58. —— "On the Nature of the General Circulation of the Lower Atmosphere" in *The Atmospheres of the Earth and Planets*, G. P. Kuiper, ed., Chap. II. Chicago, University of Chicago Press, 1949.
59. Roy, S. C., and Roy, A. K., "Structure and Movement of Cyclones in the Indian Seas." *Beitr. Phys. frei. Atmos.*, 16: 224–234 (1930).
60. Sawyer, J. S., "The Structure of the Intertropical Front over N.W. India during the S.W. Monsoon." *Quart. J. R. meteor. Soc.*, 73: 346–369 (1947).
61. Schott, G., *Geographie des Indischen und Stillen Ozeans*. Hamburg, C. Boysen, 1935.
62. —— "Klimakunde der Südsee-Inseln, *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. 4, Teil T. Berlin, Gebr. Borntraeger, 1938.
63. Schove, D. J., "A Further Contribution to the Meteorology of Nigeria." *Quart. J. R. meteor. Soc.*, 72: 105–110 (1946).
64. Scofield, E., "On the Origin of Tropical Cyclones." *Bull. Amer. meteor. Soc.*, 19: 244–256 (1938).
65. Simpson, R. H., *Synoptic Meteorology of the Tropics*. Panama, U.S.A.A.F. Tropical Weather School, 1945.
66. —— "On the Movement of Tropical Cyclones." *Trans. Amer. geophys. Un.*, 27: 641–655 (1946).
67. Solot, S. B., *The Meteorology of Central Africa*. Accra, U.S.A.A.F. Weather Research Center, 19th Weather Region, 1943.
68. Stommel, H., "Entrainment of Air into a Cumulus Cloud." *J. Meteor.*, 4: 91–94 (1947).
69. U. S. Weather Bureau, *Atlas of Climatic Charts of the Oceans*. Washington, D. C., U. S. Govt. Printing Office, 1938.
70. van Bemmelen, W., "The Atmospherical Circulation above Australasia According to the Pilot-Balloon Obvations Made at Batavia." *Proc. K. Akad. Wetenschap. Amsterdam*, 20: 1313–1327 (1918).
71. Vuorela, L. A., "Contribution to the Aerology of the Tropical Atlantic." *J. Meteor.*, 5: 115–117 (1948).
72. Watts, I. E. M., *The Equatorial Convergence Lines of the Malayan-East Indies Area*. Singapore, Government Printing Office, 1949.
73. Wexler, H., "Structure of Hurricanes as Determined by Radar." *Ann. N. Y. Acad. Sci.*, 48: 821–844 (1946–1947).

EQUATORIAL METEOROLOGY

By A. GRIMES[1]

Farnborough, Hants, England

INTRODUCTION

Equatorial meteorology may be regarded as the meteorology of the region about the equator where the Coriolis term in the equations of motion is no longer the only one which is of the same order of magnitude as the pressure term. Indeed, on the equator itself the Coriolis force is zero and the other terms must balance the pressure gradients which undoubtedly exist. One cannot quote any investigation which will indicate the limits of the region thus defined, but the general impression given by most writers is that the Coriolis force ceases to predominate between the latitudes of 15°N and 15°S, and it is to this region, roughly one-quarter of the surface of the earth, that the following discussion will be confined. The region, extending over the central portion of the Atlantic, Pacific, and Indian Oceans, is largely water broken by the land masses of central Africa, Central America and the northern part of South America, the East Indies, southern India, and Ceylon. Generally speaking, the land areas are sparsely populated and meteorological history is comparatively short, extending back only so far as the beginning of international civil aviation for most places in the region. This is not to say that no extended series of meteorological records exists in the region, but the number of stations with reliable records over a period greater than ten to twenty years is pitifully small for the large area that has to be considered. The situation is steadily improving, because of the pressing needs of international aviation, and the meteorological services in the more backward—meteorologically speaking—of the countries in the region are being more and more extended as time goes on. Nevertheless, the advancement lags considerably behind the development in the temperate zone. This is principally due to the high cost of meteorological equipment, such as the radiosonde, which puts modern methods of observation beyond the resources of the local governments. The number of scientific workers in the region is very small, and since even this small number exist almost entirely for the purpose of issuing forecasts to aircraft, it is not surprising that little meteorological information of any importance has emerged, and it is fairly certain that little will emerge until more equipment and more scientists become available to the region. The standard of observing, except at a few stations, is low in many cases because of the poorness of the equipment, and in many more cases because of the shortage of competent observers. Steps are being taken to improve the observations, but it must be many years before adequate and accurate climatological statistics will be available for most of the region.

1. Formerly Director, British West African Meteorological Services.

The greatest impetus to the study of meteorology was given during World War II when the need for developing lines of communication through the region brought an influx of men and equipment far greater than in prewar days. During this period it was realised that our knowledge of the area was meagre, and a large number of reports, both statistical and otherwise, were published to remedy this defect. Not all the reports were of great value but at least attempts were made to explain the phenomena of the equatorial zone on a scientific basis: if little in the way of coherent meteorological theory emerged, there were at least some ideas suggested that might serve as a starting point for further investigation.

THE AIR MASSES

Except for the occasional intrusion of polar air into the South China Sea area, it seems certain that the only air masses affecting the region are tropical and equatorial. The tropical air masses are those of the trade winds giving on an average a belt of northeast winds in the Northern Hemisphere and of southeast winds in the Southern Hemisphere, both derived from the subtropical anticyclones. During the summer of the Northern Hemisphere the trade wind of the Southern Hemisphere crosses the equator everywhere, so far as can be determined, and becomes what is usually known as the monsoon. In the Northern Hemisphere winter the reverse process takes place everywhere except in the Atlantic and West African areas, and it is these annual movements of air which provide the variations in the weather of the equatorial zone. When the tropical air from either hemisphere has passed into the equatorial zone and becomes relatively stagnant it is known as equatorial air and is therefore not an "air mass" as one usually thinks of the term. Such stagnant air, which frequently develops a light anticyclonic circulation over land masses, is invariably moist (or becomes moist) in the lower layers. This air is a common feature of the synoptic situation and is therefore of great importance.

The properties of tropical air are quite well known up to 4 or 5 km, but observations are too few above 5 km for reliable conclusions to be drawn. It seems that above this level the winds are mainly easterly. Some investigators consider that these winds play a large part in the variation of weather along the equator, but what part they play has not yet been demonstrated with any assurance. It is not in fact clear from the study of daily charts where the trade winds end and the overriding easterlies begin, because in most cases the maximum speed of the trade is reached between 500 and 1500 m above the surface, and there is only a gradual variation of speed in the higher levels. When the tropical air has a continental source it is usual to

regard the top of the haze layer (2–3 km) as the upper surface. This condition is confirmed by the sudden rise in humidity above this level, but when the tropical air has an oceanic source the surface of separation is ill-defined. Upper-air observations with instruments more accurate than ordinary aircraft thermometers might make it possible to define the separation of upper and lower layers with more precision.

When the surface air is equatorial, there is frequently a layer of trade or monsoon air before the overrunning easterlies are reached. The lower boundary of the trade or monsoon air is well marked by a rapid increase in wind speed with height, but the separation of trade or monsoon air and the easterlies is still ill-defined. It is quite clear that the day-to-day determination of the vertical structure of the air masses will not be possible until observations are available in greater quantity and with more precision.

FRONTAL PHENOMENA

The question of the existence or nonexistence of true frontal discontinuities in the equatorial region is one of the most controversial issues of tropical meteorology. The majority of opinion, however, favours the existence of fronts, but the very fact that many writers disagree makes it important that factual evidence should be produced. The acquisition of convincing evidence is hampered by the lack of close networks of upper-air stations, and the main arguments put forward for the existence of fronts depend on the recognition of wind discontinuities and the acceptance of such discontinuities in wind as evidence of discontinuities in the air masses. Three distinct types of discontinuity are normally recognised: the intertropical front, the meridional front, and subsidiary fronts.

The intertropical front is regarded as being formed in the intertropical convergence zone, the zone completely surrounding the earth in which the trade winds from the two hemispheres meet. In earlier days the zone was simply known as the doldrums and all the bad weather associated with the zone was regarded as due to the instability of the moist stagnant air which gave rise to convectional rain and thunderstorms. It is now more common to regard the zone as being one in which true frontal discontinuities between the northern and southern trades exist, interspersed with semi-stagnant equatorial air masses or doldrum areas. The boundary between the equatorial air masses and the trades will usually orient itself parallel to the direction of the active trade wind but with an increase or "surge" of the trade wind the relatively cold air will tend to undercut the equatorial air and give rise to frontal phenomena which are well marked and easy to identify. The front between the trades, which is known as the intertropical front, forms most frequently at the times when the trade from the winter hemisphere is beginning its penetration into the summer hemisphere and when it is receding (*i.e.*, just after the equinoxes). The advance of the trade across the equator is not a steady continuous process but is made up of surges and temporary retreats, each surge being accompanied by typical cold-frontal type phenomena; when the advancing trade dies down temporarily, the frontal effect disappears to give place to a region of equatorial air, or the trade of the summer hemisphere temporarily reasserts itself. The retreat is a little more regular, until the intertropical convergence zone has recrossed the equator, and is often accompanied by warm-front phenomena due to the flow of the freshening trade of the autumn hemisphere over the retreating trade, but more usually the activity of both trades at this period is small, and large areas of stagnant equatorial air appear on the synoptic chart.

The meridional front is the front that forms between tropical air masses from two distinct anticyclonic cells in the same hemisphere. In the Southern Hemisphere, this takes place usually between a weak ESE/E stream from the more easterly of the two anticyclonic cells and a stronger SE/SSE stream from the more westerly cell. Because of the persistence of such cells during the winters of the respective hemispheres, these fronts are of a semipermanent character and are readily identifiable. In the summer of the hemisphere being considered, the portions of the cells are more variable with a consequent greater variation in the position of the meridional front. Where the air from the other hemisphere crosses the equator the meridional front crosses from the winter to the summer hemisphere, giving rise at its junction with the intertropical convergence zone to a "triple point." There is some evidence for the belief that this "triple point" is associated with the formation of tropical cyclones.

Subsidiary fronts are of a more local character than the ones already described and are caused by local variations within an otherwise homogeneous air mass. The cause of the local variations may be orographic diversion of one portion of the air stream or the diversion of a portion of the air stream because of the breakdown of geostrophic control as it enters the equatorial region. The diverted portion subsequently meets the undiverted portion again and gives rise to a local zone of convergence, which may be very limited in extent and may fluctuate rapidly in position. The movements of these zones are very difficult to follow since the cause of the movements is not necessarily related to the winds in the neighbourhood of the zone but is related to the previous history of the air mass in which the zone has formed. The slopes of the fronts have been determined in many cases, the criterion almost invariably being the height at which the wind change occurs at various points. Most of the values lie between 1 in 300 and 1 in 500. There is no theoretical value for the slope with which comparison can be made because of the breakdown of the geostrophic equations in the area and the absence of suitable equations to replace them. It has also not been determined satisfactorily to what height the discontinuities extend, the main difficulty here being the similarity between the upper layers of the trade winds and the lower layers of the overrunning easterlies. Where a front is formed from undercutting

by a surge of the trade wind, however, the tongue of fresh air very often does not exceed 1500 m in thickness so that the frontal surface has a well-marked slope at the leading edge up to 1500 m but trails off horizontally behind. It is beyond the scope of this article to go into fuller detail in this matter of fronts; a complete account has been given by Forder [3].

THE EQUATIONS OF MOTION

The equatorial region was defined earlier as the region in which the Coriolis term in the equations of motion was no longer predominant in balancing the pressure term, so that although many of the problems of the equatorial zone are the same as those of the tropics generally, the particular problem of finding dynamical equations applicable to the movements of air within the region is peculiar to the region.

When one examines synoptic charts of the region, particularly during the northern summer, one is struck immediately by the remarkable uniformity in the speed and direction of the winds a little south of the equator and the great diversity of the winds, arising from the same air stream, after crossing the equator. Between 5°S and 10°S in the Indian Ocean–Pacific area, for example, the winds are consistently from southeast to east, whereas north of the equator the winds may be from any direction from west through south to east. At the same time, the winds to the south are not very far from being geostrophic, whereas north of the equator the winds are far from being geostrophic in most cases.

Grimes [6] showed that it was possible to obtain integrable equations of motion allowing for the variation of the Coriolis force with latitude, provided the conditions were very much simplified. The simplifications involved the assumption of steady horizontal motion with no friction and no variation in the motion with changing longitude. The further assumption that $\sin\phi = \phi$ (ϕ = latitude) was also made. Starting with an assumed value for the wind speed and direction at latitude 5°S and using Bjerknes' circulation theorem, Grimes was able to compute the paths of the air across the equator and, from the same initial wind, to arrive at a variety of paths by simply changing the value of the assumed circulation of the air at latitude 5°S. It was also possible to compute the isobaric configuration that was necessary to maintain the motion, but the computed gradients of pressure were much smaller than the ones found in practice. The equations showed, too, that if an isobar crossed the equator, it must do so at right angles, but this was a consequence of the assumption of no variation with longitude and Crossley [1] pointed out that when this restriction was removed the necessity of an isobar's crossing the equator at right angles disappeared. Crossley [2] has recently modified Grimes's treatment to obtain an exact solution using spherical polar coordinates instead of Cartesian coordinates. On the assumption that the motion is steady and frictionless and that there is no change of velocity with longitude, the solution of the equations of motion becomes

$$u = -2\Omega a\cos\phi + B\phi + C, \tag{1}$$

$$v = A\sec\phi, \tag{2}$$

where ϕ is the latitude, λ is the longitude, a is the radius of earth, and A, B, and C are arbitrary constants. The streamlines are given by

$$A\lambda + 2\Omega a\sin\phi - \tfrac{1}{2}B\phi^2 - C\phi = \text{const}, \tag{3}$$

and the pressure distribution by

$$-\frac{p}{\rho} = AB\lambda + \tfrac{1}{2}A\tan^2\phi + \Omega a[\Omega a\cos 2\phi + 2B(\sin\phi - \phi\cos\phi) - 2C\cos\phi] + \text{const}. \tag{4}$$

The arbitrary constant A can be interpreted as the northward component of velocity at the equator.

If we set $B = 0$ so that the pressure distribution is independent of longitude, we arrive at Grimes's solution in spherical polar form:

$$u = C - 2\Omega a\cos\phi, \tag{5}$$

$$v = A\sec\phi. \tag{6}$$

The constant C is related to the assumed initial velocity and vorticity.

Crossley goes on to say that if solutions of the Grimes type are found to exist, it follows that the gradient of pressure by itself is not sufficient to determine the horizontal motion of air in the tropics. In middle latitudes the assumption that the wind vanishes with the pressure gradient is in accord with experience, but this is not necessarily the case in the tropics. The theory indicates that once a motion across the isobars has come into existence on a large scale, this type of flow can continue in a similar fashion without any marked tendency to approximate the direction of the isobars unless higher latitudes are reached. It is clearly important, as Crossley says, to determine how much cross-isobar motion exists in the free air and whether there is, in fact, any wind in the absence of a pressure gradient.

Treloar [9] and Gibbs [5] have separately discussed the same problem, introducing the frictional effect but ignoring the space accelerations. Their treatment, however, cannot be regarded as altogether satisfactory.

Another method of approach to the problem may be suggested. The equations for frictionless horizontal motion may be written

$$\frac{du}{dt} - 2\Omega v\sin\phi = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \qquad \frac{dv}{dt} + 2\Omega u\sin\phi = -\frac{1}{\rho}\frac{\partial p}{\partial y}, \tag{7}$$

where all the symbols have their usual significance.

These equations may be expanded and rewritten

$$\begin{aligned}\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(\frac{u^2+v^2}{2}\right)-v\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)-2\Omega v\sin\phi \\ =-\frac{1}{\rho}\frac{\partial p}{\partial x},\\ \frac{\partial v}{\partial t}+\frac{\partial}{\partial y}\left(\frac{u^2+v^2}{2}\right)+u\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)+2\Omega u\sin\phi \\ =-\frac{1}{\rho}\frac{\partial p}{\partial y}.\end{aligned} \tag{8}$$

Equating the vorticity $(\partial v/\partial x - \partial u/\partial y)$ to 2ω and writing $u^2 + v^2 = V^2$, we obtain

$$\begin{aligned}\frac{\partial u}{\partial t}-(2\omega+2\Omega\sin\phi)v&=-\frac{1}{\rho}\frac{\partial}{\partial x}(p+\tfrac{1}{2}\rho V^2),\\ \frac{\partial v}{\partial t}+(2\omega+2\Omega\sin\phi)u&=-\frac{1}{\rho}\frac{\partial}{\partial y}(p+\tfrac{1}{2}\rho V^2),\end{aligned} \tag{9}$$

if we may assume that the density ρ is sensibly constant —an assumption which appears legitimate in the equatorial region.

Following Rossby [7], we shall call 2ω the relative vorticity and $(2\omega + 2\Omega \sin \phi)$ the absolute vorticity of the motion. Rossby [7, p. 71] has shown that after pressure has been eliminated the equations of motion reduce to

$$\begin{aligned}\frac{d}{dt}(2\omega+2\Omega\sin\phi)\\ =-(2\omega+2\Omega\sin\phi)\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right),\end{aligned} \tag{10}$$

so that with the equation of continuity $(\partial u/\partial x + \partial v/\partial y = 0)$ we obtain

$$\frac{d}{dt}(2\omega+2\Omega\sin\phi)=0. \tag{11}$$

Therefore, $2\omega + 2\Omega \sin \phi = k$, where k is a constant along a trajectory but may differ from one trajectory to another. Writing k for $(2\omega + 2\Omega \sin \phi)$ and P for $(p + \frac{1}{2}\rho V^2)$ in equations (9) (P is the dynamic pressure), we obtain

$$\begin{aligned}\frac{\partial u}{\partial t}-kv&=-\frac{1}{\rho}\frac{\partial P}{\partial x},\\ \frac{\partial v}{\partial t}+ku&=-\frac{1}{\rho}\frac{\partial P}{\partial y}.\end{aligned} \tag{12}$$

The equations for steady streaming $(\partial u/\partial t = \partial v/\partial t = 0)$ reduce to

$$\begin{aligned}-kv&=-\frac{1}{\rho}\frac{\partial P}{\partial x},\\ ku&=-\frac{1}{\rho}\frac{\partial P}{\partial y}.\end{aligned} \tag{13}$$

Equations (13) represent a pseudogeostrophic motion in which the streamlines are strictly parallel to the isobars of P and in which the space accelerations are allowed for, so that no further allowance need be made for the accelerations along the path or for the centripetal acceleration perpendicular to the path. Furthermore, the equations are independent of variations in latitude, and the variations imposed upon the motion by the change of latitude are controlled by the constancy of the absolute vorticity along the streamlines.

The constant k may vary from streamline to streamline but will be constant along any particular streamline; for practical purposes it may be regarded as constant between any pair of consecutive isobars (of P) regarded as streamlines.

If we draw isobars of P at intervals of whole millibars, we can determine the value of k with an ordinary geostrophic wind scale at any point where a value of V is known. If we write

$$k = 2\omega + 2\Omega \sin \phi = 2\Omega \sin \phi', \tag{14}$$

the value of ϕ' can be read from the scale as the value appropriate to the known value of V and the separation of the isobars at the points. Since k is constant along the whole length of the stream "tube" enclosed by the particular pair of isobars, the value of relative vorticity 2ω can be determined immediately from the equation,

$$2\omega = 2\Omega (\sin \phi' - \sin \phi). \tag{15}$$

It is thus possible to identify the areas of cyclonic and anticyclonic vorticity when the motion is one of steady streaming and where values of P can be determined.

Near the equator, where the geostrophic component of acceleration becomes small and the space accelerations become relatively more important, the advantage of explicitly eliminating changes in ϕ and in the space accelerations from the equations of motion is obvious. Practically, if time changes can be ignored, the advantage in using dynamic pressure is that the streamlines should be parallel to the isobars of P so that wind direction will indicate the direction of the isobars and vice versa, a facility so far denied to equatorial analysts.

The practical usefulness of this method depends on the feasibility (or possibility) of drawing the isobars of dynamic pressure P. Wind speeds are small in the equatorial region and so are pressure gradients, so that an accuracy in pressure reductions of at least a fifth of a millibar will be necessary if the method is to be tested. In British West Africa the altitudes of the stations above sea level are not known with an accuracy sufficient to guarantee this, and attempts to verify the deductions in this area have had to be abandoned until the altitudes of the stations have been redetermined.

It was stated earlier that the measured pressure gradients were much greater than previous theory demanded. This difficulty may now be overcome by assuming values of the relative vorticity near the equator greater than the ones Grimes assumed in his first calculations.

So far we have discussed only cases of steady streaming, but one of the most important features of equa-

torial weather is the undercutting of stagnant equatorial air by a surge of fresh tropical air, producing effects similar to those of a cold front in temperate latitudes. This situation has been treated in a most ingenious manner by Freeman [4], who draws an analogy between the undercutting of the equatorial air by a sudden surge of cold air and the supersonic gas flows resulting from the action of a piston. In his discussion Freeman assumes that the air is at rest under an inversion of constant height near the equator and that a cold-air mass, acting as a piston, moves into the area, undercutting the stagnant air. He then computes the characteristic or Mach lines for the resulting flow and shows that the theory admits of three types of disturbance oriented normal to the current, that is, a rarefaction wave, a compression wave, and a "jump." The compression wave, of course, arises from a surge of cold air, while the rarefaction wave may be caused by the falling off in the trade, although this latter case is relatively unimportant. Freeman considers that the "jump" may be the explanation of the "easterly wave" because, as he points out, when a "jump" moves into an atmosphere at rest, the energy required to maintain the "jump" is derived from the atmosphere at rest and no other source of energy is required to maintain the flow. The moving fluid will show rapid vertical motions at the "jump" which would lead to the typical weather phenomena associated with the easterly wave. The theory also provides a model for a sloping surface of discontinuity in these low latitudes which, so far, has been produced by no other theory. In the example which Freeman gives for a case in the New Guinea area, the agreement between the calculated and the observed values is most remarkable in view of the difficulties inherent in the problem.

One of the major problems that remain to be tackled is how to allow for the effect of the diurnal and semidiurnal variations of pressure which dominate the barograms of the equatorial region. The daily change in pressure is very much greater than the secular change and if one treats it loosely as the result of pressure waves that travel from east to west around the equator, the effect is to produce variable gradients of pressure which attain values of the same order of magnitude as the geostrophic and space-acceleration terms in the equations. The problem is a difficult one and not much progress appears to have been made up to now in solving it.

On the whole it is fair to say that although some progress has been made, we are still very far from a satisfactory solution to the dynamical problems of the equatorial region, but the problems are well worth attention and success in solving them might have a profound effect on meteorological theory as a whole.

FORECASTING

Since most of the meteorologists in the equatorial region owe their presence there to the necessity of providing weather forecasts, forecasts have to be made even when the methods of preparing the forecasts are only incompletely understood. Generally speaking, the forecaster labours under great difficulties, not only the economic ones of poor equipment and inadequate communications but also those which result from the fact that he is usually a stranger to the area and lacks the long familiarity with the weather changes which help to guide forecasters in temperate climates. He has no textbook methods which he can study and he has to obtain his knowledge of local conditions from woefully incomplete climatological statistics which hide rather than emphasize the variations in weather which he must account for.

For a forecaster trained in the temperate zone the greatest difficulty is caused by the breakdown of the geostrophic relationship between wind and pressure which makes the drawing of isobars within the region, especially in view of the small number of stations, appear to be a matter for imagination rather than reason. The result is that the isobaric pattern has to be drawn independently of the wind distribution and it has become usual to construct additional charts of lines of flow for the various levels independently of the pressure. By superimposing the streamline and isobaric charts it is sometimes possible to determine areas of acceleration or retardation of the streams and eventually to formulate some empirical rules for development.

One of the difficulties which have exercised the minds of many people is how to handle the diurnal variation of pressure and weather on the synoptic chart. Now that universal time has been adopted for synoptic observations the effect of the diurnal variation must be very carefully watched on any chart which has a large east–west extent so that afternoon convective effects will not be confused with frontal effects. There is a body of opinion that the pressure readings should be "corrected" for the relatively large diurnal variations of pressure in order to obtain "representative" pressures for each station, but the day-to-day irregularities, especially in the diurnal wave, make the correction of the pressures a rather doubtful project. It is indeed questionable whether such a "representative" pressure chart would have any dynamical significance and whether more might not be lost than gained in the process. It is, however, beyond question that the comparison between successive three-hourly synoptic charts is exceedingly difficult because of the diurnal variation of pressure. The tendency has been to compare each chart with the one prepared 24 hr earlier and thus obtain an overlapping series of 24-hr changes. There appears to have been no prolonged investigation into the application of corrections for diurnal variations, owing no doubt to the pressure of routine work. However, such an investigation would be of great value whatever the result.

Some workers prefer to ignore the isobaric chart almost completely and rely on the streamline charts for their analyses. It is usual to draw the streamlines for different levels in order to get a three-dimensional picture of the atmospheric structure. Usually the charts are not drawn conventionally, with the distance between the lines being inversely proportioned to the

speeds, but consist of lines of flow (direction only) with the speeds indicated by isopleths or by feathers on the arrows. It is possible, with these charts, to identify areas of convergence and divergence with reasonable reliability and also to pick out sloping surfaces of well-marked wind discontinuity. Short-period forecasts of good reliability can be made by an experienced forecaster using this method, but much more satisfactory results appear to be possible if the streamline charts are related to the isobaric charts. Sen [8] in India has advanced the method a stage further by assuming the equatorial circulation to consist of a Kármán vortex street, the location of each vortex being determined with the assistance of streamline charts and the surface isobaric chart. This conception involves the existence in the Northern Hemisphere of a line of anticyclonic vortices near the equator and a line of cyclonic vortices to the north, with the corresponding zones of weak anticyclonic and cyclonic pressure centres, respectively, to mark their positions. From the spacing ratio of the "street," the stability or instability of the system is determined, and developments are forecast accordingly. The distribution of humidity is brought into the discussion of the degree of development that is expected to take place. It is necessary, however, to assume that the geostrophic balance of wind and pressure holds to within a few degrees of the equator, which is not in accordance with the experience of most workers in the region.

The recognition of convection and orographic effects is, of course, of great importance, but since these are not features peculiar to the equatorial zone they will not be discussed here. It must be remarked, however, that by far the greater amount of rain and thunderstorms arises from these effects so that their recognition is of primary importance to the forecaster.

REFERENCES

1. Crossley, A. F., *Memo CEM/1946/1 of the Conference of Empire Meteorologists*. London, His Majesty's Stationery Office, 1946.
2. —— "On the Relation between Wind and Pressure." *Quart. J. R. meteor. Soc.*, 74:379–382 (1948).
3. Forder, D. H., *Analysis and Forecasting in the South-West Pacific Area*. Melbourne, Australian Meteorological Service (Section IV).
4. Freeman, J. C., Jr., "An Analogy between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5:138–146 (1948).
5. Gibbs, W. J., *Tropical Weather Research Bull.* No. 12. R.A.A.F. Meteorological Services, Sept. 1945.
6. Grimes, A., "The Movement of Air across the Equator." *Mem. Malay. meteor. Serv.*, Singapore, No. 2 (1938).
7. Rossby, C.-G., "Planetary Flow Patterns in the Atmosphere." *Quart. J. R. meteor. Soc.*, (Supp.) 66:68–87 (1940).
8. Sen, S. N., "Monsoon Cyclones." *Sci. Cult.*, 9:90 (1943); "Atmospheric Vortex Street." *Ibid.*, 9:453–455 (1944).
9. Treloar, H. M., *Tropical Weather Research Bull.* No. 12. R.A.A.F. Meteorological Services, Sept. 1945.

TROPICAL CYCLONES

By GORDON E. DUNN

U. S. Weather Bureau, Chicago, Illinois

INTRODUCTION

Tropical cyclones of certain degrees of intensity are known as *hurricanes* in the Atlantic Ocean, the Caribbean Sea, the Gulf of Mexico, and the eastern North Pacific Ocean (off the coast of Mexico); as *typhoons* in the western North Pacific and over most of the South Pacific Ocean; and as *cyclones* in the Indian Ocean. Locally in the Philippines these storms are called *baguios* and in Australia *willy-willies*. All have essentially the same origin, structure, and behavior.

Interest in tropical meteorology and tropical storms received considerable impetus in the 1930's with the extension of airline routes through the tropics and the consequent need of more accurate weather observations and forecasts in these areas. However, the meteorological requirements of the various military forces engaged in the tropics during World War II resulted in the collection of more observational data and in more rapid progress toward the solution of several of the more urgent and vexing problems of tropical meteorology than ever before.

A tropical cyclone is essentially a maritime phenomenon and, therefore, as a rule, aerological data are relatively sparse over large areas traversed by tropical storms. Even in the Antilles and along the South Atlantic and Gulf coasts of the United States, where raob and pilot-balloon stations are located, it has been extremely difficult to obtain upper-air observations within the storm area itself. Therefore, many of the characteristics and mechanics of the tropical storm model are still not definitely known.

SURFACE CHARACTERISTICS OF TROPICAL CYCLONES

Sources of Information. Knowledge of the surface characteristics of tropical cyclones has gradually accumulated from a half century of observations. Mitchell's work [11] on areas of origin and normal rate and direction of movement of hurricanes in the North Atlantic Ocean, together with forecasting precepts, was one of the earliest and most important studies. Several years later Cline [1] published the results of many years' study of hurricanes in the Gulf of Mexico region with considerable emphasis on their characteristics. Probably the most systematic analysis and compilation of the characteristics of tropical cyclones were those of Deppermann [2, 3, 4] for the Philippine area. Tannehill [18] has compiled a large amount of statistical information on North Atlantic hurricanes for as far back as 1494. Information on the surface characteristics of tropical cyclones from these and many other sources such as G. Norton, E. Gherzi, and others, has been summarized by Dunn [5]. These and similar sources are incomplete and in many respects not entirely satisfactory otherwise. All weather services in the tropical cyclone areas should set up procedures for a systematic tabulation of all surface and upper-air observational data to establish more definitely many of the characteristics of tropical cyclones.

Apparently very small changes in lapse rates, specific humidities, and the field of motion will result in marked deviation from the so-called "normal" features of tropical cyclones, in addition to the normal changes as they move from low to higher latitudes. In general, only those characteristics observed before recurvature are considered here.

Classification of Tropical Cyclones According to Intensity. Tropical storms are variously described as cyclones, hurricanes, storms, depressions, lows, etc. It is desirable that these terms be defined; the following classification of tropical storms has received general acceptance in the United States:

1. *Tropical Disturbance.* Circulation slight on the surface, possibly more marked aloft, one or no closed isobars. Common throughout the tropics and subtropics.
2. *Tropical Depression.* One or more closed isobars, wind force equal to or less than Beaufort 6. Frequently observed in the intertropical trough, much less frequently in the trades.
3. *Tropical Storm.* Closed isobars, wind force more than Beaufort 6 and less than Beaufort 12.
4. *Hurricane or Typhoon.* Wind force Beaufort 12 (75 mph or more).

All, of course, must originate in the tropics, but many tropical disturbances and depressions never reach hurricane intensity. An International Meteorological Conference in Manila in June 1949 adopted slightly different definitions as follows: tropical depression—winds up to 34 knots; tropical storm—winds 35–64 knots; typhoon —winds 65+ knots.

Classification of Tropical Cyclones According to Stage of Development. Like extratropical disturbances, tropical cyclones undergo constant metamorphosis from birth through maturity to decay. The general characteristics of a tropical cyclone vary considerably both as regards surface and upper-air structure as the storm progresses from one phase of development to another. Thus tropical cyclones will rather consistently exhibit different characteristics at latitude 30° than at latitude 20°. The life history of tropical cyclones may be divided into four stages:

1. The formative or incipient stage, which begins when a tropical disturbance first develops surface circulation and ends when it reaches hurricane intensity.
2. The stage of immaturity or deepening, during which the cyclone continues to deepen until the lowest central pressure and the maximum intensity are

reached. The storm is most symmetrical during this period and only a relatively small area is as yet involved.

3. The stage of maturity, when there is no further deepening, the isobars are gradually spreading out, and the area covered by gales and hurricane winds is larger than at any other period, but the intensity is gradually decreasing.

4. The stage of decay, when the storm is dissipating over land or is recurving northward and assuming extratropical characteristics.

Air-Mass Arrangements around Tropical Cyclones. Tropical cyclones develop in homogeneous air and during the early and middle stages of their existence encounter only *Tm*, tropical maritime air (the trades), and *Em*, equatorial maritime air (the equatorial westerlies and the intertropical convergence zone). During or immediately following recurvature *Np*, transitional polar air, may be encountered. There is little difference between all three air masses below 10,000 ft. Above that level *Em* is the most moist and *Np* usually the driest. The air masses around tropical cyclones are essentially homogeneous, at least until the close of the stage of maturity. There are no fronts until the stage of decay, although distortions in the wind field analogous to fronts may be found in cyclones having a strong northerly component[1] in the direction of movement.

Barometry of Tropical Cyclones. A tropical disturbance may persist for a number of days without deepening and, particularly in the intertropical convergence zone, may never develop into a depression. When the depression has intensified to Beaufort 6–7, deepening takes place rather rapidly and maximum deepening is usually reached within 72–84 hr in the Atlantic and from four to five days in the Pacific. Little change then occurs if the storm remains over water until it recurves and begins to assume extratropical characteristics and the central pressure begins to rise. There is apparently some limiting intensity, or state of equilibrium, dependent upon the moisture content and possibly the lapse rate of the lower atmosphere.

Deppermann [2] has investigated the pressure gradients of typhoons with minima less than 28.75 inches (973.6 mb) with results as follows:

Minimum pressure	Number of storms	Mean fall $in.\ hr^{-1}$	Mean fall $in.\ mile^{-1}$
Below 27.56 in. (933.3 mb)	4	0.79	0.06
27.57–27.95 in. (946.5 mb)	4	0.43	0.03
27.96–28.35 in. (960.0 mb)	7	0.45	0.03
28.36–28.75 in. (973.2 mb)	8	0.40	0.025

He concludes that, as a rule, steepness of gradient increases with depth of the barometric minimum. However, as Deppermann himself points out, the mean pressure fall per hour is determined to a considerable extent by the rate of movement of the storm and does not truly represent the pressure gradient of the storm but rather the steepness of the barograph trace. The last column does represent the actual pressure gradient, which may or may not approximate the normal pressure gradient in severe typhoons since it is based on a relatively small number of storms. Cline [1] found an average pressure gradient of 0.02 inches per mile, indicating the further spread of the isobars in the later hurricane stages which are normally found as these storms reach the coast line of the Gulf of Mexico.

There are many authentic instances of gradients much steeper than those in the table above, usually in storms in the immature stage. Several examples of steep gradients are given by Deppermann: The *Pathfinder*, anchored at Aras, Samar, P. I., recorded a pressure fall of 1.07 inches in twenty-seven minutes. In a Tacloban, P. I., typhoon there was a drop of 1.14 inches in thirty minutes, with a measured fall by mercurial barometer of 0.49 inches in five minutes. The S.S. *Virginia* in the Central Caribbean, on September 20, 1943, experienced a barometer reading of 28.74 inches at 8:00 P.M. and 27.40 inches twenty minutes later, a fall of 1.34 inches, and rose to 28.60 inches by 9:00 P.M. The pressure fell more than 2 inches in ninety minutes. The diameter of the circle enclosed by the 29.5-inch isobar was estimated at fifty miles; the calm center, from eight to twelve miles. If a flat minimum in the calm center is assumed, the pressure gradient was 2.10 inches (71.1 mb) in nineteen miles or 0.11 inches per mile. This was a young, immature hurricane which had undergone its principal development during the previous two days. In the Labor Day hurricane over the Florida Keys in 1935 there was evidence of a pressure gradient in excess of one inch in six miles.

There has been general agreement that in the immature and mature stages of the tropical cyclone isobars are nearly circular from the center outward to around the 29.4-inch (995.6-mb) isobar. However, there has never been a sufficiently dense network of accurate barometer readings to verify this completely. Indeed, evidence in recent years tends to indicate that isobars are not absolutely circular except possibly quite close to the center. Soon after the beginning of recurvature, more apparent distortion of most of the isobars is usually evident. Isobars are more truly circular in the rapidly deepening stage than at any other time.

The accepted lowest barometer reading of record in the world is 26.185 inches in a typhoon encountered by the Dutch steamship *Sapoeroea* about 460 miles east of Luzon, P.I., on August 18, 1927. The official record for the barometric minimum in the Atlantic is 26.35 inches on September 2, 1935, on Lower Matecumbe Key, Florida. Readings between 27.00 and 27.50 inches (914.3–931.3 mb) are fairly common.

Pumping—the rapid and possibly rhythmic oscillation of atmospheric pressure—has frequently been observed in tropical cyclones. Gherzi [7] claims to find a mean period of about six seconds corresponding with the period of typhoon swells, which would indicate that the oscillations in this case are produced by microseismic waves set up by the typhoon swells. Deppermann believes that oscillations of such periodicity do not occur in the Philippines, nor have they been noted in the Atlantic, although they may be masked by other

1. This expression should be interpreted to mean "a component toward the north."

oscillations. Deppermann, upon examining all Philippine barograms of tropical cyclones found [2]:

	No. of cases
1. Traces with no oscillations observable	143
2. Traces with only long-period oscillations or with long-period superimposed on short-period	79
3. Traces with only long-period oscillations (10–30 min)	21
4. Traces with doubtful oscillations, *i.e.*, doubtful due to blots and blurs	25
5. Traces which possess, almost surely, only oscillations of very small period	2

He concludes that, for the Philippines, short-period oscillations are the exception rather than the rule. This has also been considered true of the Atlantic. However, most barographs are designed to damp out minor short-term oscillations.

Deppermann found evidence of long-period oscillations in 100 out of 270 barograms with minima under 29.14 inches. He states that the simplest and most clear-cut cases resemble very much in form and period the oscillations on the microbarograph when there are thunderstorms quite near but not actually over the station.

No generally accepted explanation for these oscillations has been advanced. Short-period oscillations may be due to the pounding of the long hurricane swells on the beaches, strong rhythmic gusts of the wind, and the effect of gusts on buildings including the swaying of tall buildings. Long-period oscillations may be due to the regular march of intense squalls around the storm. In the most intense storms, violent pumping of the barometer may occur near and in the storm center. Violent kinks in the barogram are occasionally observed. In the July 1943 hurricane at Galveston, Texas, the Weather Bureau City Office barogram contained a kink just before and another just after the barometric minimum, but neither appeared on the barogram at the airport five miles away.

Temperature. At the surface no change in temperature is noted with the approach and passage of the tropical cyclone other than the lowering of the free air temperature to or near the dew-point or wet-bulb temperature incident to the heavy rain.

Wind Circulation. Winds in tropical cyclones of the Northern Hemisphere blow counterclockwise around, and incline inward toward, the center of the storm. However, the degree of incurvature toward the center is not uniform in the various quadrants. Cline noted that there is a systematic difference in the inclination of the winds in the four quadrants, and others have attempted to calculate the exact degree of incurvature. However, it appears that the incurvature varies considerably with individual storms and with their stage of development, size, latitude, and other factors. The winds in the right rear quadrant usually blow more directly toward the center (greatest incurvature) than is the case elsewhere. The intense storms of the immature stage have the most symmetrical circulation. While there is some lack of uniformity from one storm to another with regard to winds in the right front quadrant, they usually appear to move directly across the line along which the storm is moving and are tangential to the isobars. However, more definite information is needed on the surface circulation, the cross-isobar components, and the field of convergence.

If a hurricane is moving between west and northwest, the most commonly experienced wind along the line over which the central calm will move, preceding the calm, is north or north-northeast, usually the latter, and the wind may blow from this direction for hours, without veering or backing, prior to the arrival of the storm center. If the wind tends to veer or back, the center is unlikely to pass over the station. If the storm is moving in a northeasterly direction, the most usual wind preceding the arrival of the storm center is north-northeast to east-southeast.

The wind velocities encountered around the storm are affected by the stronger pressure gradient usually present to the right, since in westward-moving storms the semipermanent subtropical high is to the right and the intertropical trough to the left. Thus gales and hurricane winds extend farther to the right than to the left. In intense immature storms velocities increase until a few minutes before the calm, but in the mature and decaying stages the highest winds may occur as much as two hours before and one hour after the central calm. The dimensions of the hurricane winds will vary with a number of factors: maturity of the storm, pressure gradient, central pressure, etc. In the large Cape Verde storms, the diameter of hurricane winds may exceed 100 miles by the time the storm has reached the western Atlantic.

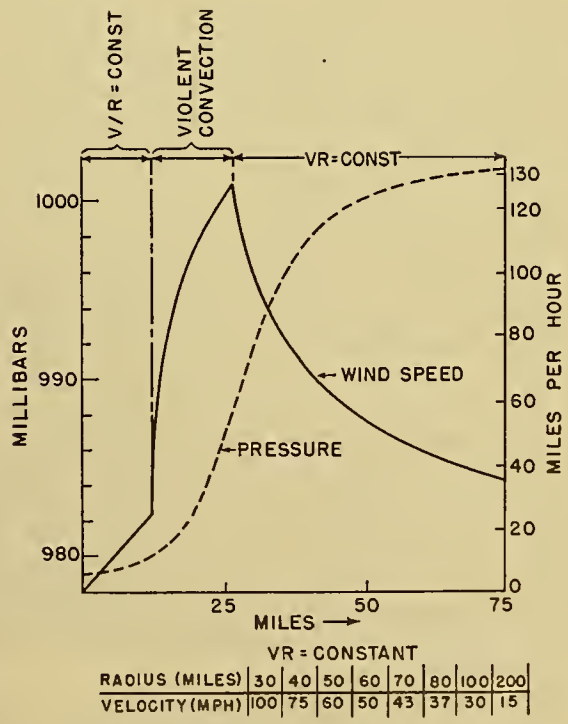


FIG. 1.—Deppermann's typhoon model.

In the Labor Day storm in 1935 over the Florida Keys, the most intense hurricane in the history of the United States, the diameter of hurricane winds was

only about thirty-five miles, therefore neither minimum barometer nor intensity of the storm is necessarily an indication of the area of hurricane winds.

Some writers have suggested various means of computing the distribution of surface wind velocities in a cyclone vortex by developing reasonable theoretical velocity profiles. Deppermann [4] has developed an idealized profile of the type shown in Fig. 1 as most typical of tropical cyclones. Here we have a Rankine vortex in which V/R = constant for the innermost cone including the eye and its adjacent areas, and VR = constant in the outer vortex. Deppermann describes the area between these two as a ring of violent convection characterized by a still different velocity variation. While this may serve as an average distribution for a large number of storms of various intensities, the range of deviations from this average will probably be quite considerable, both from quadrant to quadrant and from one stage of storm development to another. In the small immature storms maximum wind velocities will be reached closer to the eye. Since the required dense network of observing stations has never been available, no rigorous study of surface wind velocities and directions and associated pressure gradients has ever been made.

Clouds. Cloud arrangements, in common with other tropical-cyclone characteristics, will vary considerably with individual storms. As a rule the cloud sequence is much the same as in advance of a warm front in middle latitudes. First to appear are the cirrus clouds which frequently seem to radiate from a point on the horizon in the general direction of the storm. The cirrus will thicken to cirrostratus and then to altostratus and altocumulus. Soon occasional cumulonimbus, extending through the upper deck with passing squalls, will appear. Finally a dark wall of cloud, sometimes called the "bar" of the storm, appears, and squalls become almost continuous with stratus based at from 500 to 2500 ft or with a precipitation ceiling.

There is occasional reference in the literature to lurid sunsets preceding the advent of a hurricane. This condition may precede a hurricane but handsome sunsets are frequent in the tropics and subtropics where cirrus (nothus) formed from dissipated cumulonimbus are present.

In the hurricane vortex high clouds are rarely visible. Low clouds will have about the same direction as the surface wind although usually more tangential to the isobars. At sea, clouds are occasionally reported down to the water, but these reports are considered unreliable since rain and flying spray will frequently lower visibility to zero. On land, even at the height of the storm, the ceiling is usually reported as above 1000 ft, although in later stages of recurvature 500-ft ceilings are not infrequent.

Vines [19] believed that cirrus radiate from and move outward from the center of the tropical cyclone. Most of the observations during the past twenty years indicate that, in general, cirrus move forward in the general air stream in which the cyclone is imbedded. It does not necessarily follow that the hurricane circulation does not at times reach the cirrus level since cirrus cannot ordinarily be observed over the vortex. However, outside the periphery of low and middle cloud decks associated with the vortex, cirrus appear to be unaffected by the storm's circulation.

Precipitation. The distribution of rainfall around a tropical cyclone varies markedly from one storm to another. In general the more immature the storm and the lower the latitude, the more symmetrical the rain pattern from the standpoint of intensity and area. As the storm begins to recurve and reaches more northern latitudes, rainfall, from the standpoint of both intensity and area, becomes concentrated in the front quadrants. Notable exceptions, however, have occurred. In the "Yankee" hurricane of 1935, 0.04 inches of rain fell before the arrival of the center and 3.40 inches afterward. In October 1941 a hurricane passed over Miami, Dinner Key reporting a maximum wind of 123 mph, with only a trace of precipitation.

A check on three fairly recent Puerto Rican hurricanes reveals that at San Juan, 3.03 inches fell in the front half and 7.85 inches in the rear half of the storm of September 1926. This was a mature Cape Verde storm and the center passed to the south of the station. On September 27, 1932, a relatively immature hurricane passed over Rio Piedras seven miles southeast of San Juan, and 1.17 inches fell before and 1.59 inches after the center passed. On September 19, 1931, 0.71 inches fell before the calm center and 1.15 inches after it passed. This was a very young storm which had reached hurricane intensity less than twelve hours before it arrived over Rio Piedras. It is believed the distribution of rainfall around these three storms is normal for moderately low latitudes, although the total amount in the two later storms was subnormal.

Vines indicated that in Cuba rainfall extends farther in advance of the storm than to the rear, but there are also cases when the reverse is true. Most storms affecting Cuba have a strong northerly component and occur late in the season.

Deppermann has not reached very definite conclusions on the distribution of precipitation around tropical cyclones but states that most of the rain in the so-called "triple-point" typhoons in the Philippine area occurs in the rear and most of the rain in the "trade-norther" type in the right front quadrant.

Tropical cyclones in the Gulf of Mexico, according to Cline [1], have the greatest precipitation intensity 60–80 miles in front of the cyclone center and mostly to the right of the line along which the cyclone is advancing, and with relatively little precipitation in the rear half. Cline suggests that this is due to the winds of high velocities moving through the right rear quadrant and converging with winds in the right front quadrant of lesser velocity and greater cross-isobar components. He cites a rather extreme example, in the cyclone of October 10–12, 1909, when Key West received 12.04 inches before passage of the center and only 0.24 inches afterward. This storm had already recurved and was moving northeastward and, indeed, most of the storms cited by Cline were recurving and in stages 3 and 4.

The total rainfall in a tropical cyclone is dependent upon many factors, including rate of ascent of air in the storm area, location of the rain gage relative to the storm center, rate of movement of the storm, and especially topography. The rainfall averages 5–10 inches but varies greatly. Thirty to forty inches have occurred where topographical influences were favorable, but in other storms precipitation amounts have been less than an inch and in one instance only a trace was observed. Loss of rain from gages is considerable in high winds; in hurricane winds it may reach 50 per cent.

Some records of heavy rains in connection with tropical cyclones follow:

Baguio, Philippine Islands—46 inches in twenty-four hours, 88 inches in four days.

Silver Hill, Jamaica—96.5 inches in four days.

Mount Molloy, Queensland, Australia—63 inches in three days.

Réunion, Indian Ocean—47 inches in four days.

Adjuntas, Puerto Rico—29.60 inches in little over twenty-four hours.

Texas, several instances—22 and 23 inches in twenty-four hours.

In most of the foregoing cases orographic influences were pronounced. As a rule a station will not remain under the direct influence of a tropical cyclone more than two days; therefore, it is doubtful if the tropical storm was entirely responsible for the total rainfall in all of these cases. In some storms much more rain fell during the stage of advanced decay than during the period of greatest intensity.

Radar Bands. In recent years radar has been used during aerial reconnaissance to track tropical cyclones and to locate their centers. Many of the radar photographs have revealed the presence of large-scale squall or convection zones spiraling inward toward the center. A rather typical example is shown in Fig. 2. This photograph was taken at Orlando, Florida, at 0220 EST on September 16, 1945, when the storm center was some 136 miles south of the station. The quasi-circular heavy precipitation bands are clearly outlined. The lack of such bands in the southwest quadrant is real and probably due to the decreased convergence normally found in that section of the storm in this particular stage of development. Wexler [22] concluded from a study of this hurricane and of two typhoons in the Pacific that, in these three storms at least, the precipitation bands were symmetrical while the storm was over the ocean but rapidly developed asymmetry over land. Friction over land develops a deformation in the pressure and wind fields which also changes the distribution of convergence. Between these bands or zones there is a thick altostratus or nimbostratus sheet from which light rain falls continuously.

The exact mechanism of these bands has not yet been fully explained. Haurwitz [9] suggests that internal wave motion occurring when vertical wind-shear is present may lead to patterns very similar to the convection patterns observed in the laboratory. If these convection patterns are introduced into a circular vortex attended by general convergence, then the bands will spiral inward toward the center. Their width and spacing, according to Wexler, will change depending on the vertical wind and density gradients and the

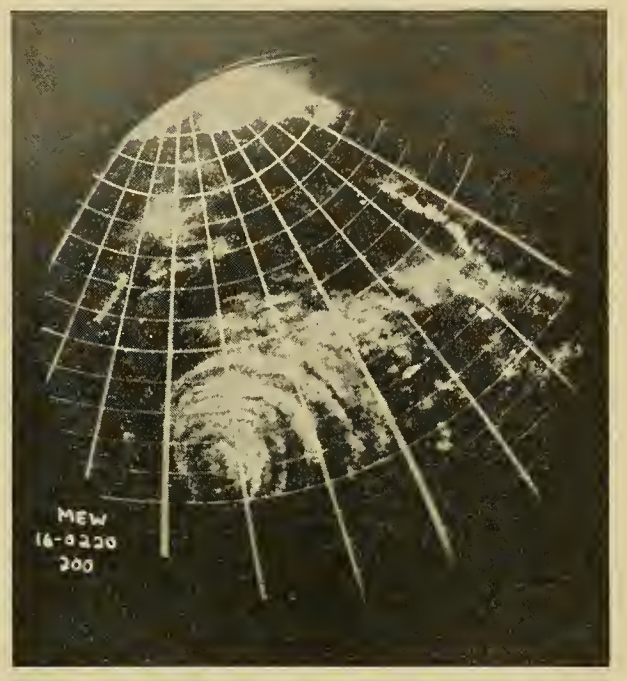


Fig. 2.—Radarscope photograph, Orlando, Florida, 0220 EST, September 16, 1945. Hurricane center 136 miles from station. (*Courtesy H. Wexler and U. S. Army Air Forces.*)

distribution and magnitude of the horizontal convergence associated with the vortex.

The "Eye" of the Storm. One of the most spectacular aspects of the tropical cyclone is the central calm or "eye" of the storm. At the center of the storm the wind rather suddenly diminishes from extreme violence to 15 mph or less. At the same time the rain ceases, the low clouds may be visible only on the horizon, and the middle deck becomes broken and often scattered. Cirrus and cirrostratus clouds are almost always present. In the middle of the "eye," winds may lessen to 5 mph and dead calms have been reported for short periods, and the sun or stars may be visible. Observers have described conditions in the relatively calm center as "oppressive," "sultry," and "suffocating," but this reaction is apparently psychological and due to the rapid transition from hurricane winds and torrential rain to relatively calm and humid conditions. The few cases of noticeable rise in the temperature at the surface and consequent changes in humidity have been explained by insolation or foehn effect.

Deppermann [2] has tabulated the duration of calms for some 59 typhoons as follows:

Central pressure	No. of cases	Mean duration of calm
Below 27.56 in. (933.3 mb)	4	18 min
27.56–27.95 in. (946.5 mb)	7	16 min
27.95–28.35 in. (960.0 mb)	5	37 min
28.35–28.74 in. (973.2 mb)	11	32 min
28.74–29.14 in. (986.8 mb)	15	39 min
29.14–29.53 in. (1000.0 mb)	17	64 min

These figures would indicate that weak storms have longer calms than more intense storms. However, it does not appear that Deppermann considered rate of movement, or whether merely a chord near the edge of the calm center instead of the diameter passed over the station. Since the rate of movement varies from one storm to another, the duration of the calm is not necessarily a measure of the diameter of the calm center. In the United States the diameter appears to be less in immature storms than in the larger mature storms. In the latter the calm center seems to increase in about the same proportion as the storm area. The smallest eye reported has been about 4 miles in diameter in a storm in the formative stage; diameters of 20–25 miles in large mature storms are not unusual. An average value appears to be 12–15 miles. Upon recurvature the eye may assume a very elongated shape in the direction in which the storm is moving. On occasion a double eye has been observed, usually when the cyclone is decaying.

Frequency of Thunderstorms and Tornadoes. Thunderstorms are frequently observed in the early and decaying stages of tropical cyclones and also around the periphery of the storm. They are not usually observed in the area of winds over 60 mph in immature and mature storms. Tornadoes have been reported in a few instances in connection with tropical storms in the Bahamas and in Cuba and fairly frequently in Florida. The tornado paths are very short (only a mile or two in length) and several appear to have formed over the ocean as waterspouts. Indeed, from the standpoint of intensity, they seem to resemble waterspouts more closely than inland tornadoes. Information is not available as to whether any particular quadrant is preferred.

Inundations. The *hurricane wave*, sometimes erroneously called a tidal wave, is the lifting up of the level of the sea at and near the center of intense hurricanes. It is usually not a series of individual waves but a rapid uplift of water and is most noticeable in the calm center. It has been noted in all areas where tropical storms occur.

Hurricane waves occur on the shores of small islands as readily as on continental coast lines. They have been responsible for heavy losses of life and precautions should always be taken against their possible occurrence, especially at and for a short distance to the right of the point where the center will cross the coast line. The hurricane wave is superimposed on the gravitational and hurricane tides prevailing at the time but will usually subside quickly within an hour or two after the center passes.

The wave has been ascribed to the decrease in atmospheric pressure, but even in the most severe hurricanes the maximum possible rise from decreased pressure has been calculated to be less than 4 ft and any such rise should be gradual. It is thought that the wave is caused by the damming effect of the wind on the forward side of the calm center against the current set up by the violent winds from the opposite direction to the rear of the calm center. The average height of the hurricane wave is 10–20 ft but higher waves have been reported, especially where the shore line permits a funneling effect.

The Hurricane Tide. There is little or no actual transfer of water in waves and swells, but sustained winds of only moderate force will set up a current flowing with the wind. The rise in the water level on the shore line may begin when the tropical cyclone is 500 miles or more distant and will continue until the storm passes inland or beyond the area. This tidal rise is called the hurricane tide and is superimposed on the normal gravitational tide. It is most pronounced in a partially enclosed body of water such as the Gulf of Mexico, where the concave coast line does not readily permit the escape of water which is thus piled up on the shore line. The strongest winds and the greatest fetch is in the right rear quadrant; therefore, the strongest current is directed with and to the right of the line along which the storm is moving.

Heights of storm tides on concave coast lines vary from 3 to 10 ft above normal. On straight shore lines the average is less and in very small island groups, around which the current may easily flow, there is little or no tidal effect of this type.

VERTICAL STRUCTURE OF TROPICAL CYCLONES[2]

Theory. Observations necessary for proof of the various theories in regard to the vertical structure of tropical cyclones are extremely sparse. Nevertheless, through contributions from Shaw, Durst and Sutcliffe, Willett, Haurwitz, Simpson, Riehl, and many others, a fairly coherent and logical theory has been developed which has found general acceptance in a broad sense but in many details is subject to further verification and, no doubt, to revision on the basis of future observational data.

The tropical cyclone is definitely a warm-core phenomenon, in contrast to all other tropical disturbances which are cold-core phenomena. The circulation is strongest near the surface, but as the storm intensifies it extends rapidly to great heights. With the approach of the vortex, temperatures and specific humidities increase at almost all levels at least up to 15–20 km. It further appears that convergence near tropical-cyclone centers is limited to a fairly shallow layer near the surface.

The problem of how evacuation of air from tropical cyclones takes place is the most controversial portion of the theory of tropical-cyclone structure. The theory of many earlier meteorologists, and even of some who are currently engaged in tropical research (see [15]), that the pressure gradient reverses its sign at some level above 10–20 km, is still inconclusive. Vines' observations in the 1890's of cirrus radiating outward in all directions from a point over the storm center have not been reaffirmed for many years. Riehl [12], while believing that a reversal of pressure gradient is

2. This phase of tropical cyclones is discussed in greater detail in "Aerology of Tropical Storms" by H. Riehl, pp. 902–913 in this volume.

likely at some altitude high above the storm center, has summarized how considerable outflow may take place otherwise. Following Durst and Sutcliffe [6] he has considered the effect of angular momentum of rising surface air. Because of the decrease of pressure-gradient force with height, the rising air seeks to draw away from the center as soon as the centrifugal force exceeds the pressure-gradient force. Values indicating the variations of the pressure-gradient force with height are tabulated below. The values in the right-hand column are the ratios (expressed in per cent) of the pressure-gradient force at different elevations to that at the surface.

H(km)	Per cent	H(km)	Per cent
Sfc	100	10	3
1	98	11	−17
2	92	12	−31
3	90	13	−35
4	80	14	−60
5	70	15	−84
6	63	16	−59
7	54	17	−131
8	28	18	−142
9	25		

(*After Riehl*)

If it is assumed that the ascending current retains its angular momentum, withdrawal from the central area would start approximately at the 3-km level, according to Riehl. At this elevation the magnitude of the pressure-gradient force directed toward the hurricane center begins to drop significantly with height. In reality, outflow will begin at even lower heights, owing to the decrease of the frictional force above the surface.

Durst and Sutcliffe also postulate an increase of the vertical velocity with height and show that such an increase may contribute to the radial outflow of the air aloft. In view of the probability that convergence is restricted to the vicinity of the surface, Riehl concludes it is unlikely that the upward vertical motion increases with height except near the surface.

Height of Tropical Cyclones. Tropical cyclones are associated with stronger pressure gradients than are observed in any other weather phenomenon where we must deal with hydrostatic equilibrium. The equalization of this pressure difference between the center and the periphery of the storm at some point in the upper air is dependent primarily upon a thermal gradient directed so that the warmest air is in the core of the storm. Haurwitz [10] has shown impressively how high this temperature must be if the circulations of very small storms are to disappear at levels at or below twenty thousand feet. For example, in a relatively moderate tropical cyclone with a pressure difference of only 40 mb from the center to the outer edge of the surface vortex, and assuming a reasonable lapse rate in the outer vortex, he found that the temperature in the eye must increase with elevation at the rate of 18C km^{-1} if the circulation of the storm is to disappear at ten thousand feet. The lapse rate must be zero, or isothermal, if the circulation is to disappear at twenty thousand feet. For the circulation to disappear at thirty thousand feet the lapse rate must be only slightly more than 4.5C km^{-1}. The two storms for which soundings in the eye are available had average lapse rates of not less than 4C km^{-1}. Since the more severe hurricanes are not infrequently associated with pressure differences of 70–90 mb between the center and the periphery, it can readily be seen that such storms must extend high into the upper troposphere.

Figure 3 shows a sounding in the eye made at Tampa, Florida, on October 18–19, 1944, and two

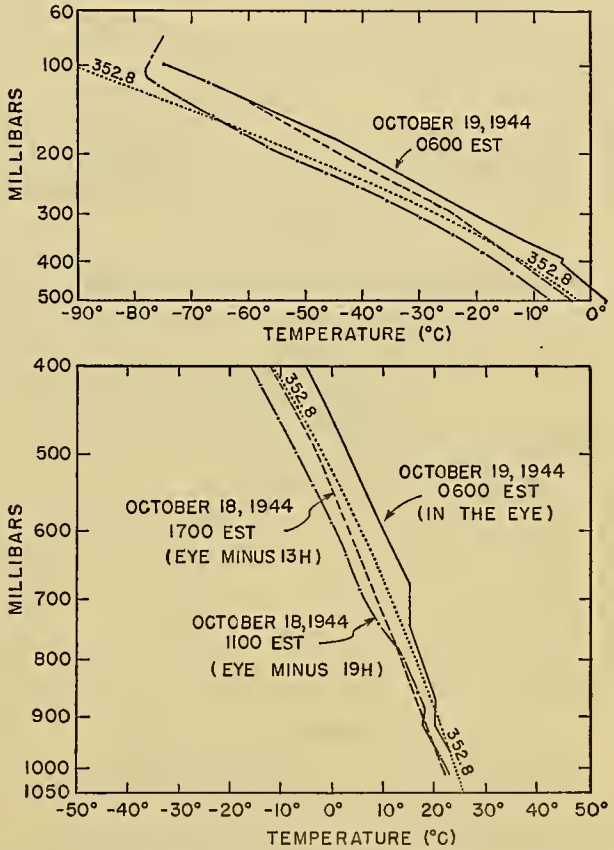


FIG. 3.—Soundings at Tampa, Florida, during the approach and arrival of a hurricane center, October 18–19, 1944.

soundings made earlier. Figure 4 presents a similar series for October 7–8, 1946. Both series represent conditions during the approach of a hurricane, and the warmth of the eye is clearly indicated. It should be noted, however, that both storms were in a state of recurvature and it is not certain that the radiosonde did not pass out of the eye.

Thermal Structure. There is as yet little actual data to verify theoretical models of temperature distribution in tropical cyclones above the surface. A number of pressure-anomaly cross-section charts by Simpson have indicated that these storms extend to very high levels and have a double cellular structure with the lower warm-core cell capped by an upper cold-core cell.

Tropopause Heights in Tropical Cyclones. Until recently it was generally thought that low tropopauses accompanied tropical cyclones. The reasons most commonly advanced for the low tropopause were the supposed absence of convection in the eye and the stronger

temperature gradients in the upper troposphere required to reverse the circulation. Of the two soundings available from centers of tropical cyclones, one reached 17½ km and the other 17 km; neither showed much evidence of having reached the tropopause, although the sounding in the October 1944 storm showed extraordinary stability in the middle troposphere. Figure 5 shows probable tropopause variation in three tropical cyclones, and in Fig. 6 Simpson indicates the probable relation between temperature in the outer and inner vortex of a mature hurricane.

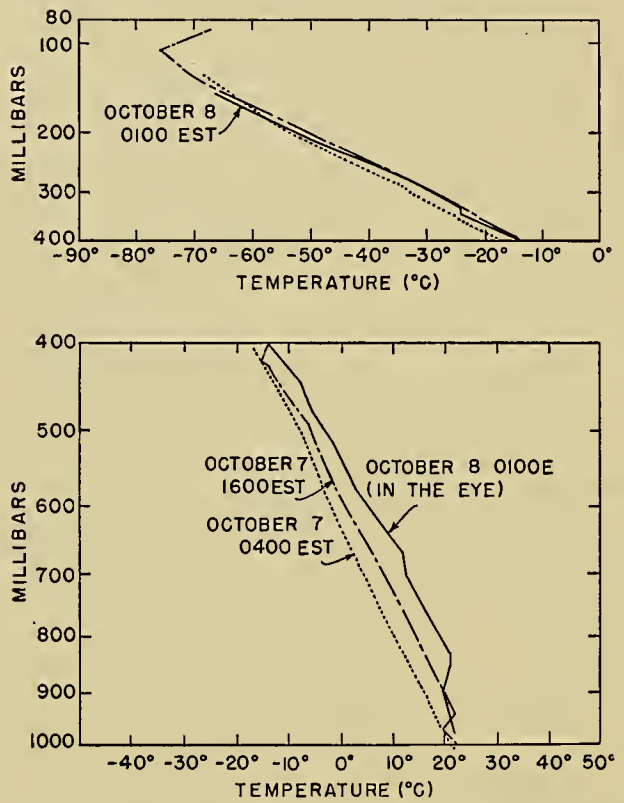


Fig. 4.—Soundings at Tampa, Florida, during the approach and arrival of a hurricane center, October 7–8, 1946.

ORIGIN OF TROPICAL CYCLONES

Theories of Tropical Cyclone Development. The convective theory of the formation of tropical cyclones was developed many years ago and was generally accepted until the early 1930's. It was believed that the following series of events occurred in the equatorial trough: The warm moist surface air rose because of both insolation and widespread convergence, resulting in numerous cumulonimbus clouds and widespread heavy showers. Then the pressure began falling slowly (for reasons never very well explained), the cumulonimbus clouds gradually coalesced, and if the equatorial trough was sufficiently far from the equator for the Coriolis force to be effective, a cyclonic circulation was initiated. The development of the circulation continued until full intensity was obtained by release of latent heat through condensation. In some areas, notably the North Atlantic, it is now apparent that many, if not the majority, of the tropical cyclones do not develop in the equatorial trough.

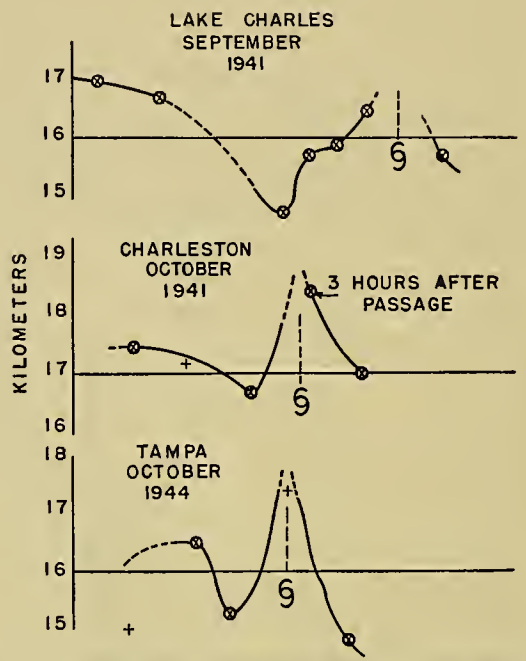


Fig. 5.—Soundings at selected stations during the near-approach or arrival of a hurricane center. (⊗—height of tropopause; +—top of sounding, tropopause not reached.) (*Hurricane Notes—Weather Bureau Training Paper No. 1, Washington, D. C., July 1948.*)

Some attempt was made in the middle 1930's to apply Norwegian methods to tropical analysis. Hurricanes were supposed to develop on the equatorial front which at that time had several different definitions but which now is considered as a narrow zone, usually associated with the equatorial trough, in which air is undergoing horizontal convergence. However, there are no significant temperature differences along the equatorial front, there is no density discontinuity, and the air is essentially homogeneous.

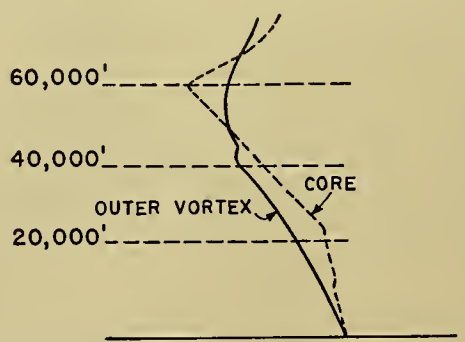


Fig. 6.—Probable relation between temperature in the outer and inner vortex of a mature hurricane. (*After R. H. Simpson.*)

Tropical cyclones originate in easterly waves, in the intertropical convergence zone, and occasionally in the trailing southerly portions of old polar troughs. Cer-

tainly it appears that tropical cyclones will develop only over comparatively warm water, probably 82F or higher. In 1890 only one tropical storm was noted in the North Atlantic and in several other years the total was only two. Therefore, it is apparent that a very special set of circumstances is necessary for their development. There is as yet no generally accepted definition of exactly what synoptic situation is responsible for the formation of a tropical cyclone.

In the North Atlantic, Riehl [12] found that intensification of low-level tropical perturbations or a change from a stable to an unstable condition resulted from the superposition of high-level extratropical disturbances. In all cases of hurricane formation noted in the course of this study, deepening began, without exception, in pre-existing tropical disturbances.

Cape Verde Island region—August and September. (It may be that most of these storms do not reach hurricane intensity until they reach longitude 50–60°W.)

Northern Caribbean Sea—late May through November.

Just to the east and north of the West Indies—June through October.

Southwestern Caribbean Sea—principally June and October.

Gulf of Mexico—June through October.

2. *North Pacific Ocean off the west coast of Mexico*—June through November.

3. *North Pacific Ocean, longitude 170°E westward including the Philippines and the China Sea*—May through December and (rarely) other months.

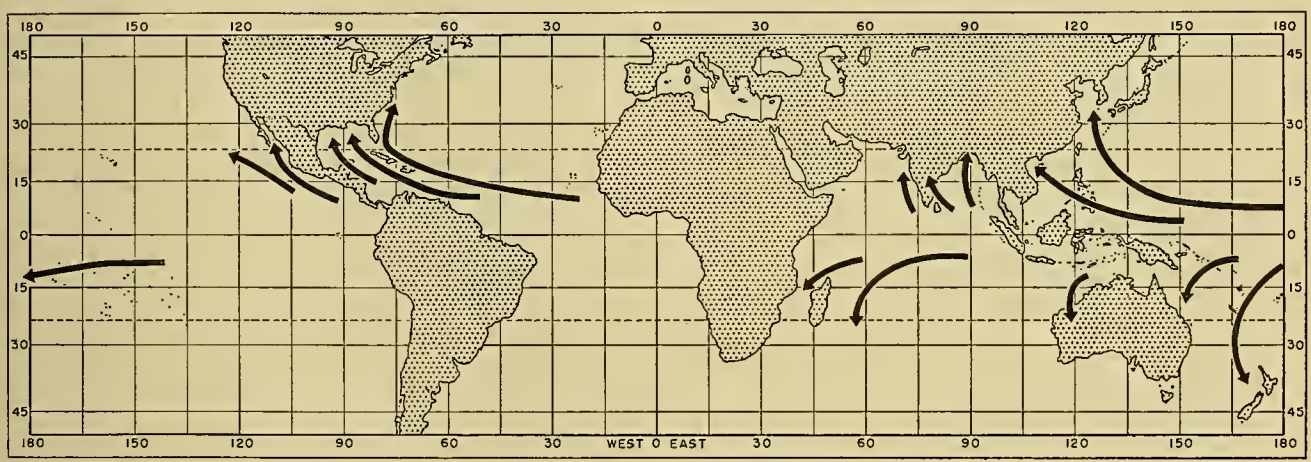


Fig. 7.—Areas of tropical cyclone development.

Tropical cyclones in the southwestern North Pacific Ocean for the year 1945, and to some extent for 1946 and 1947, have been analyzed by Riehl [13] and several other members of the Institute of Meteorology, University of Chicago. Riehl found that typhoon development can begin in consequence of instability of the Northern Hemisphere trades, not necessarily as a result of interaction between currents of the two hemispheres. Instability of the trades appeared to set in just as a low-level wave was bypassed by a broad-scale ridge in the high troposphere. Riehl suggests that a tropical disturbance will intensify when moving from an upper trough to a high-level ridge, since this pressure distribution facilitates upper outflow from the storm. So far observational data have been insufficient to test these relationships in the tropical North Atlantic.

An intensification of the observational network which will permit accurate trough and ridge analyses between the equator and latitude 30° during the tropical cyclone season is badly needed for further research on the interrelationship between upper troposphere troughs and ridges and the development of tropical cyclones.

Region of Origin and Season. Tropical cyclones develop in the following areas and seasons:

1. *South portion of the North Atlantic Ocean.*

4. *North Indian Ocean.*

Bay of Bengal—April through December.

Arabian Sea—April through June, September through December.

5. *South Indian Ocean.*

Madagascar eastward to 90°E—November through May.

Off northwest coast of Australia—November through April.

6. *South Pacific Ocean from east of Australia eastward to 140°W*—December through April.

Probably no portion of the tropical oceanic area in either hemisphere is entirely free from tropical storms, but there is no record of a true tropical cyclone of hurricane intensity in the South Atlantic Ocean or in the South Pacific east of about longitude 140°W. During the Southern Hemisphere summer, the intertropical front in these areas moves only a degree or so south of the equator, not far enough for the Coriolis force to become effective. Few, if any, tropical cyclones have been observed within 5 degrees of the equator. In Fig. 7 are shown the principal areas of the world where tropical cyclones develop and are observed.

Frequencies of Tropical Cyclones. Data on the frequencies of tropical cyclones by months are summarized in Table I.

Detection of Tropical Cyclones. Since tropical cyclones develop over water surfaces, where observational data are often sparse and occasionally nonexistent, detection is a primary problem. Ship and island observations provide information on surface conditions, and a few pilot-balloon and radiosonde reports are available. In some areas pilot reports from aircraft provide limited and often none too accurate upper-air information. To detect tropical disturbances in the early stages, a careful analysis of all observational information is necessary, but it is still possible in some areas for tropical storms to travel a number of days and hundreds of miles over water without detection.

Surface Indications of a Tropical Storm. Within the trade-wind area winds blow with great steadiness with respect to both direction and velocity. Over the ocean area the trade winds will blow from northeast to southeast at 15–22 mph. If a ship within the trades reports a wind with a westerly component, a tropical disturbance has formed; the stronger the westerly wind the more intense is the storm. If the easterly trade-wind velocities are 25 per cent or more above normal, a disturbed situation has developed.

Table I. Frequencies of Tropical Cyclones by Months

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Year
North Atlantic Ocean (1887–1948)													
All tropical cyclones	0	0	0	0	.1	.4	.6	1.6	2.4	1.8	.4	*	7.3
Hurricane intensity	0	0	0	0	0	2	2	1.0	1.3	.7	.1	0	3.5
North Pacific Ocean—off west coast of Mexico (1910–1940)													
All tropical cyclones	*	*	*	*	.1	.8	.7	1.0	1.9	1.0	.1	0	5.7
Definitely hurricane intensity	*	*	*	*	.1	.2	.2	.5	.7	.5	0	0	2.2
North Pacific Ocean—long. 170°E westward (1901–1940)													
All tropical cyclones†	.4	.2	.3	.4	.7	1.0	3.2	4.2	4.6	3.2	1.7	1.2	21.1
North Indian Ocean—Bay of Bengal (Dates not known)													
All tropical cyclones‡	.1	0	.2	.2	.5	.6	.8	.6	.7	.9	1.0	.4	6.0
North Indian Ocean—Arabian Sea (Dates not known)													
All tropical cyclones‡	.1	0	0	.1	.2	.3	.1	0	.1	.2	.3	.1	1.5
South Indian Ocean—Madagascar eastward to 90°E (Dates not known—possibly 1839–1922)													
All tropical cyclones§	1.3	1.7	1.2	.6	.2	0	0	0	0	.1	.2	.8	6.1
South Indian Ocean—Northwest Australia, excluding Queensland (Dates not known—possibly 1839–1922)													
All tropical cyclones§	.3	.2	.2	.1	0	0	0	0	0	0	0	.1	.9
South Pacific Ocean	According to Visher [20], the average number of tropical cyclones is in excess of 27 per year. However, this includes several more or less separate areas of development. A number of island groups such as Fiji, Samoa, Hebrides, Tongo, and New Caledonia average 2 to 3 per year. Data are insufficient for monthly percentages, but the tropical cyclones are apparently well concentrated in the period December through March.												

* Less than 0.1.

† Very few of the storms listed in the winter months reach hurricane intensity. The proportion of storms in other months reaching full hurricane intensity is unknown.

‡ Few of the midwinter and midsummer storms reach full hurricane intensity, and the midwinter storms may not be true tropical cyclones.

§ The intensities of these storms are not indicated.

Pressure fluctuates within narrow limits in the tropics, with the 24-hr net pressure change usually less than the diurnal variati n. The existence of a tropical storm may be recognized by a 24-hr fall of from 3 to 3.5 mb or more, or by a fall in pressure to 5 mb or more below normal.

Precipitation normally occurs in the form of scattered showers and occasional thunderstorms, with variable middle cloudiness from none to broken. If unusually

heavy or steady rain is reported from several nearby points, a developing or approaching tropical storm may be suspected. Solid cirrostratus or altostratus also indicates a disturbed condition.

In the North Atlantic Ocean, tropical storms do not as a rule develop in the equatorial convergence zone. The only exceptions may be those developing off the coast of Africa (the Cape Verde hurricanes) and in the southwestern Caribbean, north of Panama. In the Pacific Ocean it appears that some do develop along the equatorial front. Along this front a low-pressure envelope with a longitudinal extent of 5–25 degrees may appear. It will contain several depressions which drift irregularly westward until a polar trough aloft or major ridge aloft (high troposphere) is encountered, when deepening of one of the depressions may result.

Trade winds of relatively constant direction and speed set up swells which may travel for several thousand miles. The normal frequency of swells in the trade region is 8 per minute over the Atlantic and 14 per minute in the Gulf of Mexico. Hurricane winds set up swells with a greater wave length and periodicity, and in severe storms the frequency decreases to 4 per minute. Swells move outward in all directions from the storm center and will approach the observer, in areas where the swells remain unmodified by large islands and irregular coast lines, from the approximate location of the storm center. Therefore, abnormally high swells or an abnormal direction of movement will indicate the existence of a tropical storm. The height and periodicity of the swells indicate the intensity of the storm. The usefulness of swells in forecasting the movement of tropical cyclones is discussed in the next section.

Modern technique in the detection of tropical storms involves the dispatch of reconnaissance planes into the area where any observations may have indicated the possibility of the existence of a tropical storm.

Upper-Air Indications of a Tropical Storm. In the very early stages of development, tropical storms are occasionally better developed between 10,000 ft and 20,000 ft than at the surface. Therefore, streamline analysis is recommended for pibal charts at these levels, or a very careful analysis of the 700-mb and 500-mb constant-pressure charts if sufficient data are available for their preparation. Perturbations in the normal easterly flow should be carefully watched.

FORECASTING MOVEMENT OF TROPICAL CYCLONES

Monthly Variation in Normal Storm Track. Tropical storm development is a seasonal phenomenon wherever it occurs, embracing principally the summer and fall months in each hemisphere. In most, if not all, of the tropical-cyclone-susceptible areas the mean storm track, and to a lesser extent the place of most frequent origin, changes from month to month during the storm season. In Fig. 8 the change from month to month in the mean storm track is evident. However, from year to year the individual storm tracks scatter considerably.

May is an unimportant hurricane month in the tropical North Atlantic, with a tropical storm occurring once in only about every ten years. It is not included in Fig. 8. The official record includes no tropical storm of full hurricane intensity in May although at least

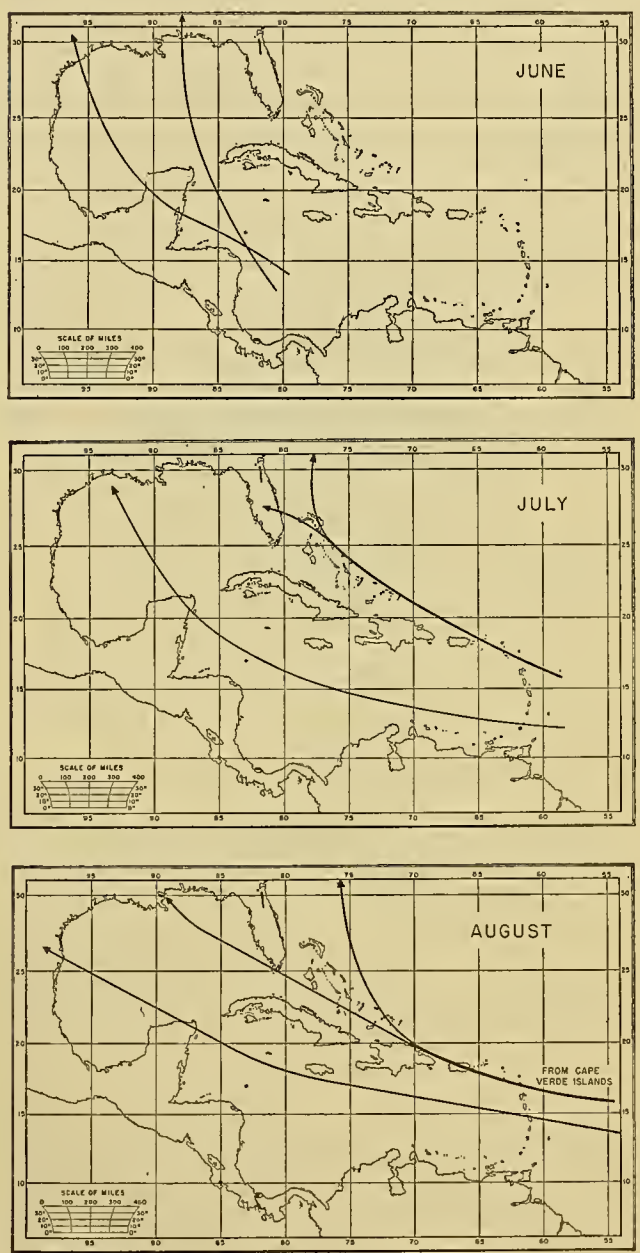


FIG. 8.—Mean track of tropical cyclones, southwestern North Atlantic Ocean, June–October. (Continued on p. 898.)

one was attended by winds of more than 60 mph for a day or more. All storms developed in the Caribbean or near Florida with one exception, the most severe, which developed to the east of Trinidad.

June storms are rather small in area but may be quite intense. A tropical storm will develop on the average every other year but one will reach hurricane intensity only once every five years. Such storms usually develop in the extreme western Caribbean or

in the Gulf of Mexico and have a strong northerly component.

Tropical storms in July are only slightly more frequent than in June and are slightly larger, but none of the outstanding hurricanes have occurred in this month. Most develop just to the north or to the south of the Antilles and have a more westerly component.

A marked increase in the frequency and intensity of tropical cyclones takes place in August and some of the most severe of record have occurred in this month. Many form in the Cape Verde region and travel across the entire Atlantic. August storms have strong westerly components and the most regular paths of any month.

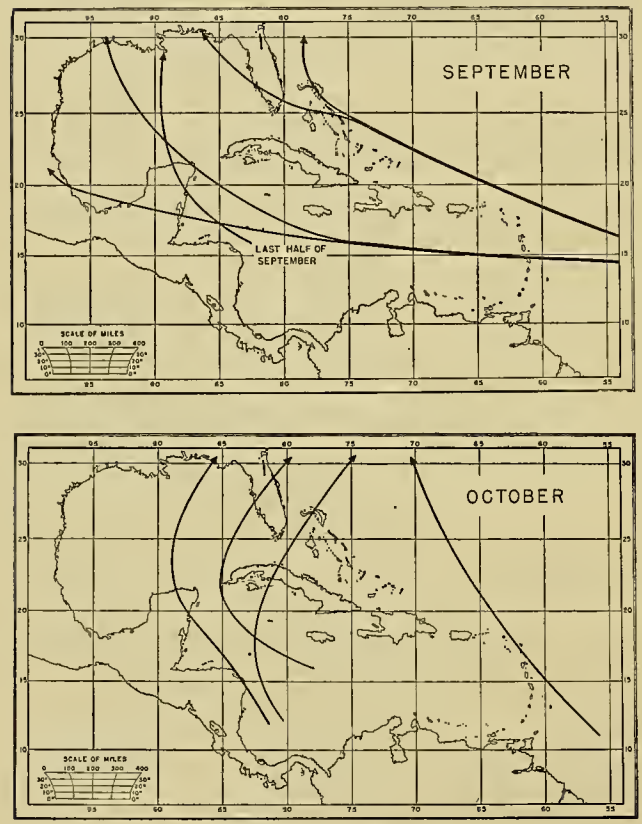


FIG. 8.—Continued

The height of the hurricane season is reached in the first half of September. During this period many still develop in the Cape Verde region. During the last half of the month the area of most frequent origin shifts back to the southwestern Atlantic and storms again develop strong northerly components as the polar westerlies move southward.

October continues to have a large number of tropical storms but the frequency and average intensity decline rapidly after mid-October. All have strong northerly components and recurve quickly. Many develop in the western Caribbean north of Panama and move northward across Cuba. Rarely do they reach as far west as Texas.

November is an unimportant hurricane month. The few storms mostly develop in the Carribbean and move northward immediately.

The stronger northerly components in the early and late portions of the tropical cyclone season and the more westerly component during the middle of the season are characteristic of all areas with the possible exception of the Bay of Bengal and the Arabian Sea. Of course, many exceptions to the mean tracks occur, depending upon the mean position of polar troughs which influence recurvature.

Steering Currents. The direction and rate of movement of a tropical cyclone are dependent upon the direction and speed of the air in which the storm is imbedded. There is no physical basis for believing the current at any single level is responsible for steering a storm, but Norton, Riehl, and others have found that the winds at or just above the top of the warm lower cell of the storm, where the vortical circulation virtually disappears, best approximate the direction of movement of the storm. This level will vary in individual storms from 20,000 ft to 30,000 ft or higher. It has been noted that most storms appear to have some deflection to the right of the steering current. The angle varies with the speed of movement of the storm, being as much as 20 degrees with speeds under 20 mph.

At times the pressure situation may be changing rapidly at the steering level and the steering current may change materially with time. Wide-scale changes of quarter-hemisphere dimensions may be taking place with resulting strong anticyclogenesis upstream, as in the Atlantic hurricane of September 17, 1947. Such a development may, in turn, alter the steering current over a large area within a 24-hr period.

Hurricane forecasters in the United States believe tropical cyclones usually move at a rate of 60–80 per cent of the velocity of the winds at the steering level, and that winds ahead of the storm are more representative than those to the rear in computing its movement. Some forecasters have indicated that the relationship between rate of movement and the 20,000-ft winds ahead of the storm is always better than the 70 per cent value.

Simpson [16] has developed a technique which he called "warm-tongue steering" for forecasting the direction of movement of tropical cyclones. This method gave good results when tested on 25 storms during the period 1941–45. An analysis of the mean virtual-temperature field between the 500- and 700-mb pressure surfaces indicated that a tongue of warmer lighter air is associated with and extends some 800 to 1200 miles in advance of the tropical cyclone. The tests seemed to indicate that a good lag-correlation exists (for 24 hr and often 48 hr) between the orientation of the warm tongue and the future movement of the storm. This correlation was least effective in small immature storms. A theoretical explanation of this relationship is not apparent. It is understood that unpublished tests by other groups fail to show substantial agreement with Simpson's results.

The problem of recurvature is the most difficult one facing the forecaster, since it takes place frequently within 24-hr striking distance of the coast line and over

oceanic areas where upper-air data are lacking. Satisfactory forecasting of recurvature is impossible without accurate constant-pressure charts. Riehl and Shafer [14] studied some 57 tropical cyclones occurring between 1935 and 1943 and classified them in groups as follows: (1) seventeen which curved north or northeastward, (2) twenty-three which continued between west and north without recurvature, and (3) seventeen which had irregular tracks, particularly motion with a southward component for a portion of the track. It should be noted that the "abnormal" storm tracks of the last group were as frequent as the "normal" tracks of the first group.

Riehl and Shafer found that a tropical cyclone will begin to react when the forward edge of the westerlies—at least as low as 8000–14,000 ft—is found 500–700 miles to the west. As the polar trough advances eastward, the storm turns from a westward to a northward path. The most difficult, though frequent, cases occurred when storms encountered strong troughs and an intrusion of the polar westerlies into the tropics similar to group (1). In that group, however, the base of the westerlies not only lowered but remained low in the subtropics and part of the tropics. In group (2), however, the westerlies died away quickly to the rear of the polar troughs and the easterlies were re-established quickly and storms continued in a westerly direction. This condition developed when a warm high follows or develops to the rear of a polar trough and also if the polar trough advances from the west against a blocking high resulting in fracture. The tropical cyclone usually responds within 24 hr after the easterlies build back to 14,000 ft and above. Group (3), the "abnormal" tracks, result from a number of causes. A storm which has already completely or partially recurved may turn back northwestward or westward when blocked by a warm high (October 13–16, 1938). A storm recurving up a polar trough may be overtaken by the trough, become imbedded in a northwesterly current, and turn southeastward (October 8–9, 1941). Storms may move southwestward when they are overtaken by a narrow trough and encounter a deep northeasterly current (November 2–5, 1935). Even at the present time sufficient upper-air information is not available to forecast or explain 5–10 per cent of the "abnormal" movement.

Forecasting direction of movement, particularly recurvature, is the most acute problem facing the forecaster. Previous checks on the correlation between the steering current and actual rate and direction of movements of tropical storms have not been completely satisfactory and, among hurricane forecasters, any further research on this problem should have the highest priority.

Forecasting Significance of Storm Swells. The appearance of a swell of a particular type may give quite reliable indications of a tropical storm as much as 500 to 1000 miles or more distant.[3] The height of the waves from which swells develop is determined by the force of the wind and the "fetch" or water distance over which the wind has blown without significant deviation in direction. According to Thomas Stevenson [17], the maximum height of a wave as a function of "fetch" is $H = 1.5\sqrt{F}$, where H is in feet, and F, the fetch, is in nautical miles.

The magnitude of waves is dependent not only upon the fetch, but also upon the wind velocity. Over oceanic areas with 600–1000 miles or more of sea room, waves 35–40 ft high are developed in ordinary storms and in more intense storms may exceed 45 ft. According to Cline [1], the quotient obtained by dividing the wind velocity (probably average for one hour) in miles per hour by 2.05 represents the average height in feet of waves developed by the wind. This should be used with caution and only as an approximation, since there are always other factors to be taken into consideration, and a wind, constant in speed and direction (in a hurricane at least), does not act on a wave for any great length of time. The breaking wave or swell is one of the most destructive elements of a tropical cyclone, since a cubic yard of water weighs about 1500 pounds and waves moving forward many feet per second may be very destructive to beaches and harbor facilities, especially when they contain debris such as tree trunks and heavy beams.

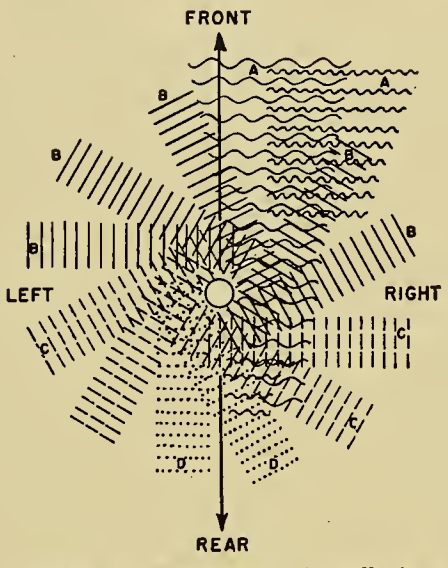


Fig. 9.—Schematic development of swells in a tropical cyclone:
A—swells of greatest length and magnitude, traveling in the line of advance of the tropical cyclone.
B—swells and waves of moderate length and magnitude in the front segment moving outward to the right and left of the line of advance.
C—swells and waves of smaller length and lesser magnitude in the rear segment moving outward to the right and left of the line of advance.
D—swells and waves of least magnitude moving outward from the rear of the hurricane. (*After Tannehill.*)

The generation of swells in a tropical cyclone is a rather complex phenomenon because of the curving wind flow. Therefore, since the wind-generated waves move in the same direction as the wind, swells tend to emanate in all directions from the center of the hur-

3. Consult "Ocean Waves as a Meteorological Tool" by W. H. Munk, pp. 1090–1100 in this volume.

ricane. Waves generated in the right half of the storm travel *with* the storm, so they are under the influence of winds with relatively little deviation in direction for a longer time, *i.e.*, they have a comparatively long fetch. Thus the strongest swells, which move through the storm and eventually a long distance ahead of the storm, are produced in this half. Conversely, swells traveling opposite to the direction of movement of the storm will be weaker. (See Fig. 9.)

It has been stated earlier that swells emanating from the region of a tropical cyclone are useful in its early detection and location and in the determination of its intensity. Swells also have further forecast value. If the direction of movement of the swells remains the same, the storm is either approaching directly or going away from the observer. If the direction changes counterclockwise, the storm is passing from the observer's right to his left; if it changes clockwise, it has passed or is passing from his left to his right, as he faces the swell. The direction from which the swell approaches indicates the direction of the storm from the observer at the time the swell was generated.

Forecasting Significance of Storm Tides. Particularly along the shore lines of partially enclosed bodies of water, such as the Gulf of Mexico, tides have some forecast value. Abnormally high tides are indicative of the presence of a storm. An area along the coast line where the tide exceeds the predicted gravitational tide and continues rising is in line with the advance of the tropical storm at the time the water started on its journey. If the point of greatest plus departure from the predicted tide shifts right or left, this indicates that the storm is changing direction toward the point where the increased rise is taking place.

Microseisms. Microseisms are more or less regular elastic surface waves which may result from a number of causes, including ocean waves and surf. Microseisms of these two types may be propagated over considerable distances, if no deep geological discontinuities exist, and may be measured at seismographic stations. Gutenberg [8] has pointed out that the location of tropical cyclones, and their intensities, direction, and rate of movement, can be determined from the differences in arrival time at three stations on a triangle, by plotting relative amplitudes and considering the changes with time. It appears that this technique might have its greatest application where airplane reconnaissance and ship reports are not available.

NEEDED RESEARCH ON TROPICAL CYCLONES

Additional Data Required. Since tropical cyclones usually develop in and often traverse remote oceanic areas, the acquisition of data required for a rigorous evaluation of many of the current theories of structure and development of these storms and for a solution of urgent forecast problems has been difficult or impossible. This is especially true with respect to upper-air information. Since tropical cyclones are relatively small, a denser network of observing stations is a prerequisite to further progress. Over large oceanic areas upper-air data have been negligible. Southern and western Florida, the most hurricane-susceptible area in the United States, would provide an excellent laboratory for hurricane study if a satisfactory network of observing stations could be set up. This network should be supplemented by a fleet of "weather-trucks" and mobile weather stations which could be dispatched and located to form any desired observing grid. Over adjacent water areas squadrons of planes, equipped with the latest meteorological devices for measuring winds, convergence, divergence, updrafts and downdrafts, should be dispatched into and over the storm area for inflight and parachute-radiosonde observations. Photographs over the top of the storm and in the eye should be taken for study purposes. For a period of at least two years, during the hurricane season, stationary ships should be placed at 300–500 mile intervals in the Gulf of Mexico, in the Caribbean Sea, and between latitudes 8°N and 30°N in the Atlantic for the purpose of securing regular surface and upper-air observations.

Urgent Research Needs for Forecasting. *Antecedent Prerequisites for Tropical Cyclone Development.* The exact combination of conditions which produces tropical cyclones must occur rarely, since the phenomenon itself is infrequent. The original causative factors which change the normally very weak and transitory pressure and wind deformations in the trades and the intertropical convergence zone to unstable waves and from the latter to intense tropical storms are not fully known. The testing of Riehl's recently published ideas on tropical cyclogenesis based on the interaction of surface perturbations and upper troughs and ridges should have very high priority as a research project. Radiosonde data, preferably from stationary ships, would be required. In the event the currently most persuasive theories are not substantiated, the new data would prove invaluable in deriving new ideas and further clues on the origin of tropical cyclones.

Differences in the intensities of tropical cyclones, which may be of the order of 80 to 90 mb, appear to depend upon apparently slight differences in the moisture and stability of inflowing air during the stage of marked development, with little change thereafter. Practicable techniques of adequately measuring and determining these differences and the establishment of criteria for limiting pressure gradients and wind shears in the developing storm would provide valuable tools for the forecaster.

Effective Steering Level. Although forecasters are using certain techniques based on steering at some arbitrary level or immediately above the storm circulation, further objective tests of these empirical rules are desirable. Recurvature is the most difficult problem facing the forecaster and, while lack of adequate observational data may always be a problem in certain areas, more definitive techniques seem possible.

Research Needed for More Complete Knowledge of Tropical-Cyclone Structure. 1. Tropical meteorology is deeply in debt to Father Deppermann for his painstaking assembly and analysis of typhoon characteristics in the Philippine area, but a further systematic collec-

tion of all observational data from tropical cyclones for statistical and physical analysis is badly needed. Evidence is far from conclusive about the actual facts with respect to the shape of inner isobars, the symmetry or asymmetry of distribution of certain elements around the storm, pressure-gradient and wind relationships, and many other characteristics which are taken more or less for granted at the present time.

2. Further studies are required to determine the exact relationship between the computed cyclostrophic term of the gradient wind and the observed wind, as a function of wind direction, velocity, storm movement, and other characteristics. More data on the actual pressure-gradient and wind-velocity relationships are needed.

3. Theories on the exact mechanics of evacuating air from the storm area are mainly deductive in nature, and data are insufficient to substantiate or disprove them. It should not be too difficult to obtain the necessary data from properly equipped airplanes.

4. A flight by Wexler [21] into a tropical cyclone in a plane equipped with a radio altimeter indicated areas of convergence considerably at variance with the normal concept of convergence and divergence in the storm area. It should be noted, however, that the storm was in an advanced stage of recurvature. Again the accumulation of data required for an answer to this question should not be difficult. Traverses should be made into storms at lower latitudes.

5. By means of planes or parachute radiosondes, more observational data, including rate of descending air, should be secured from the eye of the storm.

REFERENCES

1. Cline, I. M., *Tropical Cyclones*. New York, Macmillan, 1926.
2. Deppermann, C. E., *Some Characteristics of Philippine Typhoons*. Manila, Bureau of Printing, 1939.
3. —— *Outlines of Philippine Frontology*. Manila, Bureau of Printing, 1936.
4. —— "Notes on the Origin and Structure of Philippine Typhoons." *Bull. Amer. meteor. Soc.*, 28: 399–404 (1947).
5. Dunn, G. E., *Brief Survey of Tropical Atlantic Storms*. Univ. of Chicago, Inst. of Tropical Meteor., Rio Piedras, P. R., 1944.
6. Durst, C. S., and Sutcliffe, R. C., "The Importance of Vertical Motion in the Development of Tropical Revolving Storms." *Quart. J. R. meteor. Soc.*, 64: 75–84 (1938).
7. Gherzi, E., *Typhoons of 1936*. Zi-ka-wei Observatory, Shanghai, China.
8. Gutenberg, B., "Microseisms and Weather Forecasting." *J. Meteor.*, 4: 21–28 (1947).
9. Haurwitz, B., "Internal Waves in the Atmosphere and Convection Patterns." *Ann. N. Y. Acad. Sci.*, 48: 727–748 (1946).
10. —— "The Height of Tropical Cyclones and of the 'Eye' of the Storm." *Mon. Wea. Rev. Wash.*, 63: 45–49 (1935).
11. Mitchell, C. L., "West Indian Hurricanes and Other Tropical Cyclones of the North Atlantic Ocean." *Mon. Wea. Rev. Wash.*, Supp. No. 24, 47 pp. (1924).
12. Riehl, H., "On the Formation of West Atlantic Hurricanes." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 24, 64 pp. (1948).
13. —— "On the Formation of Typhoons." *J. Meteor.*, 5: 247–264 (1948).
14. —— and Shafer, R. J., "The Recurvature of Tropical Storms." *J. Meteor.*, 1: 42–54 (1944).
15. Schacht, E., "A Mean Hurricane Sounding for the Caribbean Area." *Bull. Amer. meteor. Soc.*, 27: 324–327 (1946).
16. Simpson, R. H., "On the Movement of Tropical Cyclones." *Trans. Amer. geophys. Un.*, 27: 641–655 (1946).
17. Stevenson, T., "Observations on the Relation Between the Height of Waves and Their Distance from the Windward Shore." *Edinb. new phil. J.*, 53: 358–359 (1852).
18. Tannehill, I. R., *Hurricanes*. Princeton, Princeton University Press, 1938.
19. Vines, B., *Cyclonic Circulation and Translatory Movement of West Indian Hurricanes*. U. S. Weather Bureau, Washington, D. C., 1898.
20. Visher, S., *Tropical Cyclones of the Pacific*. Bernice P. Bishop Museum, Bull. 20, Honolulu, Hawaii, 1925.
21. Wexler, H., "The Structure of the September 1944 Hurricane When off Cape Henry, Virginia." *Bull. Amer. meteor. Soc.*, 26: 156–159 (1945).
22. —— "Structure of Hurricanes as Determined by Radar." *Ann. N. Y. Acad. Sci.*, 48: 821–844 (1947).

AEROLOGY OF TROPICAL STORMS

By HERBERT RIEHL

University of Chicago

INTRODUCTION

One of the earliest controversies that arose in regard to tropical storms concerned their vertical extent. Some writers claimed that disturbances disappeared at heights as low as 3 km—that people standing on high mountain tops could look down on the violent cyclonic whirls beneath them. They believed that the cirrus motion at upper levels remained undisturbed and indicated the direction in which the cyclone was moving. Others insisted that hurricanes must reach to great heights, 10 km or more. They thought that cirrus radiated in all directions from the storms and used the change of direction of cirrus movement with time to locate position and track of centers.

The answer to this argument is now well known. As shown theoretically by Haurwitz [17] and empirically by high-level observations, mature storms extend through the troposphere. This affects such practical problems as routing of aircraft and forecasting storm movement. Since storms extend to the high troposphere, the upper layers should in part control the movement. We cannot hope to find the solution solely on sea-level or low-tropospheric charts, which very often do not reveal the state of the upper levels.

As we ascend from the surface to the higher atmosphere, the simple trade-wind current and the clear demarkation line separating tropics and middle latitudes disappear. Above 500–400 mb we encounter a chain of complex vortices of large dimension that are interlocked with the troughs and ridges of the long waves in the westerlies [27, 29]. Interaction between low- and high-latitude disturbances largely controls short-term developments in the tropical vortex train aloft and affects the motion of typhoons and hurricanes. It also plays a decisive role in the events that lead up to storm generation.

Formerly, formation and movement of hurricanes was forecast with detailed local analysis in a small area of the tropics. Since we know today that external forces are a major variable, we must expand the map analysis to include wide regions of the globe at all latitudes. Nevertheless the possibilities of detailed analysis are by no means exhausted. A complete three-dimensional description of hurricanes, such as has been rendered for middle-latitude cyclones, is as yet in the distant future.

In the following pages we shall consider the state of knowledge regarding the structure, formation, and movement of tropical cyclones. Since all knowledge depends on the observations and their evaluation, we shall begin with tropical data and methods of analysis.

TROPICAL DATA AND METHODS OF ANALYSIS

Observations. Tropical storms form and move over water. Therefore the number of observations available for analysis and forecasting have always been notoriously scarce. Even if perfect forecast methods were developed, the actual prediction would largely remain a guessing game for this reason. Apart from persistence and statistics on mean paths, even the best forecast method cannot help if there are no observations.

The data useful for three-dimensional analysis are cloud observations, and commercial-aircraft, pibal, rawin, and raob reports. Cloud observations and reports from commercial aircraft are most frequent, least costly, and also least useful. Wind observations from aircraft help only if pilots take double or triple drift measurements. This seldom happens. A good description of the state of the sky in surface observations still is hampered by the archaic code. Writers have pointed out the variety of cumuliform clouds that exist and have shown that broad inferences can be drawn from knowledge of the field distribution of cumulus types; but the code admits only the conventional L_1, L_2, and L_3 clouds.

Many forecasters have also complained that the quality of cloud reports has gone down with the advent of sounding balloons. This holds particularly for direction of middle- and high-cloud drift. At the beginning of this century the quality of these observations was high and the nephoscope was in frequent use. In those early days some of the articles which appeared were based on information about wind currents at high levels which we cannot match today. No wonder that some of the deductions rose far above the general level of the day and that they still form our major source of knowledge on upper-air currents over many areas, notably the tropical Atlantic Ocean.

It is unreasonable that the tropical forecaster should be deprived of the potentially largest source of information and have to rely on the few stations with expensive sounding-equipment. The writer proposes the following action:

1. Revision of the surface code to permit inclusion of at least three types of cumulus in one message. (There are about eight basic types that should be entered in the manuals on surface observations.)
2. Insistence on careful estimates of altitude and motion of middle and high clouds. (The code should allow for transmission of cloud direction in two layers.)
3. Installation of nephoscopes at all stations that make six-hourly reports, on naval vessels, and on reliable commercial ships on the principal shipping lanes.

Instrumental Measurements. Up to the present, the pilot-balloon observation has enjoyed the widest use among the instrumental measurements, but its drawbacks are numerous and well known. Balloons seldom reach the high troposphere. They are lost in clouds. In bad-weather areas where information is most needed,

wind data are often missing completely. Gradually, rawins should replace pibals everywhere.

The value of the raob in tropical analysis has been much discussed. This writer, along with others, used to think that they were useless. Horizontal gradients of pressure and temperature in the tropics are very small. At 300 mb, the height gradient may be as low as 200 ft per 10 degrees of latitude. A small error in sounding evaluation will produce an error of this magnitude at 300 mb. Moreover, local horizontal temperature gradients can be as large as the synoptic gradients. A single observation with such errors can grossly distort a whole map pattern.

In spite of these difficulties, raob analysis has turned out to be valuable when we are concerned with reports from stations where the quality of observations is high, such as San Juan in Puerto Rico, Swan Island, and Honolulu. Time sequences, especially when considered together with wind data, are reasonable. Obviously incorrect soundings are rare at such stations and stand out very clearly, especially in the case of nighttime observations. The soundings taken during the day give more erratic results and are not so useful. It is suggested that nighttime soundings be continued and expanded.

Methods of Analysis. *Area to be Analyzed.* As already noted, recent years have brought to light the strong influence that the upper layers exert on surface conditions. They have also revealed the large extent of interaction across the subtropical ridge. Thus tropical analysis should be extended to high levels and high latitudes. Writers, however, disagree on the usefulness of cross-equator analysis for short-term hurricane forecasting. Some, for example, state that West Pacific and Indian Ocean typhoons form *following* an intensification of flow across the equator. Others maintain that such an increase occurs *after* and *in consequence of* typhoon development. This writer has seen instances of the second type of situation but would not maintain that this is always the case. Experimental analysis across the equator is obviously useful. But as yet we cannot say that its value for short-term forecasting has been proven.

Contour Analysis. As in higher latitudes, the interest of the analyst centers on the fields of mass and motion. Since tropical storms decrease in intensity upward, surface analysis is a logical tool in hurricane situations. Outside of storm areas, Riehl and Schacht [32] have proposed adoption of the 850-mb (5000-ft) level as the basic level in order to eliminate many of the local features contained in surface reports.

Upper levels in regular or experimental use are 700, 500, 300, and 200 mb. The 700-mb level is situated in the center of the lower trade stream and reflects the disturbances in this current very well. At 500 mb we are in the zone of transition from trade-wind to upper-vortex regime. Winds are light and variable and height gradients are erratic. This surface often is not suitable for hurricane work, especially in the Pacific. Even the 300-mb level is not very satisfactory, although it serves well for many purposes of general tropical forecasting. Over the western portions of the oceans in summer, the high tropospheric vortices reach their greatest intensity near 200–150 mb. Analysis of the 200-mb level is essential for hurricane forecasting.

Because of the uncertainties involved in drawing contours in the tropics, described earlier, differential analysis and time sections must supplement the contour fields at all times. There is also the question of the pressure-wind relation. Opinions differ widely as to the latitude at which the gradient-wind relation becomes useless. It has been noted that aircraft dispatched with the "pressure-pattern" method [1] have arrived at their destination with considerable exactness in low latitudes. Some authors wish to apply the thermal-wind equation even at latitude 5°. Of course, there must be variations with synoptic conditions. In a strongly disturbed region, cross-isobar flow will be pronounced. It is of obvious importance to determine the true nongeostrophic flow. This could be done by using the pressure-pattern method to dispatch aircraft at different levels in selected situations and calculating the drift of the aircraft and the error between the computed and actual flight time. Until this is carried out, it is hazardous to make assumptions about the size of the angle between streamlines and contours when drawing upper-level charts.

Streamline Analysis. Because of the weak contour gradients that prevail especially at low levels outside of storm areas, streamline analysis can bring out many details of the field of motion. Some analysts, including the writer, at times have replaced contour and isobaric analysis entirely with qualitative streamline methods. The lines are drawn parallel to the wind direction and qualitative allowance for convergence and divergence is made by starting and ending some streamlines. Although this technique leads to clear and illustrative-looking charts, it has a distinct drawback: the charts are not suited for computations. An alternative is the construction of exact streamline and convergence charts. There is no doubt that exact streamline fields are one of the best potential tools for future research, especially for research concerning the structure of hurricanes.

Stability Charts. Since the thermodynamic hypothesis of tropical storms is the oldest, the literature carries many descriptions of vertical stability conditions in and near hurricanes. The trade-wind inversion disappears in storms. It has been a favorite argument to say that moisture accumulates under an inversion over the oceans. Then, as the inversion breaks down through heat accumulation below, or for some other reason, large quantities of water vapor suddenly become available for upward transport and condensation. The release of this latent heat provides an unbalanced energy source and can initiate storms.

This writer has applied this argument to the problem of driving mechanisms for the general circulation in low latitudes [29]. Local application to incipient hurricanes, however, has not proved fruitful. Storms often form in an atmosphere that has been nearly moist-adiabatic for a long time. We know that the steepest lapse rates occur in subsiding dry air above the trade inversion. Except for analysis of the inversion itself, the writer believes that stability charts are not likely

to yield results beyond the obvious. But the subject is popular, and there will be no shortage of studies of the stability consideration in the future.

Base of the Westerlies. Preparation of charts showing contours of the base of the westerlies (axis of the subtropical ridge) have been proposed by Riehl and Shafer [33]. Such charts can be a potent analysis tool. They furnish indications of temperature distribution and baroclinity of the atmosphere that greatly help to determine the upper-contour analysis where there are no raobs. The base of the westerlies is tied to many empirical forecast rules on hurricanes. For example, many storms develop from just south of the latitude where the base reaches the high troposphere (200 mb). If there is no base at all and the easterlies increase upward in intensity through the troposphere, hurricane development is not common. In the same situation, existing hurricanes will be steered strictly by the deep easterly current. The writer suggests that this simple tool be tried out in practice.

STRUCTURE OF TROPICAL STORMS

The early literature carries long and sometimes acrimonious debates on the structure of tropical storms. In recent years the arguments have abated. It has become clear that the structure varies with the age of a storm. Hurricanes that are just forming differ from those that have attained full intensity and those that are leaving the tropics. The stage of the fully grown storm has attracted the greatest amount of attention. Mature storms may travel wide distances over the tropical oceans with relatively little change of structure, and it is possible to make attempts at application of steady-state dynamics. In this section we shall also concern ourselves with mature storms.

Thermal Structure of the Rain Area. Observations taken during the last ten years have confirmed the classical idea that the air in the interior of tropical storms is less dense than its surroundings. A detailed discussion of radiosonde observations in hurricanes has been given by Schacht [37], Riehl [26], and Palmén [23]. The principal result is the following: Inside the rain area the vertical temperature distribution is the same as that obtained by raising an air particle dry-adiabatically from the sea surface to the condensation level, and then moist-adiabatically to 200 mb or even higher (Fig. 1*a*). Byers [6] adequately likens the hurricane to a huge parcel of air. Temperatures observed aloft in the rain area are the highest possible that can be obtained through ascent.

The situation is quite different from the ordinary zones of organized convection in low latitudes (wave troughs, shear lines). In these disturbances some ascent of surface air takes place. But convergence extends to relatively high levels (700–500 mb). Air initially situated anywhere between sea level and 500 mb is entrained and moves upward. Usually, this air is unsaturated at first and its moisture content is lower than that of the surface air that has risen. As shown by Stommel [41] and others, entrainment lowers the virtual temperature aloft in cumulus clouds compared to that which would prevail without entrainment. Very often air at upper levels in the convective zones is actually denser than the surrounding air. As shown by rawins, the cyclonic circulation increases upward in intensity.

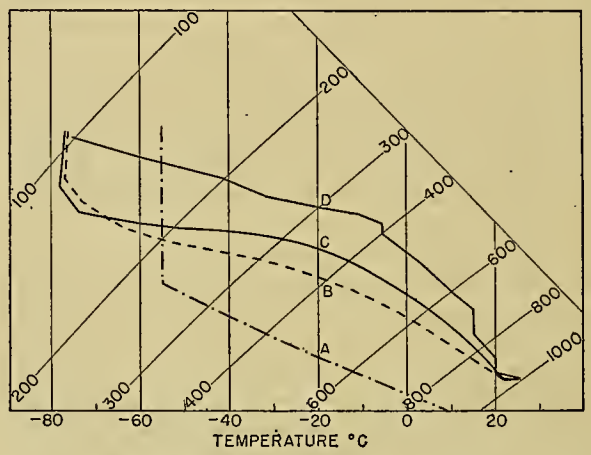


FIG. 1*a*.—Tephigram showing (*A*) U. S. Standard Atmosphere, (*B*) mean tropical atmosphere of the Caribbean in summer [37], (*C*) sounding in rain area of hurricanes, and (*D*) eye sounding [28]. In the diagram vertical lines are isotherms (°C) and slanting lines isobars (mb). Curve *C* follows a moist-adiabatic path from 900 to 200 mb.

Comparison of the vertical structure of these disturbances with tropical storms forces us to the conclusion that entrainment does not take place in the core of hurricanes and that convergence is restricted to the mixed "subcloud" layer [5].

The question arises, Does the foregoing hold true over the entire rain area? If it does, horizontal temperature gradients should not exist in the rain area,

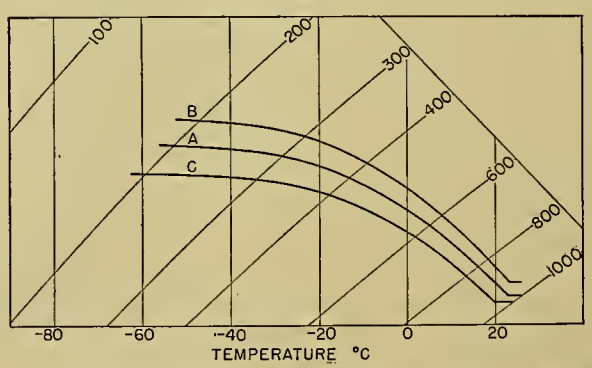


FIG. 1*b*.—Tephigram showing (*A*) ascent curve of surface air with T = 26C, spec. hum. = 18 g kg^{-1} (repetition of curve *C* of Fig. 1*a*), (*B*) ascent of the same air after isothermal expansion to 960 mb and moisture increase of 1.5 g kg^{-1}, and (*C*) ascent of surface air with T = 24C, spec. hum. = 16 g kg^{-1}.

and there should be a frontlike boundary at its outer edge. Deppermann [11] postulated such a boundary when, on the basis of empirical evidence from the Philippine area, he spoke of an outer and an inner ring of convection. The notion of a "wall" of hurricanes, often mentioned, supports Deppermann.

Palmén [23] was the first to venture a detailed vertical cross section through a hurricane. He showed, con-

trary to Deppermann, a gradual horizontal increase of temperature across the rain area, based largely on three-hourly special radiosonde observations at Miami, Florida, during the approach of a severe hurricane in September, 1947. Other series of special observations also fail to disclose a sharp boundary.

Several explanations of the difference between these authors are possible. Of these the most plausible one is that conditions are not symmetrical around hurricanes. A "wall" may be present in one or two quadrants but not the others. There is another interesting possibility. As pointed out by Byers [6], the temperature of the surface air spiralling toward a center must fall 3C or more during horizontal motion because of the large pressure reduction. Such a temperature decrease is seldom, if ever, observed. The temperature difference between ocean and surface air suddenly increases with lowering air pressure. As the ocean is greatly agitated and as large amounts of water are thrown into the atmosphere in the form of spray, very favorable conditions exist for rapid transfer of sensible and latent heat from ocean to air. Such transfer could account for a gradual increase of upper-level temperatures across the rain area. Figures 1*a* and 1*b* show that the effect referred to is very large. It roughly doubles the size of the positive area enclosed by inside and outside soundings on the thermodynamic diagram.

Microstructure of the Rain Area. There is conclusive evidence that conditions are not uniform as we circle a hurricane. Deppermann [10] showed that barometric fluctuations of shorter period are superimposed on the general barometric trace. These fluctuations have different periods, varying from minutes to hours. Deppermann [10] associated them in part with variations in cloud thickness and rain intensity.

The existence of a nonuniform cloud structure was proved by radar photographs of cloud systems in hurricanes [43]. On these photographs we can see that the rain area is composed of a number of bands of heavy precipitation that alternate with more settled conditions. The bands spiral cyclonically toward the center. Photographs of this kind have since been taken in large numbers. Further descriptive and analytical discussion of this aspect of the microstructure is one of the more interesting topics that await the investigator. A possible explanation is the development of internal waves. Another is the effect of falling rain and vertical downdrafts as discussed by Byers and Braham [7]. An analogy between the radar bands of Wexler and the thunderstorm-cell propagation proposed by Byers and Braham is by no means illogical.

Thermal Structure of the Eye. The eye of the hurricane has held the attention of all who have written on hurricanes from the earliest days. It is one of the oddest curiosities in meteorology. The precipitation ceases abruptly at the boundary of well-developed eyes, the skies clear, and the wind subsides suddenly. Observers inferred early that the vertical motion in the eye should be downward, the air therefore very warm and dry. Some thought that the tropopause was "sucked down."

Aerological observations have confirmed a part of these inferences. Pilots flying into eyes at various altitudes have reported large and sudden temperature increases. The best evidence comes from two radiosonde flights taken at Tampa, Florida, in two hurricanes [28, 39]. It is, of course, very difficult to send up balloons in an eye. In particular, there is no certainty that the balloons were still in the eye when they reached the high troposphere. In both cases, however, the eye was large. Chances are good that the balloons did not drift outside and that the temperatures reported at all levels are really indicative of the structure of vertical columns.

The sounding of October, 1945 [28] (Fig. 1*a*) showed great stability in the lower troposphere and a very steep lapse rate higher up. From the 100-mb level downward, temperatures were higher than otherwise observed in the tropics. Departures from the normal tropical atmosphere were extreme from 800 to 400 mb. They indicated subsidence amounting to several kilometers. The two Tampa soundings disagreed as to temperature in the high troposphere. In the case described by Simpson [39] the air above 300 mb was colder than that outside the eye. This observation is very difficult to interpret. The October 1945 ascent showed equalization of temperature with the surroundings only near 100 mb. Above this level, the air probably was colder inside the eye, so that the entire thermal structure relative to the outside resembled that of dynamic highs. Neither sounding reached the tropopause, and there was no evidence of "sucking down." On the contrary, it is probable that the tropopause is highest above the eye, again in analogy to warm anticyclones.

The extremely warm air of the eye is necessary to explain the very low surface pressures observed, at least statically. Haurwitz [17] has shown that application of the hydrostatic equation is entirely valid for computations of the vertical pressure distribution even in hurricanes. We can calculate, for example, what reduction of sea-level pressure is possible if the normal tropical atmosphere is replaced by the atmosphere of the rain area and if the pressure remains constant at the level where the two soundings intersect. This reduction is of the order of 30 mb and may reach 40 mb on the outside. It does not suffice to explain the very low sea-level pressures observed in many cases (900–950 mb). If we introduce an eye with a sloping boundary, however, following Haurwitz [17] and others (Fig. 2), the problem is resolved. Introduction of air still warmer than that of the rain area makes possible additional large pressure reductions.

Broadly speaking, the thermal structure of the eye, at least below 300 mb, is relatively well established. Many observations, of course, are needed to give all details, especially the slope of the boundary at different heights. Such observations would be of great scientific interest, but probably of little practical value.

Dynamical Aspects of the Rain Area. The thermal structure of the rain area permits some direct inferences regarding the dynamical structure of hurricanes. We have seen that soundings in the rain area coincide with

the computed ascent curve of air from the mixed layer. It followed that the entire rain area of a mature storm consists of air that has risen from the vicinity of the surface. The low-level convergence is restricted to the mixed layer, and a level of nondivergence is situated at its top.

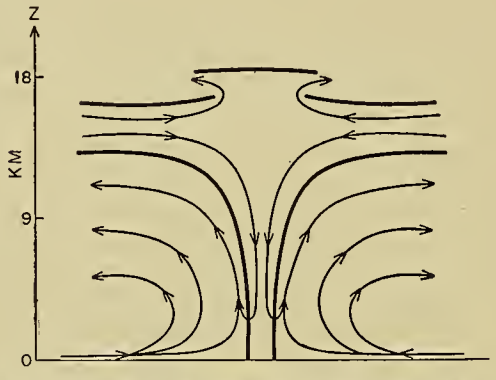


FIG. 2.—Vertical cross section through mature tropical storm showing hypothetical model of the vertical circulation. Heavy solid lines are eye boundaries and tropopauses. The model is based in part on Durst and Sutcliffe [13], Wexler [42], and Palmén [23].

Heat Transfer. We can estimate the size of the area from which a hurricane must draw surface air in order to build up the thermal structure actually encountered in the interior. Let r_0 denote the radius of this area and r the radius of the rain area as encountered. The eye may be neglected for this computation since its area is very small (10 per cent or less of the rain area). The depth of the mixed layer is about 50 mb and the clouds extend at least to 200 mb. It follows that $A_0 = 16A$, where A_0 is the initial area and A the rain area. Also, $A_0 = \pi r_0^2$ and $A = \pi r^2$. Therefore, $r_0 = 4r$. If $r = 100$ km, a moderate value, $r_0 = 400$ km, and the diameter of the initial area is about 8 degrees of latitude. This is a large value and shows how far the influence of even a small storm extends during the time of formation. In the course of its further life history, a very sizable fraction of the surface air initially situated over a tropical ocean is drawn into the circulation.

Up to now, no one has calculated the total effect that storms can exert on the general circulation as a result of the upward transfer of heat. Evidently an enormous gain of heat—both sensible and latent—accrues to the upper troposphere from the large-scale funnelling. It has been suggested that the frequency of tropical storms should be great when the normal processes of heat exchange between low and high levels are weaker than usual for some reason. But so far it has not been possible to establish a clear relation between hurricane frequency or duration and other factors of the general circulation.

Wind Distribution near the Surface. Estimates of lateral convergence of air into tropical storms and the amount of vertical motion present usually treat instantaneous values only, rather than totals integrated over some unit of time. Many observations have been published on the rate of inflow near the ground, as determined from surface-wind measurements. General laws, however, have not been formulated. The rate of inflow is quite variable. Usually it is not symmetrical, but confined to one or two quadrants. During the life of one storm, the quadrant of greatest inflow will often change.

Various forecast rules have been offered concerning the relation between storm movement and quadrant of greatest inflow. One such rule states that a storm will move toward the region of heaviest rainfall (greatest convergence); another states that the heaviest precipitation occurs to the rear of a center. Observations can be found to demonstrate practically any relation between direction of motion and rainfall concentration. This merely shows that a variety of synoptic influences are operative. A considerable advance could be gained through a serious attack on the connection between synoptic situation, distribution of convergence, and storm motion.

Our knowledge regarding the wind distribution within tropical storms and the dynamical laws that guide the air from the outskirts to the center of the cyclones is so deficient as to be deplorable. Deppermann is one of the few writers who has made a detailed effort to calculate radial and tangential velocity components. He also has presented a theoretical model [11] which discusses the tropical storm as a Rankine vortex with two "rings" of convection. Apart from Deppermann, writers have contented themselves with application of simple hydrodynamics, generally only a treatment of the vr vortex. The best discussion is found in Brunt's textbook [4, pp. 298–306]. Even Brunt, however, treats only a few elementary concepts—circulation, vorticity, and the wind distribution in a stationary vortex with point convergence at the center.

It is easy to understand why the synoptic meteorologist is discouraged from attacking the problem of the dynamics of the tropical storm. He rarely has data at his disposal for which he can claim general validity. But the apathy on the part of theorists is hard to comprehend. To the best of the writer's understanding, no laboratory experiment has been carried out to determine whether "simple" vortices can be generated in air as in liquids. It would be quite feasible to construct theoretical hurricane models, dropping some of the assumptions contained in the simple theory. If the air did move under conservation of momentum, the absolute rather than the relative motion of air about a hurricane center should be considered. The absolute rather than the relative vorticity should become zero throughout the body of a hurricane. This, however, never happens. There are a few famous cases—often quoted—in which the relative vorticity was found to be nearly zero. Since we know of no physical law that demands conservation of relative angular momentum in a hurricane, the rare cases quoted must be regarded as accidental.

It is certain that the absolute angular momentum is not conserved in the air flowing toward a center. Ground and eddy friction can be responsible for this, as well as the asymmetrical shape of many storms. None of these factors has been properly investigated, especially the

important frictional term in the equation of motion. As a prelude for further objective work, an extensive theoretical study should be executed.

Upper Outflow. The laws governing the outflow aloft (Fig. 2) have been described more satisfactorily than those governing the lower inflow. The pressure-gradient force decreases upward on account of the temperature distribution and the gradient actually reverses above 10–13 km [37]. Outflow will take place throughout the layer of decreasing pressure gradient if the air conserves its absolute angular momentum [13]. This conservation principle is more likely to hold true for the upper outflow, since ground friction is absent. Above the friction layer, the principal problem is the almost complete absence of data. There are many reasons that suggest an upward decrease of the circulation. Such wind measurements as have been made have shown the decrease and eventual disappearance of the vortices between 300 mb and 100 mb. In some cases, the circulation has actually become anticyclonic at 200 mb.

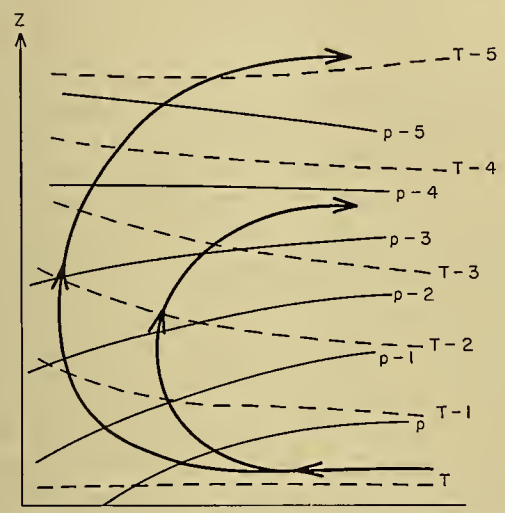


FIG. 3.—Same as Fig. 2, outside the eye. Solid lines are isobars and dashed lines are isotherms.

Dynamical Aspects of the Eye. The only explanation given for the eye so far is that the law, vr = const, cannot hold near a vortex center, as wind speeds and kinetic energy there would have to become infinite. Friction alone would act to prevent this. This negative statement is reasonable, but it hardly explains the absolute calms frequently observed or the variable diameter of eyes. We also have no knowledge concerning the rate of vertical motion in the eye. The direction evidently is downward, at least below 300 mb. Convergence must exist in the upper troposphere and divergence in the lower layers [28]. This is indicated by the stability distribution in the few available eye soundings (Fig. 1*a*). It follows from continuity reasons that air must constantly seep through the boundary of the eye in the low troposphere and enter the rain area. Beyond this statement, little progress is likely to take place in the future until actual observations of winds aloft in and near eyes become available.

Maintenance of Tropical Storms. Qualitatively, it is easier to understand the maintenance of tropical cyclones than of extratropical storms. The pressure decreases toward the center and the temperature rises (Fig. 3). Air moves inward near the ground and upward in the central region of convergence, while the outflow is aloft. Thus the circulation about the vertical solenoids is in the kinetic-energy-producing sense (excepting the eye). This solenoid field will be maintained as long as colder air with lower moisture content does not enter the center; and as long as the air evicted aloft does not descend in the immediate outskirts but is removed considerable distances. A hurricane and its surroundings cannot be a closed thermodynamic system.

As the maintenance of tropical storms does not appear to offer very difficult problems in principle, future research can turn to quantitative considerations: the measurement of heat generation, conversion into kinetic energy, frictional dissipation, etc. Again the need for accurate data stands out as the foremost requirement to advance such research.

FORMATION OF HURRICANES

Scanty as is the literature on storm structure, the number of papers written on formation is very large. Up to very recent days, investigators generally have sought some grounds leading to extensive surface convergence and attendant development of cyclonic circulation. As Brunt [4] points out, the question of the upper outflow necessary to account for the observed surface pressure falls has not been answered. Convergence near the ground should lead to pressure rise and not fall at the center. This is a basic paradox of meteorology, an obstacle to the explanation of extratropical as well as tropical cyclones. It is true that upper warming caused by convection currents will produce radial outward acceleration aloft. Yet it is not clear—and has not been shown—why this outflow should become greater than the inflow. If the low levels lead in producing storms, the surface convergence must precede the high-level divergence. It is only in the last few years that attempts have been made to reverse the problem and consider the high-level events as leading [30, 35]. If these attempts should sustain the test of time, it would follow that the low-level convergence develops as a reaction to the pressure changes imposed from aloft.

Convection Theory. The classical thermodynamic (convective) theory of hurricane formation has few defenders today. Many of the heaviest tropical rainfalls occur without the existence of a cyclonic circulation, and even closed depressions with torrential precipitation often fail to deepen. Such centers have been observed to travel in relatively steady state over distances in excess of 1000–1500 miles. It follows that convection is not a sufficient criterion for storm development.

No one, however, denies that convection is necessary for generation of tropical storms. As stated before, the air that fills the body of a hurricane is all drawn from the mixed layer near the surface. Therefore the properties of the surface air and their variation with time in part determine whether or not storms will form. Pal-

mén [23], for example, notes that the temperature difference between air ascending from the surface and the surroundings aloft is much less in winter than in summer at Swan Island (17°N, 84°W). There are no hurricanes over the Caribbean Sea in winter. Storm occurrence is also rare over the cool waters of the eastern parts of the tropical oceans. It is most plentiful in the west where the sea surface is warmest. By analogy, it is possible to argue that storm frequency should be less than average in a year when the sea-surface temperature is below normal.

It is evident that research following up this suggestion can prove most fruitful. Caution, however, is necessary in interpretation. Temperatures aloft are as much of a variable as surface temperatures. Anomalies of temperature at different altitudes may not be independent of each other but produced by some common dynamical factor. The same may be true even for sea-surface temperatures.

Frontal Theory. Ever since the formulation of the polar-front theory there have been persistent attempts, beginning with that of Brooks and Braby [3], to transplant the Bergen reasoning to the equatorial zone. Advocates of this theory visualize the equatorial convergence zone (ECZ) as a front with inverted temperature gradient. The peaks of warm sectors and developing cyclone centers should therefore be found where the ECZ bulges equatorward. Actually, there is practically universal agreement today that forming centers are initially situated near poleward bulges. This observation has led some analysts to become skeptical regarding the wholesale importation of the Norwegian ideas into the tropics. A large group of meteorologists, however, today adhere to the view that hurricanes are generated at boundaries between warm and relatively cold air. Even textbooks of high renown (Brunt [4] p. 304) carry statements to that effect. Brunt, however, is quite conscious that the frontal approach solves the basic cyclone problem no more in low than in high latitudes.

It is not possible here to examine in detail the reasoning of the adherents of the frontal theory. Only one additional point will be presented that throws doubt on the validity of the frontal concept. This is again the matter of the properties of the surface air. If a rapid occlusion of a portion of ECZ takes place, when there is a real density difference across this boundary, the whole surface layer of the nascent cyclone will soon consist of the denser air only. This must prevent deepening. It is a principal difference between high- and low-latitude cyclones that all air entering the orbit of the latter ascends in the core, while outside the tropics the polar air sinks in disturbances and only the tropical air rises. Occlusion must be an acute hindrance to hurricane development. Actually, we can observe that even mature hurricanes weaken and sometimes disappear when a current with relatively polar characteristics enters the core while it is still situated in the tropics. A forecast of filling will generally succeed in such cases.

For comparison, the data of Figs. 1*a* and 1*b* are presented in tabular form in Table I. Temperatures at certain isobaric surfaces are given for (1) the average tropical atmosphere in the Caribbean [37]; (2) the ascent of the surface air of average properties (26C, 18 g kg^{-1}), (3) the ascent of the same air after isothermal motion to 960 mb and addition of 1.5 g kg^{-1} of moisture, and (4) the ascent of air with very slight polar characteristics (24C, 15.5 g kg^{-1}).

TABLE I.—TEMPERATURES ALOFT FOR THE CONDITIONS (1)-(4) DESCRIBED IN THE TEXT

Pressure *mb*	(1)	(2)	(3)	(4)
surface	26.0C	26.0C	26.0C	24.0C
900	20.0	20.5	22.5	18.0
700	8.0	11.0	13.5	8.0
500	− 6.0	− 3.0	0.5	− 6.5
300	−33.0	−28.5	−23.5	−34.0
200	−55.0	−53.0	−47.0	−59.0

The temperature differences between the four soundings are quite large at 700 mb and very large at 300 and 200 mb. If virtual instead of dry temperature had been used, they would be still somewhat larger (up to 1C near the ground). It is seen, as stated earlier, that the temperature difference between the rain area and the surroundings doubles if we consider sounding (3) instead of sounding (2). The slightly polar air is nowhere warmer than the average tropical atmosphere, and it is cooler at most levels. This shows clearly in what measure a slight temperature and humidity drop acts as a deterrent for an incipient hurricane circulation.

These observations throw doubt on the frontal hypothesis. Some writers, however, have maintained that the frontal temperature contrast is very small and merely serves as "trigger action" to start heavy convection. But we have seen that heavy convection is not a sufficient condition for typhoon development. Moreover, if the air-mass contrast at the ECZ is as minute as suggested by these writers, it would seem that they are trying to employ a very dubious mechanism.

Synoptic Conditions During Storm Formation. Although the physical reasoning of the frontal advocates is open to doubt, their geometrical picture is to some extent correct. Storms never develop spontaneously in the undisturbed tropical currents but always in a preexisting disturbance. Such a disturbance may be of the shear-line type described (ECZ) or it may have the character of a transverse wave (Riehl [25]). Intensification of the bad weather zone attending such a disturbance and formation of a closed depression generally takes place when two or more disturbances meet [8, 26]. Considerable synoptic evidence supports this statement. The combination or superposition of disturbances usually results from motion in different directions or at different rates. Among many possibilities, a common type is westward travel of a wave trough in the easterlies that intersects an ECZ extending east-west. At other times such wave troughs become coupled with the southern extensions of eastward-moving troughs in the upper westerlies of the polar zone. The superpositions, therefore, can be horizontal and/or vertical.

To date, only the synoptic circumstances of super-

position have been explored, while the dynamics have been neglected. It is certain that most tropical depressions form in consequence of superposition. Yet the great majority of these circulations do not develop beyond a weak wind field with maximum speeds of about Beaufort 6. Sometimes such weak centers will persist in steady state up to a week and then suddenly deepen.

The Hypothesis of J. S. Sawyer. A major synoptic problem, therefore, remains: Under what circumstances will storms attain great intensity? The solution of this question must be connected with the mechanism of the upper outflow. Durst and Sutcliffe [13] have given a plausible explanation for the outflow in case of the mature hurricane. Their reasoning, however, is not applicable to the period of establishment of a storm. Sawyer [35] adopts a line of thought that has also been put forward to explain cyclone formation in higher latitudes.[1] The basic principle is given by Solberg [40], who showed that a ring of air particles may become "dynamically unstable" provided that

$$f + \zeta < 0,$$

where f is the Coriolis parameter and ζ the relative vorticity about the axis normal to the earth's surface. For an explanation of the criterion the reader is referred to the articles quoted.

Although Sawyer's approach is novel and attractive, there are serious objections. In particular, it is dubious whether zones of "dynamic instability" exist in the tropics prior to storm development. It is true that the criterion is much more easily realized in low than in high latitudes, since the value of the Coriolis parameter is much smaller. But a plausible mechanism for production and maintenance of negative absolute vorticity has not been given, nor has its actual existence in tropical currents prior to cyclogenesis been demonstrated. Within the writer's experience, a zone of strong anticyclonic shear, comparable to that on the south side of the circumpolar westerlies, does not exist in the tropics, at least within the latitude belt where tropical storms form. Sawyer himself [36] appears to be a little worried about his 1947 statements.

The Hypothesis of H. Riehl. The attempt by Riehl [30] to explain the increase in intensity, does not postulate any special internal characteristics within the zone of cyclone formation. It is based on the introduction of external forces aloft to produce the initial upper divergence. A number of observational facts are utilized for depicting the setting in which certain types of storms—but not all—develop:

1. Deepening takes place when the curvature of the high-level flow is anticyclonic, or changes from cyclonic to anticyclonic. A high-level anticyclone or wedge is situated near (commonly west of) the surface disturbance, and an upper-level cyclone or trough lies to its east (Fig. 4).

2. Superposition of these high-level centers and the wave troughs and ridges in the polar westerlies occurs so that high- and low-latitude systems are in phase. This strengthens the meridional flow components aloft and weakens the zonal component (Fig. 5). The intensification of the current above the surface depression must be brought about by acceleration toward lower pressure, therefore toward the east in the model chosen (Fig. 4). This eastward acceleration will give an initial pressure fall at the surface in the tropical disturbance.

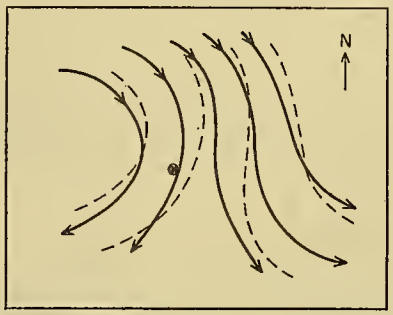


FIG. 4.—Model of streamlines and contours (dashed lines) at 200 mb during deepening of a tropical disturbance (heavy dot) in the Northern Hemisphere.

3. There is a general equatorward advance of polar air over the ocean east of the depression. This polar air does not enter the circulation. It subsides and diverges,[2] as usual, on its way toward lower latitudes. Continuity demands upper inflow above the region of subsidence. This further contributes to eastward deflection of the high tropospheric air toward the east. Convergence and descent take place at the left edge of the upper current, looking downstream (Fig. 6). Divergence aloft develops at the right edge. It produces a surface pressure fall,

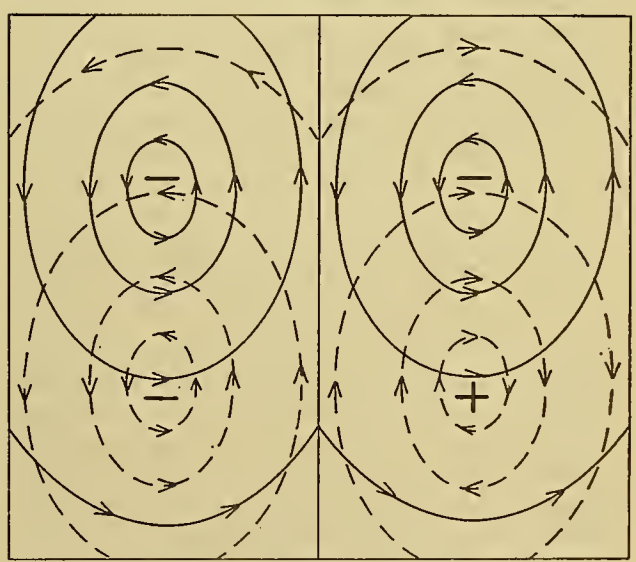


FIG. 5.—Streamlines of perturbation centers at 200 mb during in-phase (left) and out-of-phase (right) superposition of a trough in the westerlies on disturbances in the tropical vortex train aloft.

1. Consult "Extratropical Cyclones" by J. Bjerknes, pp. 577–598 in this Compendium.

2. Such divergence can also be obtained by "instability of the trades" [27].

followed by convergence at the ground and ascent of the low-level air in the initial tropical depression.

4. The vertical cross-stream circulation described is "direct," therefore kinetic-energy producing, as the air is less dense in the tropical disturbance where ascent takes place than in the surroundings. Maintenance of the vertical circulation is accomplished by continuous advection of more polar air at the left of the upper current and moist-adiabatic ascent at the right, looking downstream.

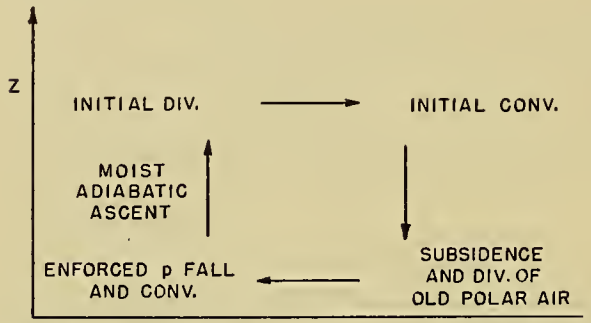


Fig. 6.—Vertical cross section taken eastward from position of tropical depression shown in Fig. 4. Model illustrates hypothetical development of vertical cross-stream circulation leading to deepening of the surface depression.

5. Several energy sources have been used in creating and maintaining the direct vertical circulation cell. The in-phase superposition of high- and low-latitude wave trains in the upper troposphere directly intensifies the speed of the high-level current above the surface depression. Use has also been made of old polar air, but in a manner quite different from that implied in the frontal theory of hurricane formation. The polar air does not enter the forming storm but moves equatorward at some distance and subsides. This subsidence is partly responsible for creating an organized circulation cell.

6. The existence of the direct circulation, initiated by the large-scale synoptic picture, increases the efficiency of the convective motion in the surface disturbance in generating kinetic energy. A priori there is no preferred direction for motion produced from condensation. The small-scale circulation cells initiated in different parts of a disturbed area tend to cancel each other. Introduction of the large-scale circulation cell, however, aligns the smaller convection cells in a definite direction. The air that rises in the zone of convection is removed a considerable distance before it sinks again.

At this time it is not possible to tell whether this latest model will withstand the test of experience. The sequence described is not the only one that can lead to cyclogenesis. Both in high and low latitudes, more than one synoptic pattern can lead to deepening. It is fair to say, however, that it is reasonable to attack the problem by placing emphasis on the upper divergence as an initiating mechanism. All efforts to explain deepening from low-level considerations have failed to date, since none of them can explain the surface pressure falls. The future evidently must concern itself with further development of models that lead to cyclogenesis. It is probable that the interrelation between different latitude belts will ultimately be recognized as an important factor. If that holds true, there is also some hope for preparation of extended forecasts of hurricane formation. The interrelation between high and low latitudes is a function of the state of the general circulation. If it should become possible to predict longer term trends of the general circulation, success in forecasting hurricanes on an extended basis could also follow.

MOVEMENT OF HURRICANES

This is the most widely discussed aspect of tropical forecasting, certainly one of the most vital. Contrary to the topic of formation, recognition of the importance of external forces for predicting the motion of storms dates from the last century. The close relation between storm tracks and the position and intensity of the subtropical highs (steering) was recognized very early [16]. Mitchell [21] was the first to provide an extensive set of rules for forecasting recurvature, based largely on the effect of travelling surface lows and highs in the westerlies on the tropical storms. Dunn [12] used similar reasoning, but applied it to 10,000-ft streamline charts. His attack was extended by Riehl and Shafer [33], who sought to determine under what circumstances hurricanes that meet a polar trough will recurve and when they will bypass this trough and continue westward. The height of the base of the polar westerlies played the principal role in their argument.

Recently, Simpson [38] introduced the "warm tongue" steering method, which from the dynamical point of view is similar to that of Dunn [12] and Riehl and Shafer [33]. Mintz [20], and Moore [22] applied the ideas of Bjerknes and Holmboe [2] to make quantitative forecasts of displacement. This is an interesting proposal and has led to some good forecasts. A purely dynamical approach is that of Yeh [44], who at first computes the path of a hurricane situated in an easterly current and then superimposes a general meridional component. He finds that storms execute oscillations of minor but definite amplitude while in the easterlies. There is, however, no total deviation of the storm track to the right of the steering current as maintained in some previous studies. When a storm comes under the influence of a polar trough, a northward displacement occurs. But this displacement will vary widely, depending on certain initial conditions, when the polar trough appears.

Riehl and Burgner [31] made a quantitative test of the steering principle. The method used was applicable to the zonal component only. An area 90 degrees of longitude long and 5 degrees of latitude wide, centered on the storm, was defined as the region of influence. Within this area, the mean zonal component of motion at 700 mb was computed and correlated against the 24-hr zonal component of storm displacement. The resulting scatter diagram showed that in the mean the zonal components of storm movement and steering current are equal. Numerous deviations, however, occurred and these were relatively large in that portion

of the graph where the zonal motion approached zero. The mean deviation was about 1 degree of latitude per day.

Steering Principle. It is no accident that the study of motion of disturbances has stood in the foreground of literature on the tropics. Contrary to other aspects of meteorology, this aspect of forecasting is more developed for the tropics than for middle latitudes. The reason for this can be seen readily when we examine what Petterssen [24] has termed the "area of uncertainty." Most tropical storms are small and have a tight circulation compared to extratropical cyclones. An error of 1–2 degrees of latitude in forecasting the position of a middle-latitude storm affects the forecast verification but little. In the tropics, however, an entire storm is often contained within this distance. An error of 1–2 degrees will make all the difference at any given station between wind of hurricane force and no wind at all.

The steering concept at first was used in a directional sense only. Hurricane tracks appeared to parallel the sea-level isobars over the Atlantic to a large extent. It is only since the late 1930's that some forecasters attempted to obtain the speed of displacement from the upper wind field. Difficulties arose at once in choice of steering level since speeds varied with height. This difficulty also applied to direction whenever the wind direction changed with height, especially near the subtropical ridgeline. Simpson [38] pointed out that there has been a tendency to shift the level thought representative of steering higher and higher as the upper-wind observations were extended upward with the advent of rawins. The difficulties of making a proper choice of steering level have led to doubts on the part of many forecasters as to the usefulness of the method in principle.

In the writer's opinion, the difficulties result mainly from application of steering techniques to situations in which they should not be used. In a strict sense, steering refers to the forces that guide a very small disturbance in a broad zonal current that is steady and usually considered uniform along all space axes. The disturbance also does not change shape or intensity with time. These restrictions never are fulfilled in reality. Two questions come up: Under what actual conditions can the simple theory be applied without much loss of forecast accuracy? and To what extent can theory itself modify the restrictions listed above? Not much has been tried along the latter line. But the problem is quite urgent. It is easy to see that formal solutions of the equations of motion will not be possible if many restrictions are dropped. But a numerical integration procedure may carry quite far, provided some good information exists as to what quantities to put into a computing machine. In this respect, studies of the kind attempted by Riehl and Burgner [31] may provide the necessary background. They can determine which theoretical restrictions are immaterial for the forecast and which impose limitations. It is almost certain, for example, that variations along the vertical axis will turn out to be important. This is amply indicated by the literature of the last ten years cited above. Probably the concept of "steering level" must be replaced by that of "steering layer." Tropical storms extend through almost the entire troposphere. It is quite reasonable to expect that some integral function of conditions at all heights will serve much better than conditions at one level alone.

Recurvature. The motion of tropical storms from the tropics into the middle latitudes can take place in several ways. Sometimes the forecast is very easy, as when upper winds are nearly uniform through a deep layer and gradually turn from east through south to southwest as we follow them around the western edge of a subtropical high pressure cell. For such cases, methods like those developed by Yeh [44] should be fully applicable. Important difficulties arise whenever winds are weak and/or when the vertical wind-shear is large. In such situations we can no longer speak of steering.

The literature on the subject, referred to above, gives empirical forecast rules for several types of tropical storms. No one, however, has claimed to have developed a perfect qualitative forecast method. The occasional failures of all systems stand out. Quantitative calculations of recurvature from empirical data have not been attempted or have been unsuccessful. Theory has bypassed the problem. Thus the state of knowledge with respect to the many situations when recurvature hangs in the balance is unsatisfactory.

It is not easy to make suggestions for future work that can readily be carried out. The trouble stems from two sources. One of the missing elements in the recurvature forecasts is a knowledge of the internal forces within a storm. These cannot be measured at all at present. The other difficulty lies in the fact that the net external force operating on storms in recurvature positions is a small resultant of large opposing forces; this is also indicated by the slow motion of storms at recurvature points. It is practically impossible to evaluate the net force from north or south acting on a storm even at one level, to say nothing of an integral from sea level to the tropopause. The station network necessary to calculate such resultant forces simply does not exist anywhere.

Thus we must again demand more data if physical analysis of recurvature is to make much headway. Unfortunately, it is probable that the great density of stations really necessary for this problem will never be realized on a routine basis because of cost. For practical purposes, future recurvature work should follow statistical rather than physical approaches. Mitchell [21], for example, has already shown the importance of the rate of motion of extratropical troughs. Riehl and Shafer [33] have emphasized the vertical structure of the subtropical high-pressure belt. Together with the paper by Klein and Winston [19] this literature yields a great number of starting points for statistical correlation. If future efforts are applied in this direction, it may be possible to bypass the obstacle furnished by the balance of forces at the recurvature point.

Large Storms. The problem of forecasting the motion of large storms occurs mainly in the Pacific Ocean in

the Northern Hemisphere. Typhoons at times can become as big as any Aleutian or Icelandic lows. Such centers move quite slowly at times. Frequently they are entirely stationary for days, usually near latitudes 25–30°. Evidently it is of greatest importance to forecast when such storms will stagnate and when and how they will resume motion.

In the whole field of hurricane work, this topic has been left untouched more than any other. From both theoretical and practical viewpoints the difficulties have appeared to be almost insurmountable. Among the older literature only the work of Fujiwhara [14, 15] is applicable. He has taken up the problem of rotation of vortices about each other. If a storm is situated as described above, it will be flanked by other vortices east, west, south, and frequently even north, especially in the upper troposphere. A mean zonal motion does not exist at high levels. Attempts at forecasting must turn to the relation between the upper vortices.

Theoretically, the problem of internal forces also moves into the foreground. This subject remained dormant until 1948. Since then two theoretical papers have appeared that treat the analysis of internal forces in vortices on the rotating earth [9, 34]. At this time, it is impossible to say how useful these particular papers will prove in practice, since again there are no direct data and application depends entirely on inference. But it is gratifying to note that the theoretical deadlock has been broken. Now that a way has been shown, dynamic meteorologists have a good opportunity to advance the understanding of the movement of large typhoons.

On the practical side, recent studies of the high-tropospheric wind structure over the tropical oceans [27] and the interrelation between high- and low-latitude circulations [8] have produced an opening. The problem of large storms evidently can be solved only by considering the atmosphere over very wide areas. Events in the entire belt of polar westerlies, the motion of long waves in the westerlies, and the structure and movement of the high-tropospheric vortex train in low latitudes must determine the fate of typhoons of huge size. A southward shift of a westerly jet stream or its formation in low latitudes, plus development of an extended trough near or west of a typhoon, might be favorable for acceleration of the typhoon into middle latitudes. Settling of a major ridge to its north could produce stagnation. These are merely suggestions which outline the direction which new efforts could take.

CONCLUSION

This report has tried to demonstrate briefly the current knowledge on structure, formation, and movement of tropical storms, shortcomings of this knowledge, and suggested paths for future research. The outstanding deficiency is the lack of observations which is so keenly felt in all low-latitude work. A network of upper-air data such as exists in the United States has never been at the disposal of the tropical analyst. Moreover, the quality of the high-level data is sometimes poor since errors in radiosonde flight evaluation can be as large as synoptic changes. If many upper-wind observations are taken, however, the quality of the reported heights and temperatures aloft can be judged easily. For this reason it is proposed to have as many rawins as possible and to revive the old and relatively inexpensive reports of middle- and high-cloud direction.

For scientific purposes, detailed observational programs should be initiated for short periods in low latitudes during the hurricane season. If this is done, it may be possible to derive analysis and forecast principles that are also applicable when less detailed data are at hand. The suggestions for the future take on several forms. Regarding structure, new observations and theoretical studies alone can bring advances. Formation is a subject that for the present should be continued with standard qualitative synoptic methods plus theoretical calculations based on the empirical inferences. As the release of several forms of energy is involved, quantitative computations appear premature at this time. Methods of numerical integration are recommended for forecasting the movement of hurricanes by steering currents. Statistical correlations are most likely to yield an advance on the problem of recurvature.

REFERENCES

1. Bellamy, J. C., "The Use of Pressure Altitude and Altimeter Corrections in Meteorology." *J. Meteor.*, 2: 1–79 (1945).
2. Bjerknes, J., and Holmboe, J., "On the Theory of Cyclones." *J. Meteor.*, 1: 1–22 (1944).
3. Brooks, C. E. P., and Braby, H. W., "The Clash of the Trades in the Pacific." *Quart. J. R. meteor. Soc.*, 47: 1–13 (1921).
4. Brunt, D., *Physical and Dynamical Meteorology.* Cambridge, University Press, 1939.
5. Bunker, A. F., and others, "Vertical Distribution of Temperature and Humidity over the Caribbean Sea." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 11, No. 1 (1949).
6. Byers, H. R., *General Meteorology.* New York, McGraw, 1944.
7. —— and Braham, R. R., Jr., "Thunderstorm Structure and Circulation." *J. Meteor.*, 5: 71–86 (1948).
8. Cressman, G. P., "Relations between High and Low Latitude Circulations." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 24, pp. 65–100 (1948).
9. Davies, T. V., "Rotatory Flow on the Surface of the Earth. Part I—Cyclostrophic Motion." *Phil. Mag.*, Ser. 7, 39: 482–491 (1948).
10. Deppermann, C. E., *Some Characteristics of Philippine Typhoons*, 143 pp. Manila, Bureau of Printing, 1939.
11. —— "Notes on the Origin and Structure of Philippine Typhoons." *Bull. Amer. meteor. Soc.*, 28: 399–404 (1947).
12. Dunn, G. E., "Aerology in the Hurricane Warning Service." *Mon. Wea. Rev. Wash.*, 68: 303–315 (1940).
13. Durst, C. S., and Sutcliffe, R. C., "The Importance of Vertical Motion in the Development of Tropical Revolving Storms." *Quart. J. R. meteor. Soc.*, 64: 75–84 (1938).
14. Fujiwhara, S., "The Natural Tendency Towards Symmetry of Motion and Its Application as a Principle of Meteorology." *Quart. J. R. meteor. Soc.*, 47: 287–293 (1921).
15. —— "On the Growth and Decay of Vortical Systems." *Quart. J. R. meteor. Soc.*, 49: 75–104 (1923).

16. Hann, J., "Ueber die Beziehungen zwischen den Luftdruckdifferenzen und der Windgeschwindigkeit nach den Theorien von Ferrel und Colding." *Z. öst. Ges. Meteor.*, 10: 81–88, 97–106 (1875).
17. Haurwitz, B., "The Height of Tropical Cyclones and of the 'Eye' of the Storm." *Mon. Wea. Rev. Wash.*, 63: 45–49 (1935).
18. —— *Dynamic Meteorology.* New York, McGraw, 1941.
19. Klein, W. H., and Winston, J. S., "The Path of the Atlantic Hurricane of September 1947 in Relation to the Hemispheric Circulation." *Bull. Amer. meteor. Soc.*, 28: 447–452 (1947).
20. Mintz, Y., "A Rule for Forecasting the Eccentricity and Direction of Motion of Tropical Cyclones." *Bull. Amer. meteor. Soc.*, 28: 121–125 (1947).
21. Mitchell, C. L., "West Indian Hurricanes and Other Tropical Cyclones of the North Atlantic Ocean." *Mon. Wea. Rev. Wash.*, Supp. No. 24, 47 pp. (1924).
22. Moore, R. L., "Forecasting the Motion of Tropical Cyclones." *Bull. Amer. meteor. Soc.*, 27: 410–415 (1946).
23. Palmén, E., "On the Formation and Structure of Tropical Hurricanes." *Geophysica*, 3: 26–38 (1948).
24. Petterssen, S., *Weather Analysis and Forecasting.* New York, McGraw, 1940.
25. Riehl, H., "Waves in the Easterlies and the Polar Front in the Tropics." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 17 (1945).
26. —— "On the Formation of West Atlantic Hurricanes." *Dept. Meteor. Univ. Chicago, Misc. Rep.*, No. 24, pp. 1–64 (1948).
27. —— "On the Formation of Typhoons." *J. Meteor.*, 5: 247–264 (1948).
28. —— "A Radiosonde Observation in the Eye of a Hurricane." *Quart. J. R. meteor. Soc.*, 74: 194–196 (1948).
29. —— "On the Role of the Tropics in the General Circulation of the Atmosphere." *Tellus*, 2: 1–17 (1950).
30. —— "A Model of Hurricane Formation." *J. appl. Phys.*, 21: 917–925 (1950).
31. —— and Burgner, N. M., "Further Studies of the Movement and Formation of Hurricanes and Their Forecasting." *Bull. Amer. meteor. Soc.*, 31: 244–253 (1950).
32. Riehl, H., and Schacht, E. J., "Methods of Analysis for the Caribbean Region." *Bull. Amer. meteor. Soc.*, 27: 569–575 (1946).
33. Riehl, H., and Shafer, R. J., "The Recurvature of Tropical Storms." *J. Meteor.*, 1: 42–54 (1944).
34. Rossby, C.-G., "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6: 163–180 (1949).
35. Sawyer, J. S., "Notes on the Theory of Tropical Cyclones." *Quart. J. R. meteor. Soc.*, 73: 101–126 (1947).
36. —— "The Significance of Dynamic Instability in Atmospheric Motions." *Quart. J. R. meteor. Soc.*, 75: 364–375 (1949).
37. Schacht, E. J., "A Mean Hurricane Sounding for the Caribbean Area." *Bull. Amer. meteor. Soc.*, 27: 324–327 (1946).
38. Simpson, R. H., "On the Movement of Tropical Cyclones." *Trans. Amer. geophys. Un.*, 27: 641–655 (1946).
39. —— "A Note on the Movement and Structure of the Florida Hurricane of October, 1946." *Mon. Wea. Rev. Wash.*, 75: 53–58 (1947).
40. Solberg, H., "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones." *P. V. Météor. Un. géod. géophys. int.*, Edimbourg, 1936. II, pp. 66–82 (1939).
41. Stommel, H., "Entrainment of Air into a Cumulus Cloud." *J. Meteor.*, 4: 91–94 (1947).
42. Wexler, H., "The Structure of the September 1944 Hurricane When Off Cape Henry, Virginia." *Bull. Amer. meteor. Soc.*, 26: 156–159 (1945).
43. —— "Structure of Hurricanes as Determined by Radar." *Ann. N. Y. Acad. Sci.*, 48: 821–844 (1946–1947).
44. Yeh, T.-c., "The Motion of Tropical Storms under the Influence of a Superimposed Southerly Current." *J. Meteor.*, 7: 108–113 (1950).

POLAR METEOROLOGY

Antarctic Atmospheric Circulation *by Arnold Court* 917

Arctic Meteorology *by Herbert G. Dorsey, Jr.* 942

Some Climatological Problems of the Arctic and Sub-Arctic *by F. Kenneth Hare* 952

ANTARCTIC ATMOSPHERIC CIRCULATION

By ARNOLD COURT[1]

Meteorologist, U. S. Antarctic Service, 1939–1941

INTRODUCTION

Antarctica's atmospheric circulation and its causes and effects are major gaps in current meteorological knowledge. Until they are filled, there can be no complete understanding of the weather processes of the earth as a whole, and especially of the radiative balance of the earth's surface and upper atmosphere and of the mechanics of variations in the general circulation. Yet there is almost no direct information on the antarctic circulation, because Antarctica is a continent, covering one-ninth of the land area of the world, for which the only meteorological data apply to a dozen isolated spots on the periphery.

This paper discusses previous concepts of the circulation over the interior and over the surrounding oceans, and of "pressure waves," then cites the additional observational material gained since these concepts were enunciated, and finally suggests a general circulation scheme which agrees with these newer data, especially concerning the upper air and the topography.

"Antarctica" is the south polar continent, and "antarctic" applies only to it and its immediate offshore islands and seas. The open ocean surrounding the continent, with a dozen or so islands, is "subantarctic." The outstanding features of the generally circular coastline are the narrow, mountainous Palmer Peninsula ("Graham Land" to the British) which extends northward to within 600 miles of Cape Horn, and two deep embayments almost opposite each other: the Weddell Sea east of the Palmer Peninsula, and the Ross Sea south of New Zealand.

At the head of the Weddell Sea is the Filchner Shelf Ice (renamed "Lassiter" by Ronne [127]), descending from the high interior; on the east side, and around the continent through almost 180°, the ice descends rather sharply to the ocean, with mountains in several areas and a few ice-free spots among them, and several ice sheets projecting into the open ocean or the various bays. More than two million square miles in the interior are wholly unknown; they may be a rocky desert, they may contain the world's highest mountain, or they may in fact be a featureless snow plain averaging 10,000 ft above sea level, as is supposed because of conditions at the geographic and magnetic poles. There is also some evidence that this area may contain a large basin.

This largest unknown area in the world ends at the massive mountain chain which borders the Ross Sea on the west and south, and up whose intervening glaciers Amundsen and Scott toiled to reach the Pole. The southern half of the Ross Sea is covered by the vast Ross Shelf Ice (or Ross Barrier), whose face rises 100 to 200 ft above open water; it is generally grounded, although extensive portions near the edge are afloat. East of the Ross Sea are complex mountain areas, and farther east the coast, as outlined by United States aircraft in 1940 [121] and 1947 [116], is quite rugged, with many mountains, one reaching 20,000 ft, and numerous bays and glaciers. The adjacent waters are so ice-filled that no ship has even penetrated within sight of the coast. The interior is an "unbroken desert of snow" about 6000 ft high [120] with a few mountains near the base of the Palmer Peninsula.

Until 1950, only three expeditions had ever wintered on the rocks of the continent proper; all others were based on small islands just offshore, on board ships frozen in near such islands, or on solid ice sheets extending out from the continent. All bases have been practically at sea level—at the edge of a continent assumed to average several thousand feet in height. No one has ever been more than a few miles inland from March through September, and even in summer no spot in the interior has ever been occupied for more than a week. All expedition bases have been on the shores of the Palmer Peninsula or of the Ross Sea, except for two on the coast facing Australia.

Exploration of Antarctica's interior began with the "Homeric period" [125] from the first wintering on the *Belgica* in 1898–99 to Shackleton's unsuccessful transcontinental attempt in 1914–16. The continent was first sighted in 1820 (but whether by an American or a Briton—or even a Russian—is argued actively) and circumnavigated around 1840 by American, British, and French expeditions, all sighting land at various places and the British one, under Sir James Clark Ross, penetrating the encircling pack ice to reach the Ross Sea in two successive years.

The "mechanical period," beginning with Admiral Byrd's first expedition in 1928, has seen several American expeditions, one British venture, and in recent years the large-scale occupation of the Palmer Peninsula by rival British, Argentine, and Chilean weather stations. During 1950, these stations were augmented by two expeditions, French and Norwegian-British-Swedish, based respectively south of Australia and south of Africa, and by weather stations on several subantarctic islands (Table I).

Most of the extant theories of Antarctica's weather processes are based on the observations of the "Homeric expeditions," generally published in detail and often discussed extensively. The "mechanical expeditions" have provided the best upper-air information, but their

1. Climatologist, Research and Development Branch, Office of The Quartermaster General. This article represents the personal views of the author and not necessarily those of the Department of Defense.

surface observations have generally been far less complete than those of their predecessors.

When Meinardus [14] compiled his comprehensive "Klimakunde der Antarktis" in 1937 for the monumental Köppen-Geiger *Handbuch der Klimatologie*, he had available observations for one or more years at only twelve places on the continental margin (only six of them for as much as two years), for three ice-bound ships drifting slowly during the winter, and for four subantarctic islands. Since then, summaries of observations made earlier at three other places have been published [25, 88] and later expeditions have made year-long observations at two new places [30, 37, 40] and one previously studied [47], besides the Palmer Peninsula stations whose data have not appeared.

Climatic summaries for all available stations are given in detail by Meinardus, whose extensive discussion of the general climate and individual peculiarities can hardly be surpassed. But recent information about upper-air conditions, antarctic geography, atmospheric dynamics, and oceanic synoptic meteorology makes it advisable to re-examine the older theories and to study critically the newer information on Antarctica's meteorology.

Table I. Antarctic and Subantarctic Weather Stations South of 50°S*

Lat. S.	Long.	Station name	General location	Index	Nationality
Antarctica Proper:					
71°03′	10°54′W	Maudheim‡	Kap Norvegia, Queen Maud Land	61903	Nor.-Brit.-Swed.
66°50′	141°25′E	Port Martin‡	Cap de Margerie, Adélie Land	95501	French
Palmer Peninsula:					
68°12′	67°03′W	Neny Fjord†	Stonington Island, Marguerite Bay	07023	British
65°15′	64°16′W	Argentine Island	Off West Coast Palmer Peninsula	88952	British
64°50′	63°31′W	Port Lockroy‡	Wiencke Island, off West Coast	88949	British
64°19′	62°58′W	1° de Mayo	Gamma Island, Melchior Archipelago	87970	Argentine
63°24′	56°59′W	Hope Bay†	Antarctic Sound, Northern tip Palmer Peninsula	10533	British
63°19′	56°54′W	Base O'Higgins	Cape Legoupil Northern tip Palmer Peninsula	85988	Chilean
West Subantarctic Islands:					
63°00′	60°30′W	Isla Decepcion	Deception Island, South Shetlands	87978	Argentine
62°56′	60°33′W	Deception Island	(Whalers Bay), South Shetlands	88098	British
62°30′	59°41′W	Bahia Soberania	Greenwich Island, South Shetlands	85986	Chilean
62°03′	58°23′W	Admiralty Bay	King George Island, South Shetlands	88934	British
60°44′	44°44′W	Orcadas del Sud	Scotia Bay, Laurie Island, South Orkneys	87981	Argentine
60°43′	45°36′W	Signy Island	South Orkneys	88925	British
South Atlantic Islands:					
54°16′	36°30′W	Grytviken	South Georgia	88903	British
54°16′	36°30′W	Georgia del Sud†	South Georgia	87992	Argentine
51°42′	57°51′W	Port Stanley	Falkland Islands	88890	British
East Subantarctic Islands:					
54°30′	158°57′E	Macquarie Island	Southwest of New Zealand	94998	Australian
53°06′	72°31′E	Heard Island	South Indian Ocean	94997	Australian
52°32′	168°59′E	Campbell Island	South of New Zealand	93944	New Zealand

* Except for seven stations in South America between 50°S and 55°S.

† Active in 1949, but not in 1950.

‡ Active in 1950, but not in 1949.

Most positions were taken from IMO Publication No. 9, and may differ somewhat from those given in other sources. Index numbers were taken from the same publication, except that the numbers for the active British stations are revisions to be effective in 1951.

THE ANTICYCLONE

Antarctica's basic atmospheric circulation pattern is generally assumed to be anticyclonic at the surface and cyclonic above, but there has been marked diversity as to the areal and vertical extent of the anticyclone, and even as to its permanence. Until the end of the 19th century, meteorologists could not agree whether pressure at the Pole was high due to cold or low due to "the excessive equatorward centrifugal force of the great circumpolar whirl." Expeditions early in the 20th century found a barometric trough around 60°S with pressure increasing southward, but Shaw [110] remarked in 1909 that "The Antarctic anticyclone, if it exists,

is a comparatively superficial effect attributable to the surface cold."

Others, however, postulated a vast continental anticyclone around which rotate a succession of huge oceanic cyclones [103, 105]. To this anticyclone Hobbs gave a unique character and a distinctive name [104]: a "fixed glacial anticyclone, modelled on the form of an hour-glass, . . .roughly centered near the South Pole," in which a continuous downdraft supplies air which flows out radially in bursts of strong winds. The vast snowfall presumably required to maintain the icecap against its own slow outflow and the outward sweeping of the winds comes from "the ice spicules of the cirrus and other closely related cloud forms . . . adiabatically vaporized in the downdraft and reprecipitated as they approach the glacier surface," which is radiatively cooled. This precipitation mechanism did not appear quantitatively sufficient to most other meteorologists studying Antarctica, and the analogous permanent anticyclone of Greenland is not accepted *in toto* by many meteorologists today.[2]

Meinardus [107] interpreted the strong easterly winds common at the *Gauss*, frozen in for a year some 40 miles off the coast at 90°E, on the 1902–4 German expedition, as due to passages of cyclones to the north, and felt that "If there is an Antarctic anticyclone it can exist only in the inner portion."

> The Antarctic anticyclone, so much discussed in the past, is a pressure distribution peculiar to the lower atmospheric strata only, appearing with distinctness only in the sea-level pressure distribution. On the other hand the low Antarctic temperature must produce such a rapid vertical decrease in pressure that above a certain level the Antarctic pressure must be lower and not higher than that of surrounding regions. Thus the sea-level anticyclone must be overlain by a cyclone, the so-called "polar whirl" in the general circulation of the globe.

From available information, Meinardus estimated that this transition occurs at about 2 km above sea level, and that most of the continental interior is higher, at the level at which cyclonic conditions prevail, with moisture-laden air flowing in east of the Ross and Weddell Seas, and a cold dry outflow on their western sides. This concept was attacked by Simpson, meteorologist of the last Scott expedition of 1911–13, who insisted [112] that a continental mass like Antarctica would establish its own regime above its surface, and not penetrate into the circulation like an isolated mountain peak:

> A statement of the general air circulation over the Antarctic is now quite simple. Over the snow-covered surface of the Antarctic whether at sea-level or at the height of the plateau radiation is so strong that the air is abnormally cooled especially in the layers of air immediately above the surface. This cooled air is heavier than the surrounding air and therefore the pressure increases from the exterior to the interior of the Polar area; in other words the pressure distribution is anticyclonic and the air motion is in general outwards. Above each anticyclone a cyclone forms on account of the relatively rapid vertical pressure change caused by the cold dense air. These cyclones convey air from higher latitudes over the Polar region and supply the air which passes outwards near the surface. In the normal steady state the air circulation takes place slowly and th descending air is warmed up dynamically so dissolving cloud and giving clear cloudless skies, thus accounting for the decreasing cloud amounts observed as one penetrates the Antarctic.
>
> The clear skies in their turn facilitate radiation as also does the small absolute humidity of the air. In consequence the air and the snow surface become abnormally cold and there is a great tendency to the formation of temperature inversions especially in the lower atmosphere . . . . The abnormally cold surface air is forced upwards in these [forced ascending] currents, rapidly cooled in the ascent, and the water obtained is precipitated as snow, which when combined with the high surface winds produces the typical Antarctic blizzard.

Simpson also disputed Meinardus' interpretation of the *Gauss* winds, considering them too constant in force and direction to be due entirely to cyclonic passages. His general model was followed, in preference to that of Meinardus, by Barkow, meteorologist on the *Deutschland* during its year-long drift in the Weddell Sea, and Knoch, editor of the report [101] after Barkow's death:

> The circulation over the Antarctic continent is dominated by an anticyclonic cap of air which probably flows outward in a series of waves moving in a south to north direction. Above this cap of cold air there is a cyclonic stratum connected with the circulation of temperate latitudes and in this wandering depressions travel in a west to east direction. The two air-systems mutually influence each other; on the whole the lower system dominates the upper at least so far as atmospheric pressure is concerned. The surface winds are dominated on the whole by the lower system for the most part on the continent and in less degree as the distance from land increases. In the Cirrus region, and especially in the stratosphere, cloud movements appear to show that there is a current of air moving right across the continent from the Indian Ocean to the region of the West Antarctic. . . .

However, Shaw [111] evened the score by abandoning his thin anticyclone of twenty-five years earlier in favor of Meinardus' model rather than that of his own colleague, Simpson, or that of Hobbs:

> The circulation would correspond with a permanent anticyclone covering the continent if the space which the anticyclone should occupy were not already filled by a huge mass of land. . . . We should not expect to find evidence of the so-called [glacial] anticyclone on the top of the land-mass; the conditions there may be the reverse of those which are indicated by the margins at sea-level.

In turn, Kidson, although opposing Simpson's theories on some points (see p. 927), accepted his surface anticyclone rather than Meinardus' peripheral sea-level anticyclone or Hobbs's glacial anticyclone. In his discussion [75] of the results of Mawson's 1911–13 expedition (he had previously edited and analyzed the data of Shackleton's 1907–9 expedition), he said:

2. Consult "Some Climatological Problems of the Arctic and Sub-Arctic" by F. K. Hare, pp. 952–964 in this Compendium.

The cold surface of the Antarctic continent is continually removing air from the general circulation. This lies over it as a cold layer of varying thickness which flows outward to the periphery under the influence of gravity, giving the well-known katabatic winds. . . . The cold air accumulates especially along the coast of the continent where it appears to raise the air pressure by about 0.2 inch above what is observed over the open sea, far from the shore, in corresponding latitudes (for example, over the Ross Sea). It is responsible for a prevalence of easterly winds along the antarctic coastline and its effect extends, on the average, for about 200 miles from the coast. Pressure systems, fronts, waves, cyclones, etc., move to a large extent, over and independently of the cold layer. Fluctuations in the katabatic flow are, of course, produced.

Seasons. These various models of Antarctica's circulation make no great distinction between summer and winter, yet all seem in general to be based on winter conditions. Meinardus [14] did suggest that the north-south pressure gradient above 2 km is especially strong in winter. As will be explained later, it is likely that both the temperature and pressure gradients aloft reverse from summer to winter, and it is gross oversimplification to say, as did Haurwitz and Austin [133], "Even though the south polar sea-level anticyclone is very pronounced, the strong north-south temperature gradient probably results in a polar cyclone at low altitudes."

By contrast, most of the comments of recent years imply some seasonal variation in Antarctica's circulation. Thus Serra and Ratisbonna [42] considered that "The cold antarctic anticyclone [is] greatly weakened in summer and forced back toward the pole," and that the upper cyclone is found above 2 km in summer, and above 3 km in winter. Coyle [27], like them primarily interested in antarctic influences on South American weather, reasoned similarly. Gentilli [9] recently assumed that "the area of prevailing polar easterlies contracts in summer and expands in winter."

Meteorologists of the U. S. Navy's Operation Highjump[3] during the summer of 1946–47 suggested [2] that "The south polar anticyclone is broken up into a number of cells with their centers displaced to the northward." Two such cells, representing persistent outflows of cold air, were found around 120°W and 145°E, plus a narrow wedge along the Palmer Peninsula (65°W), a semipermanent wedge along 80°E, and a migratory or transitory wedge between the MacKenzie and Weddell Seas (20°W to 70°E). The 145°E position was given by the Ross Sea group; the western group placed this "Balleny Island wedge" along 155°E.

No such cells had been shown by Meinardus [14] on his map of mean annual pressures, but Lamb's revision [83] suggested three areas with pressures greater than 995 mb centered on 80°S, at 20°E, 130°E, and 90°W. Earlier Lamb concluded [81] that "we should think of the south polar anticyclone not as a permanency but as a recurrent feature subject to most of the same laws of life and decay that apply to high pressure systems everywhere."

Upper winds at Little America during two full years [50] show the circulation to be far more complex than would result from a permanent surface anticyclone with upper cyclone, or any other simple model. Grimminger, who tabulated and summarized the data, said [49]:

> The annual means show very clearly winds with south and east components at the low levels and with north and west components at the higher levels and agree qualitatively with the concept of the outward drainage of cold air below and a compensating inflow aloft. . . . The south components are distinctly larger in winter than in summer up to 1,000 m; this is even more pronounced during the coldest period and extends to 2,000 m. In the upper layers. . . there are larger north components in winter than in summer at 7 and 8 km and in the coldest period from 5 to 9 km.
>
> In general, the seasonal variation of the north-south components agrees with what would be expected if the circulation over the continent were controlled mainly by the drainage of the cold lower layers. The seasonal variation of the east-west components in the upper layers also agrees with this; but that of the lower layers does not, since as pointed out above the east components are smaller in winter and are lacking entirely during the 3 coldest months. . . . Just why these east components do not appear in the coldest months when the drainage should be a maximum is not clear.

From the same data, Palmer [62] found deep easterly winds, *i.e.*, continuously between north-northeast and south-southeast up to 5 km, on 38 per cent of all days with flights to 5 km or higher, and on 44 per cent of days with flights to 3 km or higher. These easterlies aloft "are far more frequent than would be expected from the theory of the polar vortex."

It is obvious from its derivation that the idea of a "polar vortex" is an abstraction, adequate, perhaps, as a summary of average conditions, but misleading if applied to the analysis of individual synoptic situations in the far south. To find, by the manipulation of mean values, that the antarctic anticyclone is on the average a shallow pressure feature may be justified, but to deduce from this that all easterly, southeasterly and southerly winds, and in particular the blizzards, are "katabatic" or are shallow surface flows is to obtain a result that is really beyond the capacity of climatological methods.

Similarly, Meinardus pointed out repeatedly that mean annual pressure maps are merely averages of daily weather patterns, and [14] that "the direction and force of the air motion on the inland ice are governed not only by the direction of the slope but also by the general pressure distribution," which changes from day to day.

The 426 pilot balloon and rawin ascents made on Highjump [2], operating at sea all around the continental margin, were summarized for three levels: "2,000

3. While the full aerological reports of Operation Highjump [2] and Task Force 39 [4] are classified as "Restricted" by the Navy Department, the actual observations [3] are unclassified, and the discussions and conclusions in the Restricted reports are considered unclassified, so that one unclassified summary [1] of conclusions has appeared. The generous co-operation of Capt. H. T. Orville, Chief of Aerology, permitted access to the full reports and obtained official clearance for their use and quotation here.

m, predominantly under the influence of the polar anticyclone; 3,000 m, the zone of transition between the low-level polar anticyclone and upper-level cyclone; and 5,000 m, definitely under the influence of the polar cyclone. The winds show a steady increase in velocity with heights to a marked maximum at the tropopause, 25,000 to 30,000 ft, [and also] a slow increase in average velocity as fall approached, coinciding with an increase in the intensity and rate of movement of low-pressure areas."

The Interior. Only four areas of Antarctica more than 100 miles from the coast have been explored on the surface: two converging routes to the South Pole, two routes from opposite directions to the south magnetic pole, the plateau due west of McMurdo Sound, and the center of Ellsworth Highland midway from the Palmer Peninsula to the Ross Sea. In all four areas, midsummer weather is variable, with clear days and cloudy, strong winds and light, snowfall and ablation.

Prevailing winds observed by parties in these four areas are shown in Table II, adapted from Taylor [114], who considered both winds and *sastrugi*, the ridges or dunes of snow which indicate wind direction. Since convergence of the meridians makes it difficult to compare directions in different longitudes, directions are indicated by the parallel direction at the South Pole, measured in degrees eastward from Greenwich.

Table II. Prevailing Winds of Interior Antarctica in Summer.
(Referred to directions at the South Pole: 90°E = 90°, 90°W = 270°)

Party	Year	Area	Wind
Mawson and Bage	1912–13	Adélie Land to the magnetic pole	300°
David	1908–09	Victoria Land to the magnetic pole	300°
Scott	1903–04	West of McMurdo Sound	360°
Shackleton	1908–09	South of Beardmore Glacier	320°
Amundsen and Scott	1911–12	South Polar Plateau	326°
Ellsworth	1935	Hollick-Kenyon Plateau	20°

These half-dozen average wind directions might indicate a general surface air motion in summer from the Weddell Sea toward the Ross Sea or Australia, opposite to the upper-air flow deduced by Barkow. They are compatible with a single surface anticyclone centered around 80°S, 70°E. By themselves, the prevailing winds around the magnetic pole and west of the Ross Sea are compatible with an anticyclonic cell centered around 70°S, 155°E, roughly the position given by the High-jump aerologists and by Lamb for one cell of the polar anticyclone. The south polar plateau winds similarly would agree with another lobe, or perhaps the central cell, centered around 85°S, 70°E.

Neither of these postulated cells, however, can be even semipermanent. "The wind," David [118] reported, "had helped us by blowing from the southeast, just before we reached the Magnetic Pole, and now it was blowing in the opposite direction, helping us home." On 22 January 1909, "a clear day with bright sunshine, the wind started soon after 5 A.M., constantly freshening, as it usually did in this part of the plateau, till about 3 P.M., then it gradually died down by about 10 P.M."

On the south polar plateau, wind and weather in December 1911 and January 1912 were similarly variable. There, Simpson's painstaking analysis [112] showed, 63 per cent of all the winds recorded by Amundsen and Scott, representing 75 per cent of the total movement, were from 320°, 340°, or 360°, with resultant 326° at 8.5 mph. But there were calms, winds from other directions, wide variation in the strength of the prevailing winds, and fresh snow as well as blowing snow.

Most surprising, and least compatible with a theory of anticyclone cells, are the winds observed by Ellsworth [120] at three camps (at 80°S between 104° and 115°W) made to wait out blizzards whose "hard, fine-grained" snow "is dry, fine as flour, sifts into everything, and packs hard as rock." The first landing was made when clouds appeared and visibility grew poor; after nineteen hours there, "the weather was fine, though the horizon ahead looked thick" and after 30 minutes' flying, weather forced a second landing.

"Three days of varying thick and clear weather" ensued; after a 50-minute flight "the weather became so thick we could scarcely see to land" and a three-day blizzard followed, with east to southeast winds estimated at 45 mph. On the fourth day, "December 1, the storm moderated . . . with intervals of sunshine . . . . Buffeting wind, snow, and thick weather . . . bitterly cold" came on the fifth day, and heavy snow that night. On the sixth day, another "heavy storm broke from the southeast—thick snow and a high wind.".

Finally on the seventh day, although "the weather was anything but promising, with thick horizons all around and a sullen sky overhead," the flight was resumed, and an hour later they encountered "blue sky with a clean horizon all around" at the western edge of the plateau. "Evidently the storm area had been hanging stationary near the western edge" of the plateau. Both in the air (usually at 10,000 ft, some 4000 ft above the surface) and on the ground, during these eleven days of the only visit to date to this area, all winds were from east or southeast, except for two brief periods of north wind; "it never did blow from the west"; south winds were first encountered as the Ross Shelf Ice was reached.

These weather conditions are similar to those along the coasts and seem out of keeping for the center of an "unbroken desert of snow" 6000 ft above sea level. Ellsworth thought the then unknown coastline was "450 miles to the north," but explorations in 1940 [121] and 1947 [116] revealed a large bay or a large island-choked, ice-filled gulf at the head of Amundsen (formerly Roosevelt) Sea. This indentation, as yet unnamed and unmapped in detail from the available aerial photographs, appears to reach to about 77°S at 105°W, or only some 200 miles from Ellsworth's first camp, explaining "the water sky observed by Hollick-Kenyon on Nov. 23 in the conjectured position of 76°S and 100°W."

If this bay or gulf is as large as preliminary reports indicated, the weather sequence could readily be explained in terms of a family of occluded cyclones entering from the north and stagnating there. In fact, the weather reported by Ellsworth is evidence that this coastal feature is extensive. There is a strong possibility that the general pressure field during this period was considerably lower than at Dundee Island, where the flight started in good weather (high pressure), so that the altimeter readings of the heights of the camps (and the various mountains) may be too high by as much as 1000 ft or even more.

At any rate, the variable winds, chiefly southeast, recorded by Ellsworth from 23 November to 4 December at 80°S, 104° to 115°W, do not fit Lamb's anticyclonic cell at 80°S, 90°W [83], and are hardly compatible with "the anticyclone located near 120°W" inferred by the Highjump aerologists [2] twelve years and two months later. Perhaps those two months are significant, and the anticyclone, varying from 90°W to 120°W, develops only in late summer; perhaps the 1946–47 summer saw a northward displacement of an anticyclone which in late 1935 was centered much farther south.

These are some of the problems of Antarctica's anticyclone, which exists only around the edges according to Meinardus and Shaw, is a thin layer according to Simpson, Barkow, Kidson, and Grimminger, is broken up into several cells according to Lamb and the Navy aerologists, shrinks markedly from winter to summer according to Serra and Ratisbonna, Coyle, and Gentilli, and is a major feature fed by an upper cyclone according to Hobbs, with Palmer considering the upper cyclone to be an abstraction derived from averages of widely varying conditions.

CYCLONES

Mariners encountered and named the "roaring forties" several centuries ago, and later reached and named, with more alliteration, the "furious fifties" and "shrieking sixties." But only slowly has the meteorological nature of these subantarctic "brave west winds" become known, and there is still much to be learned about them. The type and extent of the frontal zones surrounding Antarctica are far from established, the number and character of air masses are uncertain, and the formation, travel, and dissipation of subantarctic cyclones are still subjects of speculation. Yet understanding of these processes is fundamental to any theory of the meteorology of Antarctica, because almost all the weather phenomena thus far encountered (around the coasts, the only part so far studied) appear related to systems impinging on the continent from the surrounding seas.

Knowledge of weather processes over the southern oceans has kept pace with Northern Hemisphere findings as well as the scanty observations have permitted. As late as 1929 Barlow [102] said that the subantarctic "depressions travel from west to east around the globe in nearly unbroken succession," as had been similarly assumed for the Northern Hemisphere. Emergence of the concept of cyclone life cycles and cyclone families was traced in detail by Palmer [89], as background for his preliminary (1942) theory of subantarctic meteorology.

This theory appears in more mature and concise form in the *Handbook of Meteorology*, in an article [18] credited to "Civilian Staff, Institute of Tropical Meteorology, Rio Piedras, Puerto Rico," but actually written in 1944 by Palmer. It describes quasi-permanent anticyclones lying along 30°S in the central South Atlantic and Indian Oceans and the eastern Pacific Ocean, with "a train of warm migratory anticyclones" between each pair; between these migratory anticyclones are found "inverted **V**-shaped" troughs extending northward from great closed cyclonic systems "somewhat similar to the stationary Icelandic and Aleutian lows of the Northern Hemisphere but, unlike those depressions, moving from west to east with the anticyclones to the north."

Even more extensive than Palmer's first treatise was Kidson's series [75] of daily weather maps for 1912 for all Australia and New Zealand and the region southward almost to the Pole, using data of the Australasian Antarctic Expedition and others. A pioneer in the application of frontal analysis to the Southern Hemisphere, he drew very involved systems to account for all the meagre observations, but did not "make analysis conform very strictly with kinematic laws" and tended to ignore the contrast of sea and land.

Kidson died on 12 June 1939 without writing the final summary discussion of his work. His charts and analyses, made during the 1930's but not published until 1946, seem out of date, but as Gibbs [73] pointed out, "he was a pioneer of frontal analysis in the southern hemisphere and was using a new and somewhat unfamiliar tool of analysis, with a particularly scanty observational network. . . .His work should be accorded recognition for its pioneering aspect but his conclusions may be discarded where later, more soundly based evidence disproves them." Gibbs also emphasized that he studied conditions throughout the year while most other reports deal with summer conditions only.

Three extensive reviews [67, 72, 80] of Kidson's work agreed as to its historical value, but took issue with various aspects. Lamb cited summertime "evidence of a distinct change of regime in the atmospheric circulation in the higher southern latitudes east and west of about the hundredth meridian. Points west of 100°E come under the influence of systems circulating southwards from the Indian Ocean and often recurving towards the west in their later stages," involving a westward movement which Kidson did not conceive.

Fronts. Kidson felt that his charts proved that there is no "continuous polar front encircling the southern hemisphere in high latitudes"; G.C.S. (Simpson?) [72] doubted this conclusion. Meinardus [14] was not sure that the boundary between continental and marine air "should be considered as the polar front," but conceded that such boundaries were quite marked on the east

side of the Palmer Peninsula, along the Adélie Land Coast, and near McMurdo Sound, and must influence weather processes there.

Actually, there are two frontal zones in the subantarctic oceans: the antarctic front, encircling the continent at least in segments, and the southern polar front, trending diagonally southeastward across each ocean and interrupted over the continents [41, 70, 133]. Palmer [18] identified

> . . .three major frontal systems in the Southern Hemisphere (1) that which runs from a point a little south of Tahiti southeastward along the southern boundary of the quasi-stationary subtropical anticyclone of the South Pacific, (2) that which crosses South Africa somewhere between 30 and 40°S latitude and extends southeastward on the poleward side of the South Indian anticyclone, and (3) that which lies in the vicinity of the mouth of the Plata River and stretches across the South Atlantic Ocean and far to the south of South Africa.

While the southeastern ends of these polar-front segments enter subantarctic waters and may even reach the continent, they are less directly connected with Antarctica's circulation than the antarctic front, about which far less is known, especially in winter. For the area south of Cape Horn, Serra and Ratisbonna [42] postulated that

> The Antarctic front follows a SW-NE direction. It passes from the Pacific to the Atlantic, extending from the pressure trough of the Belgica [Bellingshausen] Sea around through the low of the Weddell Sea. It is most intense in winter when the gradient is particularly steep and the circulation most active because of the deepening of the depressions over the Antarctic seas.

This last assumption was not accepted by Coyle [27]. Allowing for wintertime semipermanent lows in the Weddell and Ross Seas, he said:

> We get a very sinuous picture of the Antarctic front during the winter months and here we meet one of the strange Southern Hemisphere facts, namely, that contrary to expectation there seems to be a lesser easterly stream behind the front in the wintertime than there is in the summertime. It is during the summertime that the Weddell and Ross Sea lows disappear, thus theoretically allowing the Antarctic front to straighten out and recede poleward as the summertime weakening of the Antarctic anticyclone becomes effective. . . .
>
> We therefore come to a picture of the Southern Hemisphere wintertime westerly band (30° to 60° south) being composed of two parts, an equatorward warm westerly stream between 30° and 45° south and a cool band of transitional Antarctic air, seen as a west to southwest stream between 45° and 60° south. The Southern Hemisphere "Polar" front then would seem to be this boundary between the . . . currents. . . .
>
> It seems that the cold southwest streams which invade the [South American] continent are breaks through the Antarctic front of cold air drawn up about the westward side of the cyclones, and these streams, with continental transport equatorward, become deflected to south and southeast streams to form the migratory anticyclones which then move up over Argentina and Brazil. It is only when this cool transitional Antarctic air moves north of latitude 45° that it becomes well defined and the "Polar" front becomes well developed as it moves against the tropical continental masses. . . . It is difficult to say just when the leading edge of a cold mass ceases to be the Antarctic front and becomes the Polar front.

On their January and July maps, which extend only to 70°S, Haurwitz and Austin [133] showed an antarctic front northwest of the Palmer Peninsula in July, northeast of the Weddell Sea in January.

No continuous antarctic front, but only segments of it, were found by those in the best position to observe it, the aerologists [2] of Operation Highjump; on their mean surface pressure chart, "The mean position of the zone of convergence between the polar easterlies and maritime westerlies, [which] represents the antarctic front, is not accurate because the methods of summarizing and presentation of data obscure the minute details of the pressure trough." In summer, semipermanent segments of the antarctic front were found along the pack-ice edge in the Bellingshausen Sea and south of Australia, extending west from 170°E along 65°S.

Each section was either warm or cold, depending on the relative strengths of the warm moist air to the north and the cold dry air to the south. The segment south of Australia was normally 50 to 150 miles north of the ice pack, sometimes 500 miles north, and occasionally pushed onto the continental plateau itself by extremely strong northerly circulation. Strong outbursts from the anticyclone along 145°E, which was bounded by this front, at times drove it eastward to trend southeastward across the Ross Sea. As a cold front, it had a line of snow showers, a wind shift of as much as 90°, and a temperature drop of 2 to 3F, occasionally 5 to 7F. As a warm front, it had a well-defined cloud shield, but without cirriform clouds, and snow showers ahead of the snow shield.

As long as the anticyclone located along 120°W during the 1947 summer was strong, its cold air mass did not "meet a barrier or front that would separate it from the warmer air mass to the north," but when it was reestablished after an intense low had penetrated the southeastern Ross Sea, a weak cold front formed near 65°S.

The following summer another group of aerologists found [4] "the antarctic front . . . observed on numerous occasions . . . to be very similar to the Equatorial front . . . a zone of convergence of two air masses with similar properties. The slope of the front varies according to which air mass is the colder."

Air Masses. The Highjump aerologists operating south of Australia assumed that the cold dry polar continental (*cP*) air of the high antarctic plateau was transformed by slight humidifying of the lower layers and subsidence aloft into continental antarctic (*cA*) air of the coastal area and solid pack ice, then into maritime antarctic (*mA*) air by rapid humidification as it passed over open pack ice or open water for 50 to 100 miles, after which it became fresh polar maritime (*mP*) air, with "the antarctic front the northern boundary of this *cA* or *mA* air mass."

Of the eleven air masses affecting South America, distinguished by Serra and Ratisbonna [42] in their pioneer study, "the originally continental mass, *A*, changes into *At* (transitional Antarctic) and finally into *Pm* (polar maritime)" as the air is humidified and warmed in leaving the continent, generally west of the Weddell and Ross Seas. In winter, *Pm* stagnating over Patagonia becomes continental polar air (*Pc*); active polar masses (*Pk*) moving northward were distinguished from returning polar masses (*Pw*) returning poleward over South America.

Although basing his discussion to some extent on this study, Coyle [27] did not use these names or indicators, preferring to consider currents rather than air masses. Apparently without reference to this or other previous Southern Hemisphere classifications, Robin [38] identified three air masses in the area south of Cape Horn: "Antarctic air (*A*) . . . over the continental and barrier ice of Antarctica, Polar maritime air (*Pm*) . . . in the southern half of the westerly wind belt of the southern hemisphere, [and] a more temperate type of maritime air (*Tm*) was also occasionally experienced, this air coming from a more northerly portion of the westerly wind belt." The antarctic front separated *A* and *Pm* air, and was strongest when it lay along the edge of the pack ice.

Generalizing at second hand, Haurwitz and Austin [133] concluded:

> The air masses of the continental antarctic resemble the *cP* air of northeastern Asia. The principal differences are the high frequency of steep lapse rates near the stirred surface layer and the very cold temperatures at all levels over the southern continent. During summer, antarctic *cP* air is still very cold and, except for a gradual rise in temperature, appears to resemble the winter air mass. Throughout the year, low specific humidities are characteristic of the southern *cP* air masses. . . .
>
> Because of the high topography, *mP* air masses affect only the periphery of Antarctica, where the cyclonic circulation is reasonably intense in summer and winter. Consequently, the summer air mass probably lacks the stability of the Northern Hemisphere air mass. Therefore, the lapse rate is likely to be approximately moist adiabatic throughout the year. Besides increasing the temperature and cloudiness of the coastal area, *mP* air must also supply the moisture for the light precipitation of the interior plateau.

Six air masses were deduced for the Southern Hemisphere by Gentilli [9] by applying to Shaw's classic maps of mean temperature and pressure [111] the air-mass criteria in use in the Northern Hemisphere:

Air Mass		*A*	*Pm*	*Tm*	*Eq*	*Tc*	*Pc*
Area of source in millions of square miles	summer	8	20	34	34	?	
	winter	11	30	31	16		?

The greatest contrasts, and hence most active fronts, occur between the polar maritime air (found between 40° and 68°S in summer and between 34° and 65°S in winter) and tropical continental air (found typically in Australia in summer). Polar continental air occurs only in South America in winter, equatorial air is formed as far as 24°S in summer, and superior (*S*) air is presumed to descend under certain conditions.

While conflicting in notation and definition, these various air-mass classifications can be harmonized. Despite Kidson's warning [75] that there is no "air mass to which the application of the term 'antarctic' could be justified," there seem to be three such masses: continental antarctic (*cA*) in the interior, transitional antarctic (*nA*) along the coasts, and maritime antarctic (*mA*) over the pack ice and the ocean south of the pressure trough, convergence zone, or antarctic front. To the north is a vast mass of maritime polar (*mP*) air. The symbols of various authors apparently fit this classification as follows:

	cA	*nA*	*mA*	*mP*
Highjump	*cP*	*cA*	*mA*	*mP*
Serra and Ratisbonna	*A*	*A*	*At*	*Pm*
Robin	*A*	*A*	*Pm*	*Tm*
Gentilli	*A*	*A*	*A*	*Pm*
Haurwitz and Austin	*cP*	*cP*	*mP*	*mP*

Proof of the existence of these masses (perhaps *nA* is really *cAk*), and better definitions and criteria, must come from meteorologists drawing regular synoptic maps of extensive antarctic and subantarctic areas.

Storms. Waves formed on the antarctic front west of the Ross and Bellingshausen Seas, the Highjump aerologists found, moved eastward and occluded, often merging with deeper systems which had approached from the northwest. Some antarctic frontal waves dissipated near the Balleny Islands, others curved southeastward into the Ross Sea and dissipated before reaching its southeastern corner; in the Bellingshausen Sea they merged into old Pacific Ocean occlusions. These storms were minor compared with those from the north:

> Wave formations developing on the polar fronts of the Atlantic, Indian, and Australia-New Zealand areas occlude and approach the continent usually on an east-southeast track [at 20 to 25 knots in January, 35 to 45 in March]. During periods of low circulation index these lows, upon approaching the continent, tend to stagnate in. . .semi-permanent low pressure areas. . .in the Ross Sea, near 120°E along the edge of the ice pack, in the western part of the MacKenzie Sea [60°E], and in the eastern part of the Weddell Sea. During high index conditions these lows stagnate only intermittently in these areas. . . .
>
> When a migratory low pressure center stagnated in the Bellingshausen Sea, a new center developed off the northern tip of the Palmer Peninsula,. . .deepened slowly but moved rapidly off to the eastward. . . . A northward outbreak of cold continental air [occurred] along the east coast of the Palmer Peninsula and behind the low as it moved into the Weddell Sea.

These summertime conclusions are reinforced by those reached during the ensuing months by Robin [38], meteorologist in charge of the British station at Signy Island, South Orkneys, during 1947–48 (and currently with the Norwegian-British-Swedish expedition): "Cyclones observed in the Antarctic front were either well-developed cyclones which approached Graham Land and from the west, or were young cyclones which formed near the north tip of Graham Land, the latter being more in evidence in late autumn and spring." Both types moved rather rapidly (about 600 miles per day) east-

ward and not, as previously presumed, southeast into the Weddell Sea. Both young and old cyclones occur in groups of twos and threes followed by an "injection of a cold air mass over the South American continent" which may reach Brazil, as others [27, 32, 33, 41] have found.

Mean monthly pressure charts for the same area, based on unpublished observations of Rymill's 1935–37 expedition [40] and available reports from South America and the South Orkneys, indicated to Mirrlees [36] either that the main track of depressions is farther south than previously described or that there may be actually two such tracks.

Two such separate tracks were found in the southeast Indian Ocean by Gibbs [73] who studied the reports of Palmer, Kidson, and Operation Highjump, and 6-hourly maps of the southern Indian Ocean during the first year (1948) of complete weather observations at Marion, Heard, and Macquarie Islands. Gibbs concluded that depressions in this area originate north of 45°S, as wave cyclones in winter and spring but frequently as frontless cyclones in summer and autumn. They travel southeast, intensifying between Marion and Heard, occlude and retard after passing Heard, become old cyclones by the time they pass Macquarie, and "probably move to the Ross Sea area, where they weaken and finally disappear"; a second track runs from western Australia to Macquarie.

The Antarctic front is generally impelled northward into the latitudes of the westerlies by the approach of one of the great depressions to the vicinity of the Antarctic coastline. Subsequently on reaching lower latitudes wave development may occur on it and the waves so developing may become one of the great southern depressions.

Substantially similar concepts of fronts, air masses, and weather processes over the oceans south of Australia and New Zealand were reached independently by Japanese meteorologists[4] on board whalers in recent years.

Circulation. In Palmer's concept of Southern Hemisphere circulation [18], cyclones are formed in groups off the east coasts of continents by the interaction of polar maritime air of the westerlies and tropical air flowing southward around the three subtropical anticyclones. These cyclones travel southeastward, first in the general circulation around the highs and then with the westerlies, developing, occluding, and finally filling as they approach Antarctica. New depressions form as waves on their fronts and continue somewhat farther. This model implies that cyclones which actually reach Antarctica are the second or third generation of systems which formed originally in low latitudes several thousand miles to the northwest

Frontal zones were of less interest to Lamb [81, 82, 83] than the general circulation of the Southern Hemisphere. By the end of his three summer months in subantarctic waters he was drawing usable maps for the entire hemisphere, although he was unable to obtain reports from whalers other than those in his own group or from the contemporaneous Operation Highjump. Analyses were based on North Atlantic experience and the postulate that "stable, warm anticyclones exercise a controlling influence on the circulation patterns far beyond their own limits."

These 1947 summer maps showed groups of occlusions symmetrically located, with a pattern of six such groups spaced 60° of longitude apart (and six intervening subtropical anticyclones) characteristic of low-index conditions, four groups 90° apart with higher zonal index, and a hint that as winter approached the pattern would reduce to three groups 120° apart.

Earlier, Kidson [75] had found "a strong suggestion that the moving anticyclones originate as some periodic function of the upper atmosphere, or of the atmosphere as a whole, which is continually tending to produce them at intervals of about 45° of longitude." This was based on his daily charts of 1912–13, when the mean speed of anticyclones, 8.8° per day, "was certainly above normal"; "more anticyclones passed during the year than usual," and their "mean latitudes were ... unusually high." Such 45° spacing implies a hemispheric pattern of eight anticyclones and intervening occlusion groups and, by extension of Lamb's conclusions, an unusually low zonal index with greatly increased meridional exchange of air.

The period from 2 February 1912 to 31 January 1913, studied by Kidson, may well have been different in general characteristics from the late summer of 1946–47, on which Lamb based his analysis. In the Northern Hemisphere, the latter period was quite abnormal; Namias [137] found pressures in the Arctic basin somewhat below normal during December and January and greatly above normal in February and March. Throughout this winter, the zonal index was below average and the total mass of air over the Northern Hemisphere above normal. On the other hand, during the 12 months studied by Kidson, the Northern Hemisphere zonal index varied from slightly below average to exceptionally above average; on the whole it was nearly 0.6 m sec^{-1} stronger than average, or about one-third stronger.

These departures from average of the Northern Hemisphere zonal index are directly opposite to those suggested by Lamb's and Kidson's analyses, and it is tempting to think that the indexes vary in opposite senses in the two hemispheres. If the negative correlation between zonal index and total atmospheric mass north of 20°N, found by Brier [128], represents interhemisphere transfer, a complementary relation of the hemispheric indexes is to be expected; but it may merely represent mass transfer from low to high latitudes. Opposing any inverse relation between the zonal indexes of the two hemispheres is the belief that all climatic changes, from ice ages to unusual years, are due basically to similar changes in the atmospheric circulation; since the larger climatic variations were simultaneous

4. Two Japanese reports were received after completion of the bibliography: Oceanographical Section, Central Meteorological Observatory, "Report on Sea and Weather Observations on Antarctic Whaling Ground (1947–48)." *Oceanogr. Mag.* (Tokyo) 1: 49–88 (1949); "Report ... (1948–49)." *Ibid.*, 1: 142–173 (1949).

in both hemispheres, parallel variations in circulation would be expected.

As yet, no series of Southern Hemisphere maps is adequate to give zonal index data for comparison with the Northern Hemisphere, but current research should provide preliminary answers. With any correlation thus established between indexes of the two hemispheres, the Southern Hemisphere index can be estimated, from Northern Hemisphere values available back to 1899, for those periods of detailed antarctic observations. Pending such further study, the possibility remains that the two periods whose weather has been studied intensively, 1912 and the 1946–47 summer, were both rather abnormal, 1912 having a very low zonal index with more storms and more meridional air exchange than usual, and the 1946–47 summer having a very high zonal index with fewer storms and much less meridional exchange than usual. Furthermore, variations in the circulation must be considered in evaluating any meteorological data from Antarctica.

WAVES

Characteristics. Nonperiodic *waves* of atmospheric pressure, found on the barograms (actual or reconstructed from barometer readings) of all antarctic stations with an amplitude of about 0.6 in., are a very controversial phenomenon of antarctic meteorology. Only the nature of the polar anticyclone has aroused more discussion than these waves, which are defined by agreement as having a change from minimum to maximum of at least 0.2 in. or 5 mm (6.7 mb).

That such waves can be found is not denied; the argument is whether they are "true pressure waves traversing the upper atmosphere,... travelling outwards from the center of the antarctic continent," as postulated by Simpson [112] and supported by Loewe [87] and Lamb [81, 82, 83], or merely reflect the passage of ordinary fronts and depressions, as vehemently maintained by Meinardus [14], Reuter [98], Kidson [74, 75], Palmer [89], and Ramage [64].

Unfortunately, these waves, about whose cause and exact place of origin Simpson did not speculate, have been often confused with the long-period *surges*, for which he gave both cause and place of origin. These surges, found from 10-day running means of pressure, have a constant period of about one month, but decrease in amplitude about 0.162 in. per thousand miles as they travel outward in all directions from 80°S, 120°W, where the extrapolated amplitude is 0.677 in. They are "due to increases and decreases in the intensity" of the upper-level cyclone "with the centre of lowest pressure over the low-lying region which we have reason to believe exists in the Pacific quadrant of the antarctic"; the epicenter was determined so that the observed amplitudes would fit the assumed law of linear decrease with distance. The surges explain why monthly pressure departures throughout Antarctica tend to vary similarly, and inversely from pressure departures in South America and the southern portions of Africa and Australia; surges are not reflected in local weather phenomena.

The only visit thus far to the presumed epicenter of these surges, by Ellsworth [120] in the spring of 1935, encountered twelve days of variable bad weather, apparently from occlusions penetrating a deep bay to the north (see p. 921). Since it is likely that low pressure prevailed during this period, the daily pressure readings from Rymill's Argentine Island base (65°15′S, 64°16′W) and Laurie Island, the only antarctic and subantarctic stations operating during 1935, could be tabulated in 10-day running means to determine whether Ellsworth's visit corresponded with a surge minimum.

Waves and surges were discussed by Simpson separately, the only connection between the two being that the apparent line of wave propagation, extended backward, "passes very near to the position 80°S, 120°W about which the pressure surges were found to have their maximum intensity." Yet Kidson, Ramage, and Lamb have given the surge epicenter as that of the *waves*, whose origin Simpson did not particularize any more closely than "the center of the antarctic continent"; even Meinardus confused waves and surges.

Table III. Antarctic Pressure Waves

Location (westward)	No. of years*	Period (hours)	Amplitude (inches)
Framheim	1	163	.624
McMurdo Sound	4 S	152	.572
" "	5 L	152	.560
Cape Adare	1	119	.549
Macquarie Island	2	98	.66
Cape Denison	2 L	102	.57
" "	– M	108	.61
Queen Mary Land	1	131	.59
Gauss	1	122	.642
Kerguelen Island	1	69	.634
Deutschland	1	124	.61
Laurie Island, South Orkneys	–	91	.642
Snow Hill Island	2	107	.567
Belgica	1 S	124	.630
"	1 M	134	.63

* Values cited by Simpson (S), Meinardus (M), Loewe (L).

Average amplitudes and periods of waves found for all the earlier antarctic stations, and the number of years on which they are based, are given in Table III. Data of expeditions of the last two decades have not been analyzed for such pressure waves, largely because their hourly values have not been published, or even computed in many cases. Although from 1935 to 1949 there were nearly 40 station-years of observations in the Palmer Peninsula area by the British, Americans, Argentines, and Chileans, the only data published so far are brief monthly summaries for Rymill's 2-year expedition [40] and Ronne's recent wintering [37], and a note on the British observations [38]. Three years of hourly pressures for Little America have been published [47, 50], but they have not been analyzed for waves.

Movement. The crux of the pressure-wave dispute is the asserted northwestward travel of the waves. This direction of propagation is taken to prove that the pressure changes which they cause (and from which they are found) cannot be due to the eastward motion of subantarctic pressure systems. Constancy of wind speed

and direction, despite marked pressure fluctuations at several antarctic stations, notably the *Gauss*, has been cited to prove that these waves, and not cyclones, control the weather processes.

Northwestward travel of the waves was assumed by Simpson because of their progressively later occurrence at stations at greater and greater distances northwest of Framheim, the easternmost of the three 1911 Ross Sea stations. Superposition of barograms for the three stations shows good correspondence, with such time lags in occurrence of crests and troughs that Simpson concluded the waves "travel with a linear wave front . . . along a direction parallel to a line joining Cape Adare and Framheim" at about 35 mph. In all, Simpson found waves in the barograms of eight antarctic and subantarctic stations, but only the three Ross Sea locations were simultaneous and close enough to indicate the direction of travel.

From detailed analyses, Simpson concluded that west-moving pressure waves alter the existing local pressure distributions so as to cause blizzards and other phenomena, not readily explained otherwise, at various antarctic stations. This he considered stronger proof of the waves' existence than the apparent motion; for his own station, "it is difficult to see any close relationship between the winds at Cape Evans and the actual pressure waves, as would be the case if the pressure waves were due to the passage of high- and low-pressure systems."

Using data subsequently published for three stations to the northwest of the Ross Sea, Loewe extended the general westward movement by comparison of barograms for Macquarie Island, Cape Denison, and Queen Mary Land with each other and with barograms for Cape Evans. He also found a higher statistical correlation between pressures at pairs of these stations, assuming westward motion at the rate indicated by this comparison, than for eastward motion corresponding to the average travel times of subantarctic pressure systems.

However, since "the southwest-to-northeast direction of wave fronts in the Ross Sea, as found by Simpson, cannot be extended as far west as Cape Denison," Loewe postulated that "the crest and trough lines of the waves west of the Ross Sea run rather from west to east, the waves radiating outwards from the continent," so that the 35-mph speed would be maintained. But his waves reach Macquarie Island, twice as far from Cape Evans as is Cape Denison, too quickly for such a pattern, and his later finding that waves reach Queen Mary Land 19 hours later than Cape Denison, 1300 miles due east, requires either northwestward motion or rapidly accelerating westward progress.

Objections. Travelling pressure waves were deemed unnecessary by Meinardus, Reuter, Kidson, Palmer, and Ramage to account for observed conditions in Antarctica. Meinardus considered the waves found for the different stations as indicating the "unrest of the atmosphere," and insisted that the *Gauss's* weather phenomena were adequately explained by the eastward passage to the north of pressure systems which are not circular; Simpson had assumed circular systems in proving that the *Gauss* winds were too steady to be cyclonic. However, Meinardus admitted [14] that in a few cases pressure waves from the interior could account for the *Gauss* weather.

Palmer was not so generous: after reviewing Meinardus' discussion, he redrew some of the 1902 maps on the same patterns he used for a 1932 series based on whaler [96] reports and concluded that Simpson's "interpretation of the *Gauss* observations was, to say the least, far-fetched," and that "there is no evidence for the existence of Simpson's 'pressure waves' in the far southern Indian Ocean."

In his discussions of Shackleton's 1907–9 expedition and Mawson's 1911–14 data, as well as elsewhere, Kidson maintained that passage of complex frontal systems could account for all the observed pressure and weather phenomena, including the apparent westward motion of pressure maxima. This motion he verified for Simpson's original three stations by superimposing curves of the average pressure difference for each two hours before and after pressure maxima and minima at Cape Evans and the corresponding values at Framheim 9.1 hours earlier and at Cape Adare 10.7 hours later, finding "remarkably close relationship between the pressure changes at the three stations."

For the 1912–13 data, in which Loewe "found pressure waves radiating outward from the antarctic continent," Kidson agreed that:

> On the average, the changes at Adélie Land lag considerably behind those at Cape Evans. . . . The amount of this lag is, however, very variable. Sometimes it amounts to practically nothing and there are even occasions when changes occur earlier at Adélie Land than at Cape Evans. It would be extremely difficult to account for all phases of the relationship between the two stations on the basis of Simpson's waves. . . . It is not possible to account for the Queen Mary Land pressure variations as being due to waves which have passed Framheim, Cape Evans, Cape Adare, and Adélie Land. . . . [It is] still more difficult to imagine. . .that the waves which have reached Adélie Land 13½ hours after Cape Evans pass Macquarie Island only 7 hours later.

He explained the whole mystery by assuming that "the fronts are, on the average, inclined at a large angle to the meridian. In the Ross Sea region, indeed, they must lie almost east and west." Kidson's table of the mean travel times of fronts indicates "the average interval in hours after the passage of Queen Mary Land at which each station would be passed": Framheim 39, Cape Evans 48, Cape Adare 55, Adélie Land 57, Perth 68, Macquarie Island 114, etc.

All the phenomena which Simpson explained as due to pressure waves, Ramage attributed to the passage of Types *A* or *D* of the six types of cyclones which he found in a detailed single-station analysis of two years of Little America data; these two types, both moving eastward over the Ross Sea north of Little America and one (*D*) later curving southward, "make up 35 percent of the cases considered," and two pressure-wave examples cited by Simpson were, to Ramage, *D* types:

Fluctuations of the meteorological variables in the Ross Sea area can be attributed to a succession of depressions not differing materially from those of lower latitudes. . . .Apart from some brief periods, the pressure systems of the Ross Sea are connected with, and have a definite effect on, the systems further north. . . .Blizzards set in when the isobars (usually in the rear of a depression) coincide with the direction of the down slope winds.

Except for some rather forced analyses, Ramage's interpretations are rather convincing: there is an obvious relation between the Ross Sea cyclones and westward-moving pressure waves, but whether the relation is one of simple cause-and-effect or one of steering-and-intensification cannot be determined.

Suggestions. Little attention has been given to one aspect of the pressure waves: while their amplitude is constant (about 0.6 in.) throughout Antarctica and even at subantarctic islands, their period decreases steadily from Framheim to Cape Denison. (Simpson and Loewe confusingly called the duration in hours the wave length.) If these are classical waves, in which the product of wave speed and period gives wave length (distance, not duration), then waves travelling from Framheim to Cape Denison with period decreasing from 163 to 102 hours must either increase in speed or shorten in length, or both.

As Lamb has suggested, it is likely that Antarctica's coast is affected both by westward-moving pressure waves and "the familiar effects of fronts and depressions;. . .these apparently conflicting influences can be seen as interlocking parts of the same atmospheric circulation." Much of the wide variation in pressure waves, and the several anomalous weather sequences which have forced wave opponents into rather tortuous synoptic analyses, can be explained by assuming that both processes are at work.

Arctowski, the first meteorologist to winter in antarctic regions (on the *Belgica*, 1898–99), has suggested [5] that

> Even the formation of lows and highs of pressure, their tracks and their changes in form and extent, may be directly produced by atmospheric waves. . . .' Most probably the pressure changes observed on Kerguelen Island are the product of two intercrossing systems of waves, one of them originating on the Antarctic continent.

Synoptic evidence of a westward-moving antarctic pressure wave (misnamed a surge), and a possible explanation for it, were found by Lamb [81] on the maps covering all Antarctica which he drew at the end of a summer on Antarctica's Indian Ocean coast:

> Between 19 and 23 March 1947 a pressure surge [wave] moved from east to west across Antarctica, starting from James W. Ellsworth Land, but. . .seems to have come in from outside, originating in the breakdown of the subtropical high-pressure system of the south-west Pacific east of New Zealand. This view carries with it the further suggestion that warm air of subtropical origin was transferred, particularly in the levels of the atmosphere above about 4,000 feet, far to the south and possibly right over Antarctica. No information exists to show what temperatures might be observed in this air upon its arrival over Antarctica or how rapidly it would be chilled thereafter by the radiation conditions of high latitudes; such invading warm air would in any case normally be separated by a shallow surface layer of extremely cold air from the actual ice of the antarctic plateau, but it is likely that the more abrupt mountain peaks and ranges would penetrate it. (See also [76])

Another example of a warm polar anticyclone, "its warmth due to subsidence in the whole air mass of which it was constituted" and maintained by inflow of air at high levels from the northeast above a vigorous and extensive disturbance over New Zealand, was studied in detail by Palmer [62]. However, he did not recognize that this anticyclone, or a part of it, caused a typical pressure wave at Little America on 29 November 1929 as it moved northwest.

The anticyclone, deduced from hodograph analysis of two long balloon runs on the 29th to be "south of Little America," was actually over the polar plateau. Byrd's South Pole flight [115] through it on the 29th (GMT) encountered southerly winds throughout, and reported clouds forming to the east over the mountains. Gould [124], heading the geological party 350 miles south of Little America at the southern head of the Ross Shelf Ice, noted on the 27th that "our bad weather was moving northward for, as conditions improved with us, they had grown less favorable at Little America." Gould's camp had "perfect weather" all of the 28th (GMT); it remained warm and clear until late on the 30th, when a slight south breeze freshened into "a veritable antarctic blizzard" which was "the warmest southern wind we had ever faced and grew warmer as we neared the mountains"; southerly winds usually were biting cold, but this one was uncomfortable because of its strength, not because of its temperature.

Hypothesis. To incorporate this information, Palmer's three synoptic maps for 28, 29, and 30 November may be revised and extended south and west of Little America to show a lobe of a polar anticyclone moving out along about 140°E (one of the anticyclone cells postulated by the Highjump aerologists, representing frequent outflow, is along 145°E or 155°E; one of Lamb's high pressure centers is at 130°E). Palmer's hodograph for 1600 on the 29th indicates "that the anticyclone was becoming less intense and was probably re-oriented to extend off the continent somewhere in the northwest."

Pressure at Little America reached a maximum of 29.55 in. from 1000 to 1200 on the 29th (GMT), a rise of 0.22 in. in 26 hours, then fell 0.21 in. in 21 hours as an occlusion approached from the northwest. Both these changes exceed 0.20 in., the criterion for a "pressure wave." Winds remained easterly during the entire two days between pressure minima, but were 12 to 18 mph during the day and a half preceding the maximum and decreased to 3 mph or less for ten hours during the subsequent pressure fall, then turned NE and increased as the occlusion passed.

In this case, passage of a relatively weak warm anticyclone several hundred miles to the westward did little

more than affect the pressure and modify the strength of the prevailing easterly winds. A much stronger anticyclone, or one passing closer, would give Little America the "deep southerly current aloft (which) attended or preceded settled clear conditions and good visibility," as Harrison [55] found on the first Byrd expedition, and as successive meteorologists at Little America have verified.

To attribute all weather phenomena of the Ross Sea to dying cyclones from the northwest, as Palmer and Ramage have done, is to consider only half the picture. Perhaps the antarctic anticyclone develops lobes which become separate centers, one or more of which flow out into the westerly circulation. This process may be caused by the invasion of air from lower latitudes in another quadrant, as Lamb suggested, or by the impinging of cyclones on the anticyclone in such places as the Ross and Bellingshausen Seas; possibly such impinging occurs only because the anticyclone already has broken into separate cells. In the present state of meteorological ignorance, these points cannot be determined definitely.

Regardless of the first cause, such interplay of subantarctic cyclones and polar anticyclones is a plausible explanation of observed conditions. The anticyclones, travelling north or northwest, will give pressure maxima as they progress, so that pressure waves may be found moving in the same direction. The cyclones, moving east or southeast, will give pressure minima moving in that general direction.

Probably most of the minima of the controversial antarctic pressure waves are due to subantarctic cyclones, and many of the maxima may represent only the wedges between these cyclones, but the more pronounced maxima are apparently caused by outbreaks of the polar anticyclone. Application of this hypothesis to the available data, and examination of the newer data, particularly those for the summer of 1946–47, from this viewpoint, should provide much information on antarctic meteorology and help settle the acrimonious dispute over antarctic pressure waves.

TEMPERATURES

Upper-air temperatures and pressures were first measured over Antarctica in 1911 (Table IV), but only the nine-month series at Little America III during 1940–41 [47] and the 1947 summertime observations are sufficiently extensive, in number and height, to shed much light on the atmospheric circulation. Simpson's 12 balloon meteorograph flights at Cape Evans in 1911 [112], Barkow's 128 kite meteorograph flights from the *Deutschland* in 1912 [101], and the 57 kite and airplane meteorograph ascents at Little America in 1928 and 1934 [50] all ended well below the tropopause, and have been used chiefly to study the height and variations of the surface inversion [63, 99]. Antarctica's first radiosonde ascents, made in 1934 by Dr. Jörgen Holmboe, meteorologist on Lincoln Ellsworth's attempted transantarctic flight, were so erratic that they have never been worked up. Records of the first successful radiosonde ascents close to Antarctica, the 36 made from the *Schwabenland* in 1939 [94], were destroyed during the war, and only partial extracts are available [97] for the analyses which have recently appeared [8, 91].

Table IV. Antarctic and Subantarctic Upper-Air Temperature Soundings
(Information to 1 January 1950 for all places south of 50°S, except South America.)

From	To	Location	Method*	No.	Highest level (m)
13 Aug. 1911	25 Dec. 1911	Cape Evans, McMurdo Sound, Ross Sea	B	12	6743
— Mar. 1912	— Jan. 1913	*Deutschland*, Weddell Sea	K	128	2400
22 Sept. 1929	20 Dec. 1929	Little America, Ross Sea	K	24	2559
13 Nov. 1929	7 Dec. 1929	Little America, Ross Sea	A	5	3548
3 Sept. 1934	9 Jan. 1935	Little America, Ross Sea	A	28	4929
— Oct. 1934	— Nov. 1934	Deception Island, Palmer Peninsula	R	6	—
20 Jan. 1939	6 Feb. 1939	*Schwabenland*, South Atlantic	R	36	—
26 Apr. 1940	15 Jan. 1941	Little America, Ross Sea	R	189	33750
15 Dec. 1946	15 Mar. 1947	"Highjump," all coastal waters	R	348	19900
— Dec. 1946	— Mar. 1947	*Willem Barendsz*, Northeast Weddell Sea	R	90	—
15 Dec. 1947	15 Feb. 1948	*Edisto and Burton Island*, 90°E to 70°W	R	104	19183
— Dec. 1947	—	Heard Island, 53°01′S, 73°08′E	R	cont.	—
— Jan. 1948	—	Macquarie Island, 54°30′S, 158°57′E	R	cont.	—
— Feb. 1949	— Feb. 1949	*Commandant Charcot*, 136° to 164°E	R	25	—

* Balloon meteorograph (B), kite meteorograph (K), airplane meteorograph (A), radiosonde (R).

Meteorological observations in phenomenal quantity were made in antarctic and subantarctic waters during January and February of 1947, including 348 radiosonde ascents by the three groups of the U. S. Navy's Operation Highjump and about 90 by the Dutch whaler *Willem Barendsz* in the north Weddell Sea. All the Highjump ascents have been published [3], albeit only by standard pressure levels and arranged chronologically regardless of location; the Dutch soundings mentioned by Lamb [83] have not yet appeared.

The following summer, two U. S. Navy icebreakers obtained 104 more radiosonde observations around the continental edge; these are available on microfilm but have not been published because of the lack of prompt

response to the thorough publication of the more extensive Highjump data, whose use has just begun [60, 91]. At Macquarie and Heard Islands, the Australians [69] began daily radiosonde observations early in 1948, and at Marion Island, the South Africans [92] began regular radiosonde ascents in May 1949. During February 1949, twenty-five flights were made from the *Commandant Charcot* [71] off the Adélie Land coast, south of Australia between 136°E and 164°E, as it unsuccessfully tried to land a French expedition which did establish a base a year later.

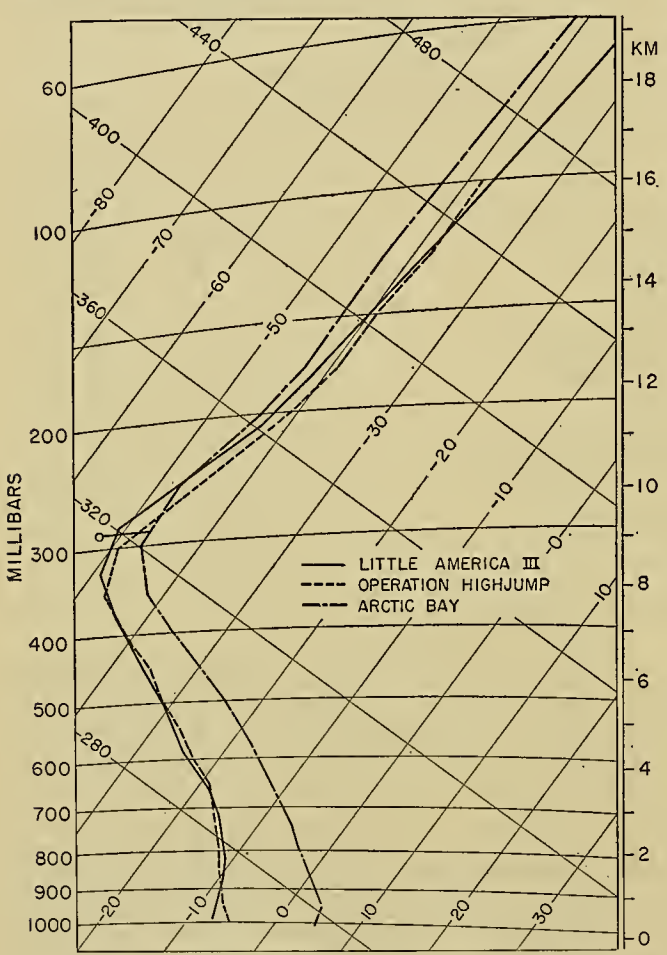


FIG. 1.—Antarctic and arctic summertime soundings. The curve for Little America III (78°30′S, 163°50′W) is based on thirteen soundings during the period 1–15 January 1941 [47]. The curve for Operation Highjump (70°–75°S, 160°W–165°E) is based on thirty-four soundings from 14 January to 8 February 1947 [3]. The data for Arctic Bay, Canada (73°16′N, 84°17′W) are from 152 soundings during July, 1943–47 [134]. The tropopause at Arctic Bay was computed separately.

Despite these extensive summertime soundings, the 1940–41 series provides the only information to date on the entire troposphere and lower stratosphere throughout the year. The 265 days covered by the 1940–41 data can be extended another twenty-two days by the 13 Highjump flights made from ships in the Bay of Whales and 21 others made in the Ross Sea south of 75°S. Close agreement between the averages of these 34 shipboard ascents and the 13 at Little America six years earlier is shown in Fig. 1; also plotted for comparison are the average temperatures of the warmest month, July, at North America's coldest aerological station: Arctic Bay, Canada (73°16′N, 84°17′W, 11 m m.s.l.), the only station on the continent at which the mean monthly temperature at 850 mb is never above freezing [134].

Wintertime. The Little America soundings show the tropopause in summer at roughly 9 km, −50C, and slightly higher and much colder in winter, at 10 km, −70C. The annual variation in stratosphere temperatures and lapse rates is so great that in summer it is warmer than −40C above 14 km, while in winter the stratosphere lapse rate approaches that of the troposphere and the tropopause tends to disappear [45].

Interpretation of the wintertime soundings is difficult because the extreme cold imposed an effective ceiling on the balloons, so that the higher soundings were all in warmer-than-average air. This tendency is shown by a diagram (Fig. 2) of the extreme temperatures at each

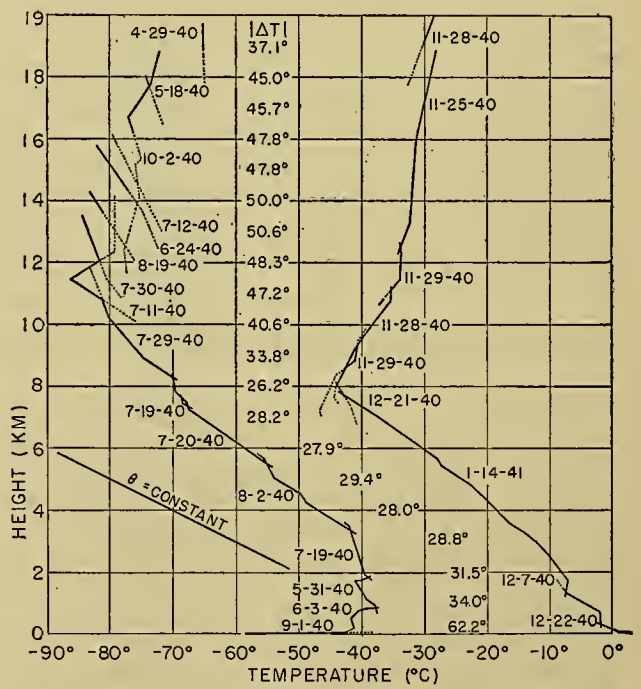


FIG. 2.—Observed temperature extremes and absolute range $|\Delta T|$ at standard levels over Little America III [47]. Dates are those of individual soundings. A dry-adiabatic (θ = const) is drawn for comparison.

kilometer level, with portions of the soundings in which they occurred. The envelope thus formed is smooth for both extremes in the troposphere and for the warmest values in the stratosphere, but shows that the coldest stratosphere temperatures always occurred in soundings which terminated soon thereafter.

From June through September, none of eighty-eight soundings (one with a 700-gram balloon) rose to 18 km, only two to 17 km, three to 16 km, and five to 15 km. At 17 km, temperatures were −70.8C on 5 June and −75.8C on 14 September. At 16 km, these same soundings gave −69.2C and −75.0C, but the third

sounding, on 12 July, gave −79.3C. At 15 km, the coldest value, −79.5C, was reached on 24 June in a sounding ending at 15,770 m, −82.1C, and was almost 3C lower than the next coldest reading (on 12 July). At 14 km, the coldest reading, −82.1C, was reached on 19 August; only one other August sounding reached 14 km, finding −80.7C, a temperature colder than found at this level in any other month.

Thus it seems that at 14 km and for an unknown distance above, winter temperatures must often be below −80C. Consequently, the absolute ranges of temperatures given in the diagram for each kilometer level are valid only up to about 13 km, where they attain the phenomenal value of 50.6C. They are misleading at higher levels: at 20 km, where the warmest temperature was −28.4C on 28 November, the lowest in late winter was probably colder than −100C, for an absolute range of more than 70C in three months. In the mid-troposphere, however, the absolute range was only about 29C and the range between extreme monthly means is only some 15C, about 40 per cent of the greatest such range in North America, over Hudson Bay [131].

This wide variation of temperatures in the lower stratosphere, unparalleled in the Northern Hemisphere, is for a station not in the interior of Antarctica, but for one several hundred miles from the continent proper, at the very edge of the thick Ross Shelf Ice with open water four miles away in summer and not more than twenty or thirty miles distant in winter. Conditions over the continent can only be inferred from the Little America data, until soundings from the interior are actually available. And inference, with so many factors unknown, can be very misleading.

For instance, Grimminger [49] deduced from the 969 pilot balloon ascents of the first two Byrd expeditions that the tropopause was about 1 km higher in summer than in winter and averaged 7.7 km; his computation assumed that, with increasing latitude, troposphere temperatures decrease and stratosphere temperatures increase, as they do in middle latitudes. Later measurements, however, showed the tropopause lower in summer than in winter, and about 2 km higher than the level of strongest winds, which was around 8 km in 1940 as in 1928 and 1933.

Summertime. Temperatures do decrease with latitude all the way to the Pole in *winter*, in both troposphere and stratosphere, as Grimminger assumed. This is shown in the first meridional atmospheric cross section [60] to use winter data from high southern latitudes, those for Little America already discussed as well as one year's values for Macquarie Island. But the *summer* diagram, for which in addition the Highjump data were used, is open to question: not only do stratosphere temperatures *increase* continuously from Equator to Pole, as it shows, but in the troposphere it is more probable that the lowest temperatures in midsummer occur at about the margin of Antarctica, and that toward the interior temperatures actually increase again.

Poleward increase of temperature in summer in the upper troposphere and in the stratosphere is indicated by the *downward* progress of the springtime warming over Little America, from October at the upper limit of observation, in the mid-stratosphere, to December in the lower troposphere. The pattern is revealed clearly in the individual observations: at 4 km warming does not start until December, at 8 km (just below the tropopause) it occurs chiefly in November, at 12 km in late October and November, and at 16 km in October and early November. At 20 km the warming had already occurred by the time the lower layers became warm enough for a balloon to reach that height on 1 November. The trend is shown even in monthly means (C):

Height (km)	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.
20	...	...	...	−31	−33	−34
16	...	−79	−65	−37	−38	−37
12	−74	−75	−69	−46	−43	−42
8	−64	−63	−62	−53	−49	−49
4	−38	−38	−33	−32	−25	−25

Obviously, advection cannot be the major cause of the stratospheric warming, since the temperatures become warmer than those at the same elevations farther north. The warming at 16 km of about 1C per day could be caused by subsidence of some 100 m per day, about one-third the rate of the downward progress of the start of the warming, which is around 2 km per week. However, the most plausible explanation of the rapid warming is that it is due chiefly to absorption of solar radiation: in the summer half-year, the duration of sunlight increases not only with latitude, but with elevation [130]. At 16 km over Little America, continuous sunshine is received from 10 October onward, but at the surface it does not start until 20 October; at the South Pole, it starts about 10 September at 16 km, and about 20 September at the surface.

This downward progress of available solar energy can explain the apparent heating from above of the atmosphere over Little America, provided the energy is absorbed. Presumably, it is absorbed by ozone, which is known to exist in greater concentrations and at lower levels in the arctic atmosphere than in lower latitudes, and to have its maximum concentration in spring. No measurements of ozone concentration have been reported from Antarctica, but the atmospheric temperature regime seems to indicate that it may be higher than in arctic regions.

This explanation of the springtime warming over Little America implies that farther south, where continuous sunshine begins earlier, the temperature rises sooner, and thus that in spring and up to the solstice the temperature increases poleward at all levels above the surface layer. Consequently, any diagram of summer temperatures over Antarctica, such as those of Loewe and Radok [60] and Flohn [91], should show the −50C isotherm as sloping downward from lower latitudes to about 60° or 70°S, then bending upward vertically; likewise the −40C isotherm probably does not reach the pole, but bends upward similarly around 85°S. At lower levels, the warmer isotherms extrapolated poleward from Little America (78°30′S) should

bend upward, rather than continue downward as in the published figures.

These temperatures, warmer in interior Antarctica than at the periphery, coupled with slightly higher surface pressure, would cause pressures near the Pole to be higher at all levels than around the coasts. Consequently, the isobaric surfaces should be lowest at 60° to 80°S and then rise slightly toward the interior, rather than continue to slant downward (as in Fig. 4 of Loewe and Radok, and up to 300 mb in Flohn's Fig. 8). More precisely, the trough in the isobaric surfaces in early summer probably inclines poleward with height up to the tropopause, so that the 300 mb surface is lowest (5 km) at about the coastline and the 200 mb surface lowest (11 km) some distance inland; Flohn's diagram suggests such a pattern, with a trough around 65°S at 150 mb and higher. Thus upper-level west winds at coastal stations may not indicate an upper-level polar cyclone, but the southern edge of the westerly flow. Implications of these suggestions are discussed in the final section.

The Loewe-Radok cross section for 150°E shows the tropical tropopause extending to 40°S and the polar tropopause beginning under it at 30°S, where it is at 12.6 km in summer, 8.5 km in winter. At 55°S the height is the same, 10 km, throughout the year, and farther south the height decreases in both seasons, but is slightly greater in winter. The twenty-five soundings from the *Commandant Charcot* [71] around 65°S in the coastal icepack, some 400 miles south of Macquarie Island, during February 1949 found the tropopause at 6 to 9 km, −50 to −57C, slightly lower but in general agreement with this diagram. Flohn's summer cross section for the South Atlantic, with Little America data added, shows only one continuous tropopause.

The Loewe-Radok cross section shows tropopause height as nearly constant between southern Australia and Macquarie Island (36° to 55°S), but 1000 miles to the east, in New Zealand, a winter "reversal of the slope of the tropopause" has been noted [57], the average height during July 1945 increasing[5] from 9.7 km at Auckland (37°S) to 10.1 at Hokitika (43°S) and 10.3 at Taeri (46°S). Earlier, Kidson [11] assumed that the "annual variation of temperature in the stratosphere can be little less than that at the surface in subantarctic latitudes and...the base of the stratosphere must vary little," because ocean temperatures between 45°S and 60°S change little through the year.

Other Aspects. Water temperatures at Campbell Island (52°32′S, 169°08′E), about 400 miles south of New Zealand, vary [56] only 6F through the year, from 42.8F in July to 49.1F in February (1942–47). Air temperatures there are 2F warmer in winter, 7F warmer in summer than at Macquarie Island (54°30′S, 158°57′E), 500 miles to the west, which in turn is noticeably warmer [73] than Heard Island (52°01′S, 73°08′E), 3400 miles farther west. Marion Island (46°51′S, 37°52′E), 1600 miles farther west-northwest of Heard, appears from preliminary reports [92] to be substantially warmer than Heard, but colder than Macquarie.

Month by month during 1948, Macquarie was 10F or more warmer than Heard [73], with the difference gradually diminishing with height so that at the tropopause, around 300 mb at both stations, the modal temperature at Macquarie was −50C, Heard −53C. Despite local topographic effects (both stations are a few feet above sea level on the low, narrow necks of peninsulas extending northward from their main rocky islands), winds at Macquarie were northerly, at Heard southerly, showing that the wind regime is far from zonal.

Why should the maximum temperature on clear days during the antarctic night occur, almost without exception, in the hours after midnight? Simpson pondered this problem longer than any other of the many he investigated in compiling the meteorological reports [112] of the last Scott expedition, without developing a satisfactory hypothesis. "In McMurdo Sound there is an excess of temperature at 4 A.M. compared with 2 A.M. and 6 A.M. [only bihourly readings were made] in every month from April to September, and the same effect is possibly present in October, November, and March,...is clearly seen in the temperature observations at the *Gauss*...[and] the Snow Hill observations also show the same effect in July and August." In all cases it was characteristic of cloudless or partly cloudy weather; on cloudy days "during the three winter months when there is no direct solar radiation the day is warmer than the night."

Similar relations were found by Rouch on the Charcot expedition and later in the data of the two Byrd expeditions [65], where the average daily variation of 2C on 33 clear days "is a simple curve with maximum around 1 or 2 A.M. and minimum after noon...I cannot give a satisfactory explanation." Nor is van Everdingen's advective explanation [24] very plausible.

Less mysterious, but still not satisfactorily explained, is the annual march of temperature at Little America [47], which shows a single maximum in early January, but three distinct and progressively colder minima: in early May, in July, and in early September. Evidenced in monthly means, this is shown most clearly by 10-day means (Fig. 3). Temperature averages at other antarctic stations do not show such triple minima, although individual years at many stations do show double minima as found, although not adequately explained, at arctic coastal stations.

Other aspects of Antarctica's temperature regime offer no peculiarities. The interdiurnal variation and the mean daily range [16] change little through the year between 50° and 60°S. They are greatest in winter south of 60°, greatest in summer north of 50°. The

5. This anomaly is shown also on an excellent winter cross section along 170°E from Little America to the equator, published since preparation of this article. Otherwise it and the companion summer diagram are quite similar to those of Loewe and Radok for 20° farther west, but do not attempt extrapolation poleward from Little America. (See Hutchings, J. W., "A Meridional Cross-Section for an Oceanic Region." *J. Meteor.*, 7: 94–100 (1950).)

meridional temperature gradient is strongest in winter south of 55° or 60°S, and also north of 30°S, and weakest in winter between 30° and 55°S.

Atmospheric pressure around Antarctica is exceptionally low; readings as high as 1000 mb are unusual and at Little America on 24 September 1940 sea-level pressure was only 932.5 mb. Meinardus [14] distinguished two types of annual pressure variation: a deep minimum in the Ross Sea area in late winter or early spring (July to October) with a very steep rise to a sharp maximum in midsummer (December or January), and a double period with minima around the equinoxes and maxima near the solstices in the Palmer Peninsula area. Whether this difference is regional, latitudinal, or

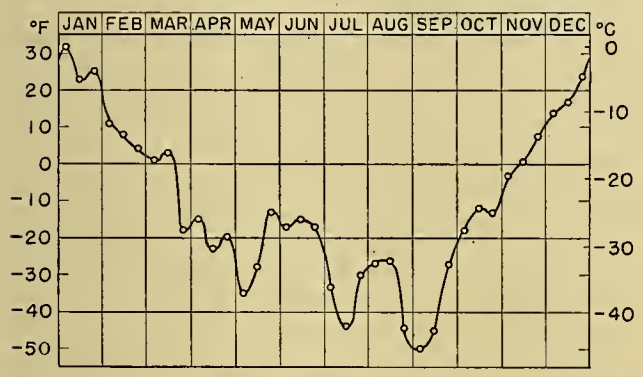


FIG. 3.—Mean surface air temperature for 10-day periods at Little America, based on three years of data.

due to the differing years of observation is not certain; the more recent data from the Palmer Peninsula [36, 37, 40] show no pronounced pattern. In the Ross Sea, the lowest monthly values at Little America have always come in September, while in McMurdo Sound each year has had its minimum in a different month.

Antarctica is the only part of the world where the atmosphere appears to have less oxygen than "normal." Everywhere else oxygen represents 20.95 per cent of the dry constituents of surface air, but the only analyses made of antarctic air [58] gave an average of only 20.56 per cent in summer, and 20.64 in winter except for one day (22 July 1940) which averaged 20.50. Although two leading authorities on atmospheric composition [132] do not regard this deficiency "as well established, since no check analyses with normal air were carried out," these are the only analyses to date and their implications must be considered until they are proven definitely erroneous.

'Atmospheric oxygen is being removed constantly by rock weathering, while burial of organic material results in a net gain. Rock weathering in Antarctica, however, seems insufficient to maintain the observed deficiency, even if all the presently unknown continental areas contained bare rocks undergoing weathering. A possible link between antarctic air and that of the middle-latitude stratosphere, which was considered in 1936 to have an oxygen content of about 20.50 per cent, has been destroyed by recent indications [129] "that the chemical composition of stratospheric air at 70 km is practically the same as that of the troposphere."

PRECIPITATION

Precipitation in Antarctica cannot be measured as it falls, but its total accumulation over a year can be determined by measuring the daily changes in height of the snow surface, using a graduated stick imbedded in the snow. At Little America during 1940 [47], such a program, involving four pairs of sticks several hundred feet distant in four directions from the camp, showed a total accretion during 348 days of 170 cm due to new snow, hoarfrost, or drift snow, and total decrease of 85 cm due to ablation, deflation, or settling. The net increase of 85 cm had an average density of 0.35 and therefore represented a net accumulation of 30 cm or 12 inches of water during 11½ months, most of it in March, June, and July.

In the western Ross Sea the 1902–1909 average (computed from the accumulation over a depot on the Ross Shelf Ice) was 19 cm (water equivalent); other estimates for various years were: Cape Evans 41 cm, Cape Royds 23 cm, Cape Adare 36 cm. Annual precipitation at the *Gauss* was estimated at 80 cm, at Charcot's two Palmer Peninsula bases 34 and 38 cm, and at the *Deutschland* only 11 cm.

At the South Pole, Meinardus [14] estimated annual precipitation to be 2 cm, ablation 1 cm, for a net accumulation of 1 cm; independently, Kidson [74] found it "difficult to imagine that the precipitation at the Pole is greater than 2 inches per year," or 5 cm. For all Antarctica, Meinardus estimated precipitation 7 cm, ablation 3 cm, net accumulation 4 cm; Kidson deduced total continental precipitation at less than 9 cm.

Kidson's estimate assumed that precipitation of all the moisture in the lowermost 2 km of an air mass initially saturated at 32F would give 0.5 cm, and that this process would require twenty days. An entirely different approach was used by Meinardus [107]: "If the interior of Antarctica is covered with snow and ice and there is a discharge of the same to the marginal oceans, it follows that over Antarctica as a whole the precipitation exceeds the evaporation." Outflow of the inland ice, 250 m high and 17,000 km in perimeter, at 150 m per year, totals 640 km³ of ice or 550 km³ of water, requiring an annual net precipitation of about 4 cm over Antarctica to maintain balance [108].

Although Meinardus considered his figures conservative, each of the three factors in this computation seems excessive in the light of recent exploration. Furthermore, there is considerable doubt that the ice which does flow out around Antarctica originates very far inland.

Height. Eighty years ago, an ice thickness of 24 miles at the South Pole was estimated by Croll [117], assuming that the ice cap extended to 70°S and had a one-degree slope, as required "to produce the necessary outflow of ice." However, "to avoid all objections on the score of over-estimating the thickness of the cap," he used a thickness of only 6 miles in his computations, explaining that "the greater thickness of an ice-cap at its centre is a physical necessity not depending on the rate of snowfall," which anyhow would not be negligible

because "the currents carrying moisture move in from all directions towards the pole."

Forty years ago a mean height of 2000 m for Antarctica was deduced by Meinardus [106], since the annual variation in the amount of air over the rest of the world, as computed from mean station pressure values, would be just compensated if the entire surface inside the Antarctic Circle were 1350 m above sea level; if one-third the area inside the circle is at sea level, the rest must be at 2000 m—a value which has been used extensively and uncritically by geographers, geologists, and others. Using the same pressures but different antarctic mean surface and free-air temperatures, Simpson [112] obtained an average height of only 852 m, which he considered just as plausible as 1350; Meinardus' recomputation [109] for the area inside the circle, based on later data, was 1500 m, and for the continent, 2200 m.

Meanwhile, the altitude of the snow surface at the Pole itself was found by Simpson [112], from exhaustive study of Amundsen's and Scott's barometric readings, to be 9172 ft or 2.8 km, 700 ft *lower* than the ice divide crossed by both explorers at 88°30′S, some 200 miles back of the crests of the Queen Maud Mountains which form the south side of the Ross Sea. A similar ridge was found behind the Admiralty Range on the west side of the Ross Sea.

The German *Schwabenland* expedition in 1939 found a new chain of east-west mountains athwart the Greenwich meridian about 250 miles inland. Behind this range, which stretches at least 500 miles and has peaks up to 4 km, the ice surface rises even higher. Herrmann [126] offered an antarctic cross section along the 0°–180° meridian, showing a trough between these mountains and the Queen Mauds, with 1 to 2 km of ice in the trough so that the highest point, at 80°S, is more than 5 km above sea level, and the lowest point of the snow surface is 3 km m.s.l., at the Pole. This range, extending along 80°S eastward to at least 20°E, the limit of German exploration, may be a continuation of one postulated[6] by Lamb [81] to explain the apparent behavior of weather along Antarctica's Indian Ocean coast, and which he found "for my meteorological purposes. . .was a satisfactory working assumption":

> The whole trend of recurving depression tracks, both from the Ross Sea-Wilkes Land sector and in the Indian Ocean sector,. . .lend support to the belief in a topographical barrier in the interior of the continent of similar order of magnitude to the Alps or Rocky Mountains, say 12,000 to 15,000 feet in height. There is no evidence to suggest how far the barrier may extend west and south-west beyond 80°S, 80°E, nor whether it may be a great range of mountains or the main crest of the antarctic ice-cap. Nor can one say whether its crest presents an unbroken contour or is interrupted; but I think its north-eastern end is likely to be nearer 70°S and 97° to 100°E than at the coast itself. . . . The principal south polar anticyclones tend to take up positions centered. . .near the line of this supposed summit and show signs of dividing across it when they decay. . . . Very occasionally a depression moved right across the ice-cap on one side or other of the line of this supposed barrier, but probably never crossed it.

The *Schwabenland* findings, Manley pointed out [12], "go far to alter the simple concept of an antarctic dome more or less symmetrical about the pole." Similarly, the chain of east-west mountains found by Operation Highjump along 76°S from 110° to 140°W have one peak some 6 km high, although the Hollick-Kenyon Plateau behind them, on which Ellsworth landed, is less than 2 km high. Thus recent exploration has generally invalidated the simple model of a high East Antarctica plateau and a low West Antarctica plain, the model on which Simpson based his theory of antarctic weather processes.

Even more important, if there is no central dome to the ice cap, but instead a central trough or basin, then "the greatest problem of the antarctic anticyclone, namely, the origin of the precipitation within the anticyclone," as Simpson put it, disappears. Instead, the continental interior is an area of little or no precipitation, which can well have a perpetual anticyclonic regime, as long as the circulation provides snowfall up to the observed ice crest.

Recession. Even this peripheral snowfall need not equal the annual ice outflow, because Antarctica's present ice is merely "the remains of the Pleistocene glacial ice-sheet and. . .at present a complete recession is taking place." (The rapidity and course of this recession are major problems of the present Norwegian-British-Swedish expedition.) Although in his *Handbuch* monograph [14] Meinardus assumed climatic equilibrium in estimating Antarctica's precipitation, his earlier study [108] pointed out that equilibrium probably does not obtain. There he concluded "that Antarctica during the ice age had, in association with an increased air circulation, higher temperatures, augmented water vapor turnover, greater precipitation, and greater evaporation." Higher antarctic temperatures arose chiefly from increased circulation, so that elsewhere temperatures were lowered, but whether "increased precipitation or lowered temperatures [were] the primary cause of glaciation" in lower latitudes was uncertain.

Manley's mechanism for the original glaciation of Antarctica [12] starts with a relatively ice-free continent surrounded by seas without permanent ice, so that the resulting heavy precipitation along the coasts, and particularly in the mountains adjacent to the Ross and Weddell Seas, would cause glaciation. Outflow glaciers would so chill these seas that their present ice sheets

6. An earlier geographic postulation from meteorological evidence proved unfortunate: "Before the *Scotia* had left the Antarctic Seas, Mr. Mossman was able to demonstrate meteorologically the existence of the land reported by Johnson and Morrell, extending northward to about 65°S at 44°W," largely by the coldness of SW winds at Laurie Island. "Since that time, with the additional data furnished by the Scotia Bay Station during eight years, it has become more than ever certain that New South Greenland, as Johnson called it, really exists." Within a year after Dr. W. S. Bruce wrote thus in *Polar Exploration* (New York & London, 1911), Dr. Wilhelm Filchner sledged 40 miles from the beset *Deutschland* and found no bottom at 600 fathoms where New South Greenland was postulated; three years later Shackleton's *Endurance* drifted across the supposed land area and finally sank in the middle of it.

would be the result (although glaciologists [119] estimate that less than a third of the present Ross Shelf Ice is derived from glacier outflow).

Snow and ice flow away more rapidly from the seaward faces of the coastal mountains than from their backs, "so that after a time the ice-shed lies inland from the mountains," a phenomenon observed in Antarctica and deduced for the Pleistocene ice sheets of Scandinavia and Canada. Behind this divide, any inflowing air is descending, so in the interior "precipitation is small and...the ice in that region must have largely accumulated at a time when the Ross Barrier did not exist."

Finally, Manley reasoned, glaciers from the coastal mountains so cool the surrounding ocean that the open water, moisture source for the precipitation, is found farther and farther away, and precipitation decreases. This involves a cycle perhaps several centuries long: heavy coastal precipitation, glacial outflow, sea cooling and freezing, reduced precipitation, starved glaciers, less sea ice and warmer seas, increased precipitation.

Manley's glaciation cycle is self-generating, as opposed to others which require some external change, be it variation in solar energy, in mountain heights, or in submarine topography affecting ocean currents. Simpson has maintained consistently [138] that "the glacial epochs were the consequence of an increase and not a decrease in the solar radiation," and that "higher temperatures and greater general circulation of the atmosphere are absolutely essential to produce increased outflow of ice in the Antarctic."

Kidson [74] questioned this precept, because "in temperate, and probably also antarctic, regions of the southern hemisphere, precipitation is heaviest in winter," so that "a marked fall of temperature would be necessary to produce the conditions at the maxima of glaciation." Flint [7], who has reviewed Pleistocene glaciation in all parts of the world, pointed out this conflict. His excellent treatise [123] expresses the prevailing view of geologists that Pleistocene glaciation was due to a general lowering of the earth's mean temperature, presumably by reduction in solar radiation, so that a greater proportion of the earth's precipitation fell as snow.

Modern meteorological principles led Willett [139] to support Simpson: "The general circulation and worldwide precipitation are greatly intensified during an ice age, results which cannot be brought about by a general lowering of the earth's mean temperature." From a detailed study of the present temperature and precipitation regime along 49°N, Leighly [136] concluded that the North American glaciation was due to increased wintertime transport of warm moist air from the Gulf of Mexico to the St. Lawrence region, such as obtains with low-index conditions.

The various hypotheses about Pleistocene atmospheric processes, summarized extensively by Willett [139] and more briefly by Landsberg [135], do not help materially to define Antarctica's circulation; instead, they require its independent elucidation for their full development. In one way, however, they have removed one of the unknown variables from the circulation equation: production of precipitation in the continental interior is not a requirement for an acceptable description of Antarctica's present circulation.

CONCLUSION

Antarctica's atmospheric circulation has been described in terms of various versions of a polar anticyclone—glacial, laminar, cellular, peripheral, and abstract. Their review in the light of present knowledge of topography and precipitation distribution, of pressure waves and oceanic storms, of upper-air temperatures and surface temperature and pressure regimes, shows first that Antarctica's circulation cannot be described adequately by its annual pattern alone.

Variations in the extent of ice around the continental edge, effectively expanding the land area in winter and spring and decreasing it in summer and fall, have profound influence. More important are the changes from continuous summer sunshine to continuous winter darkness with outgoing radiation. Because the snow surface absorbs little solar energy but readily radiates away its own heat, winter cooling is primarily at the surface while summer heating occurs in all layers, beginning in the stratosphere.

Throughout the year, the planetary pattern causes pressures to be lower around 60°S than at the Pole; the lower temperature of the continental snow surface as compared to the subantarctic waters intensifies the surface pressure gradient. But in the upper air, as already mentioned, the gradients apparently reverse with the seasons: temperature and pressure in winter decrease from middle latitudes to the Pole, while in summer their decrease stops at about Antarctica's margin and over the continent they increase poleward. The insolation which warms the lower stratosphere over Little America to almost −30C in November and December should cause even higher temperatures farther poleward, so that on the average Antarctica is overlain by a deep warm semipermanent anticyclone in early summer.

By late summer, insolation is no longer able to maintain midsummer temperatures in the upper air; possibly the ozone concentration decreases markedly by midsummer, stopping the warming. The upper layers then cool slowly but steadily so that by midwinter the lapse rate in the stratosphere approaches that of the troposphere and the tropopause disappears. Cooling lowers all the isobaric surfaces, creating over the surface anticyclone an upper-level cyclone which deepens as the cooling progresses; at the surface, pressure falls during fall and winter to a September minimum. Spring sunshine affects the stratosphere first, warming it rapidly; the cyclonic layers expand and rise, and by summer the cyclone disappears, with surface pressure rising rapidly.

In summer and winter, the surface distribution may have the form of a large central anticyclone, or one with several pronounced lobes, or several cells around a weak center. During periods of high zonal index the polar anticyclone is apparently central, during low-

index periods several cells may exist. Each of these patterns may thus be found in different years, or they may appear in succession as the hemispheric flow pattern changes. Consequently, isolated observations in different years may indicate conflicting results.

Cyclones travel in families south-southeastward across the three great oceans to stagnate in the seas which indent the coast to greater or lesser degree: Ross, Amundsen, Bellingshausen, Weddell, and "Mackenzie" (70° to 80°E). These cyclones may go inland when the zonal index is low and there are several anticyclonic cells, especially in summer, but they are confined to the coast when the index is high and there is an extensive central anticyclone. Those which do go inland provide some precipitation in the interior, but fill rapidly as they leave open water and lose energy.

Exchange. Expansion of the atmosphere in summer by the warming of both troposphere and stratosphere causes a net outflow of air from Antarctica. This outflow occurs chiefly in those areas where anticyclonic lobes or cells form during moderate- or low-index conditions: in the southeastern corners of the three oceans washing Antarctica, or roughly 10°W to 20°E, 130° to 160°E, and 120° to 100°W. A fourth area is along the Palmer Peninsula (60° to 70°W), for topographic reasons.

Outflow in these areas is reinforced by meridional exchange, which occurs throughout the year at a rate inversely proportional to the zonal index. It is not a continuous flow, but occurs as outbreaks of anticyclones from the interior; as they progress outward, these anticyclones cause "pressure waves." Their formation, in turn, may be associated with the poleward motion in other longitudes of warm anticyclones of subtropical origin, so that the waves at times appear to have travelled across the entire continent. The warm anticyclones from the subtropics provide some of the compensating air influx.

Even under very-low-index conditions, when the meridional exchange between antarctic and subantarctic regions is greatest, it is very slight compared to the amount of air exchanged between continents and oceans in the rest of the world. The first report [45] on the extreme cold of the winter stratosphere over Little America suggested that "lack of circulation between Antarctica and lower latitudes would permit air there to cool unmolested in the winter until it attained lower temperatures than anywhere else on earth," and that the subantarctic westerlies "may largely confine the antarctic air."

This hypothesis and those of Hobbs, Meinardus, Simpson, and others were doubted by Ramage [64] insofar as they involved a "disconnection between antarctic and temperate circulations," because "before such a circulation theory, unparalleled anywhere else in the world, can be put forward successfully, detailed and accurate observations must be made in the region where the discontinuity is thought to exist." However, he conceded that "a temporary interruption of atmospheric interchange with lower latitudes would be sufficient to cause the tropopause inversion to disappear in winter."

Strong support for the hypothesis of a general lack of air exchange between Antarctica and the rest of the world is provided in an "exclusion principle" postulated by Starr and reported by Willett [140]: "A strong circumpolar cyclonic vortex or zonal westerlies should necessarily repress" transport of heat and kinetic energy. Even at their weakest, the subantarctic westerlies are far stronger than the Northern Hemisphere westerlies [8], so that by this principle the meridional flow in high southern latitudes should be far less than in the north polar region—as Bjerknes and others had suggested much earlier. After several months of drawing and analyzing complete Southern Hemisphere maps, Willett and his staff felt tentatively that "real polar outbreaks of the large-scale type involving the equatorward movement of large polar anticyclones, such as we observe in the Northern Hemisphere, do not occur on the same scale in the Southern."[7]

Slight as this exchange is when compared to that of the Northern Hemisphere, variations in it affect the general hemispheric circulation. Seasonal changes in the pressure field induce less meridional exchange in some seasons than in others. The flow apparently is least in late spring and early summer (October and November) because of the rapid heating at all levels, and isobars then should be most nearly zonal without pronounced lobes on the vertically expanding continental cyclone. Consequently, the zonal westerlies are strongest in spring, as Gabites [48] has found. This seasonal variation may account for the early summer cyclonic weather encountered by Ellsworth in an area later postulated as having an anticyclonic wedge or cell in late summer.

Implications. The general hypothesis outlined above differs from earlier versions of the polar anticyclone postulated by others in allowing two types of modification: seasonal and circulational. In winter it is similar to Simpson's laminar model, and during low-index periods not markedly different from Hobbs's glacial anticyclone—except that no artifices are needed to provide central precipitation since recent evidence indicates that little occurs. In summer it follows the cellular or lobar pattern suggested by Lamb and the Highjump aerologists, recognizing that their limited period of analysis was one of moderate westerly flow. And in recognizing the penetration of warm air masses, it agrees with Palmer's contention that the anticyclone is not a permanent feature but a model picture, though somewhat more than a statistical abstraction. The only model with which it is completely at variance is the peripheral anticyclone of Meinardus, which has been vitiated by more recent findings that Antarctica is not a simple elevated plateau.

Although the processes suggested above explain the upper-air temperatures and the surface pressure regime at Little America, they seem to offer no ready explanation for the three progressively colder temperature

7. Private communication, 4 November 1949.

minima there, and certainly none for the anomalous post-midnight diurnal temperature maximum of clear days during winter at various coastal stations.

As to specific situations, it is suggested that 1912 was a year of low zonal index with stronger-than-usual meridional exchange, so that there were abnormally strong south gales at Mawson's "Home of the Blizzard" in Adélie Land, in the path of one of the preferred outflow paths. Simpson [113] long ago suggested that February and March, 1912, were unusually cold on the Ross Shelf Ice, imposing unexpected hardships which proved fatal to Scott and his party returning from the pole, and Kidson [75] pointed out that "1912 was the windiest year experienced in McMurdo Sound." Conversely, the summer of 1908–9, when David found rather variable winds near the magnetic pole, may have been a period of reduced outflow.

Obviously, the variations in hemispheric circulation patterns which are considered to be reflected in the pressure and circulation distributions of Antarctica are intimately related to the variations in the amount of ice in the subantarctic seas. Coasts accessible by ship in some years are completely ice-jammed in others; some ships entering the Ross Sea have fought hundreds of miles of ice, others at the same time of year have found no pack ice at all, as shown in English's [122] exhaustive tabulation. The relation between the hemispheric circulation and sea-ice conditions is a complex problem, but one of the most important in Antarctica's meteorology.

Prospects. For the first time in history, prospects are bright that a good understanding of Antarctica's atmospheric circulation, and thus its climate, can be achieved in the near future. Two well-equipped expeditions have established bases on the coast south of the string of full-scale weather stations on subantarctic islands (Marion, Kerguelen, Heard, Macquarie, Campbell), and the British, Chilean, and Argentine stations in and around Palmer Peninsula are continuing (although the southernmost, in Marguerite Bay, was abandoned by air in February 1950 because ice conditions made it inaccessible by sea for two consecutive summers).

These, however, are still coastal stations. Sorely needed are a few year-round stations well in the interior of the continent, located for example along 80°S at 0°, 90°E, and 90°W; occupation of the South Pole itself, while spectacular, would be of less meteorological benefit than a single station at the world's "pole of inaccessibility," around 80°S, 80°E. Such stations, operating simultaneously with those active during 1950, and making radiosonde flights (and rawin flights if possible) at least twice weekly, would finally provide enough information on the seasonal changes in Antarctica's atmosphere to permit the circulation to be established definitely—and to permit a full solution of the puzzle of the pressure waves.

Meanwhile, there are several problems which require study. Ozone concentrations must be measured throughout the year at several places, and air samples should be procured weekly at antarctic bases, for later laboratory analysis, until the apparent oxygen deficiency is established or disproved. Pressure surges (not waves) can be sought in the more recent data (the 1935 problem has already been cited) to determine whether any benefits to forecasting can be obtained, such as the correlation cited by Diaz [28] that "polar air invasions over Argentina are preceded by . . . a barometric maximum over Antarctica (South Orkneys and Little America) from 6 to 5 days before." The anomalous post-midnight maximum of temperature on clear winter days requires explanation; unfortunately, clockwork deficiencies precluded a thermograph record during the 1940 winter at Little America, so that as yet there are no examples of this anomaly with simultaneous radiosonde observations.

Of course, further exploration of the continent is needed, since the meteorologist must know the nature of the land (or ice) surface in order to understand the weather processes above it. However, merely flying over the interior in a camera-equipped airplane will not give adequate information: the elevation is important, and to obtain it with sufficient accuracy requires ground surveys. As long as the basic circulation is uncertain, elevations cannot be determined by pressure altimeters hundreds of miles from the nearest barometer of known height or from the nearest ocean.

Expeditions which do go to Antarctica should have meteorologists not only competent to obtain the needed observations under adverse conditions, but meteorologists who understand the problems which they are to help solve. "It is not enough to give a man an instrument and then send him to some outlandish place to expose that instrument and take readings," said Sir George Simpson. He must know what he is doing, how others did it before, and above all why he is doing it and the significance of the results. Yet too many expedition meteorologists have been chosen more for availability than for ability, more for backpacking than for background, more for eagerness than for erudition.

Finally, the problems of Antarctica's meteorology cannot be solved in Antarctica. Observations must be made there, both by accurate instrument and by understanding eye, but they must be published quickly and in full detail and then synthesized and interpreted in the calm of office and library. An expedition cannot be termed scientific if it does not provide its meteorologists (and other scientists as well) as much time for study of the results, after the expedition, as was spent on the entire trip itself, and under the most favorable conditions for study; arrangements for prompt publication must also be made. Only through such prior commitments can any expedition hope to help materially in dispelling the present uncertainty concerning the climate and atmospheric circulation of Antarctica.

REFERENCES

In the past twenty years, more than one hundred books, articles, and notes concerning antarctic meteorology have been published. The most important of these are listed in

Section I of the extensive classified bibliography below, which overlaps Meinardus' list [14] by five years to include all results of the "mechanical period" expeditions beginning in 1928.

These hundred entries are arranged alphabetically within each of five regional groupings: General, Palmer Peninsula Sector, Ross Sea Sector, Australian and South Indian Ocean Sectors, and African Sector. Some entries with nonspecific titles, not discussed in the text, have short explanatory notes. Asterisks indicate 33 entries considered most significant because of either discussion or data.

The bibliography's second portion contains 40 other references cited, arranged alphabetically in each of three groups: Antarctic Meteorology before 1930, Antarctic Geography and Glaciology, and General Meteorology.

I. ANTARCTIC METEOROLOGY SINCE 1930

A. General

1. Aerology, Flight Section, Chief of Naval Operations, "Reports from Operation Highjump." *New Developments in Naval Aerology for Reserve Aerologists*, (NAVAER 50-50T-1) 1: 1–9 (1947). (Extract of [2].)

*2. —— *Aerological Aspects of Operation Highjump.* (NAVAER 50-45T-6, Restricted), 99 pp., Washington, D. C., 1947.

*3. —— *Aerological Observations and Summaries for the Antarctic from 15 December 1946 to 15 March 1947.* (NAVAER 50-1R-214), 404 pp., Washington, D. C., 1948.

4. —— *Aerological Aspects of the Second Antarctic Development Project.* (NAVAER 50-45T-10, Restricted), 24 pp., Washington, D. C., 1948.

5. Arctowski, H., "Notice sur les pseudo-ondes barométriques observées dans les régions antarctiques et ailleurs." *Kosmos, Lwów*, 52: 318–327 (1927); "On Weather Changes from Day to Day." *Mon. Wea. Rev. Wash.*, 67: 322–330 (1939); "On Solar-Constant and Atmospheric Temperature Changes." *Smithson. misc. Coll.*, Vol. 101, No. 5, 62 pp. (1941). (See pp. 9–12)

6. Constantino, C. E., "Clima de la Antártida." *Bol. Cent. nav., B. Aires*, 64: 271–294 (1945). (Elementary but thorough.)

7. Flint, R. F., "Glacial Climates in the Southern Hemisphere." *Amer. J. Sci.*, 244: 861–862 (1946).

*8. Flohn, H., "Die Intensität der zonalen Zirkulation in der freien Atmosphäre aussertropischer Breiten." *Beitr. Geophys.*, 60: 196–209 (1944).

9. Gentilli, J., "Air Masses of the Southern Hemisphere." *Weather*, 4: 258–261, 292–297 (1949).

10. Johansson, O. V., "Der jährliche Gang der Temperatur in polaren Gegenden." *Geogr. Ann., Stockh.*, 21: 89–118 (1939). (Analyzes annual temperature patterns shown on maps in [14].)

11. Kidson, E., "Some Problems of Modern Meteorology. No. 8, Problems of Antarctic Meteorology." *Quart. J. R. meteor. Soc.*, 58: 219–226 (1932). (Later published by the Royal Meteorological Society in *Problems of Modern Meteorology*, London, 1934.)

*12. Manley, G., "Recent Antarctic Discoveries and Some Speculations Thereon." *Quart. J. R. meteor. Soc.*, 72: 307–317 (1946).

13. Mecking, L., "Dic antarktische Treibeisgrenze und ihre Beziehung zur Zyklonenwanderung." *Ann. Hydrogr., Berl.*, 60: 225–229 (1932).

*14. Meinardus, W., "Klimakunde der Antarktis," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. 4, Teil U. Berlin, Borntraeger, 1938.

15. —— "Zum Klima der Antarktis." *Forsch. Fortschr. dtsch. Wiss.*, 16: 6–8 (1940).

*16. —— "Die interdiurne Veränderlichkeit der Temperatur und verwandte Erscheinungen auf der südlichen Halbkugel." *Meteor. Z.*, 57: 165–176, 219–233 (1940).

*17. —— "Die jährliche Periode der meridionalen Luftdruckgradienten und der Windstärken auf der südlichen Halbkugel." *Meteor. Rdsch.*, 1: 1–4 (1947).

*18. Palmer, C. E., "Southern Hemisphere Synoptic Meteorology" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See pp. 804–812)

19. Priestley, C. H. B., "Air Circulation and the Antarctic." *Aust. J. Sci.*, 10: 129–131 (1948). (Need for Australian stations in Antarctica.)

20. Ruthe, K., "Das Klima der Antarktis." *Polarforsch., Kiel*, 11(2): 1–3 (1941); "Vom Klima der Antarktis." *Ibid.*, 12(1): 1–4, 12(2): 1–4, (1942), 13(1): 1–5 (1943).

*21. Skinner, T. C., "Problems of Antarctic Meteorology." *Quart. J. R. meteor. Soc.*, 59: 21–22 (1933). (Comments on [11].)

22. Sparn, E., "Segunda contribución al conociemento de la bibliografía meteorológica y climatológica del cuadrante americano de la Antártica y Subantártica." *Bol. Acad. Cienc. Córdoba*, 34: 183–201 (1938); "Tercera contribución . . . ." *Ibid.*, 37: 332–341 (1945). (List 202 papers for 1924–1937 and 67 papers for 1938–1944, respectively, on climatology, meteorology, oceanography, and glaciology of all parts of Antarctica, despite title.)

*23. Sverdrup, H. U., "Diurnal Variation of Temperature at Polar Stations in the Spring." *Beitr. Geophys.*, 32: 1–14 (1931).

24. van Everdingen, E., "Der tägliche Gang der Temperatur in der antarktischen Polarnacht." *Beitr. Geophys.*, 32: 271–274 (1931).

B. Palmer Peninsula Sector

25. Bagshawe, T. W., *Two Men in the Antarctic.* Cambridge, University Press, 1939. (Meteorological appendix, pp. 209–229, has monthly means for 1921 expedition.)

26. Bustos Navarette, J., "Estudio Meteorologico de la Region antartica y su Importancia para la Prevision de Tiempo en la America del Sur." *Rev. Meteor.*, 3: 176–181 (1944).

*27. Coyle, J. R., *A Series of Papers on the Weather of South America, Part II.* Rio de Janeiro, Pan American Airways, 1943. (Reprinted as NAVAER 50-1R-105, Aerology Section, Chief of Naval Operations, 1944.)

28. Diaz, E., "Some Researches on the General Atmospheric Circulation." (Abstract) *Bull. Amer. meteor. Soc.*, 25: 370–371 (1944); "Algunas Investigaciones sobre Circulacion Atmosferica." *An. Soc. cient. argent.*, 137: 241–272 (1944).

29. —— "Posibilidad de establecer una Estacion meteorologica en el Pacifico antartico y su probable Rendimienta." *An. Soc. cient. argent.*, 139: 195–208 (1945); *Rev. Meteor.*, 5: 89–100 (1946). (Advantages and feasibility of station on Peter I Island.)

30. Dorsey, H. G., Jr., "Meteorology at East Base of U. S. Antarctic Expedition, 1939–1941." *Bull. Amer. meteor. Soc.*, 22: 389–392 (1941).

31. —— "An Antarctic Mountain Weather Station." *Proc. Amer. phil. Soc.*, 89: 344–362 (1945).

32. Escola, M. Z., "La Antártida en nuestra Meteorologia. Conveniencia de una Exploración Argentina al Interior de esa Continente." *Bol. Cent. nav., B. Aires*, 58: 711–725 (1940); "Antecedentes para una Expedición Científica Argentina a la Antártida." *Ibid.*, 59: 73–103 (1940).

*33. Giuffra, E., "La Circulación de la atmósfera en el hemisferio sur." *Rev. Meteor.*, 2: 123–198 (1943).

34. Howkins, G. A., "Developments in the Antarctic." *Weather*, 4: 157–158 (1949). (Activities of Falkland Islands Dependencies Survey.)

35. Minchin, C., "Weather and Pack Ice Encountered during the Season 1948–49 in Antarctic Waters." *Mar. Obs.*, 20: 50–51 (1950).

36. Mirrlees, S. T. A., "Notes on Southern Hemisphere Circulation." *Meteor. Mag.*, 78: 315–321 (1949). (Rymill 1935–36 data compared with South American data.)

37. Peterson, H.-C., *Antarctic Weather Statistics; Atmospheric Refraction Project; Results of the Solar Radiation Project; Weather Observing Program; Guide for Stonington Island Aviation Meteorology.* Ronne Antarctic Research Expedition. Office of Naval Research, Washington, D. C., 1948.

*38. Robin, G. de Q., "Notes on Synoptic Weather Analysis on the Fringe of Antarctica." *Meteor. Mag.*, 78: 216–226 (1949).

*39. Rouch, J., "La pression barométrique dans l'Antarctide américaine et l'anticyclone polaire." *Rev. gén. Sci. pur. appl.*, 41: 424–432 (1930). "La variation du niveau de la mer en fonction de la pression atmosphérique, d'après les observations du *Pourquoi-Pas* dans l'Antarctique." *Bull. Inst. océanogr. Monaco*, No. 870, 7 pp. (1944).

40. Rymill, J., *Southern Lights.* London, Chatto, 1938. (Monthly means of weather elements during 1935–37 on pp. 274–275; *cf.* [43].)

*41. Serra, A., *La Circulation générale de l'Amérique du Sud.* Rio de Janeiro, Servicio Nacional de Meteorologia, 1939. "The General Circulation over South America." *Bull. Amer. meteor. Soc.*, 22: 173–178 (1941).

*42. —— and Ratisbonna, L., *As Massas de Ar da America do Sul.* Rio de Janeiro, Servicio Nacional de Meteorologia, 1942. Translated as Rep. No. 403 of Weather Information Branch, Headquarters Army Air Forces, 1943. (Reprinted as NAVAER 50-1R-79 by Aerology Section, Chief of Naval Operations, 1944.)

43. Stephenson, A., "Notes on the Scientific Work of the British Graham Land Expedition: Meteorology." *Geogr. J.*, 91: 518–523 (1938). (*Cf.* [40].)

C. Ross Sea Sector (including New Zealand and Southwest Pacific)

44. Court, A., and Dorsey, H. G., Jr., "Antarctic Weather Observations." *Bull. Amer. meteor. Soc.*, 21: 386–387 (1940). (1939–41 plans.)

*45. Court, A., "Tropopause Disappearance during the Antarctic Winter." *Bull. Amer. meteor. Soc.*, 23: 220–238 (1942).

46. —— "Weather Observations during 1940–41 at Little America III." *Proc. Amer. phil. Soc.*, 89: 324–343 (1945).

*47. —— "Meteorological Data for Little America III." *Mon. Wea. Rev. Wash.*, Supp. No. 48, 150 pp. (1949).

48. Gabites, J. F., *South Pacific Meteorology.* Paper read at Amer. Meteor. Soc. Annual Meeting, New York City, Jan. 1949.

49. Grimminger, G., "Preliminary Results of Pilot-Balloon Ascents at Little America." *Mon. Wea. Rev. Wash.*, 67: 172–175 (1939).

*50. —— and Haines, W. C., "Meteorological Results of the Byrd Expeditions 1928–30, 1933–35: Tables." *Mon. Wea. Rev. Wash.*, Supp. No. 41, 377 pp. (1939). Grimminger, G., "Meteorological . . . : Summaries of Data." *Ibid.*, Supp. No. 42, 106 pp. (1941).

51. Guerrieri, E., "Climatologia dell'Antartide." *Ann. Inst. sup. nav. Napoli*, 9: 91–110 (1940). (Several tables from [50] compared with Abruzzi-Spitsbergen data 1899–1900.)

52. Haines, W. C., "Winds of the Antarctic." *Trans. Amer. geophys. Un.*, 13: 124–128 (1932).

53. —— "Meteorological Observations in the Antarctic." *Bull. Amer. meteor. Soc.*, 12: 169–172 (1931).

54. —— "The Green Flash Observed Oct. 16, 1929 at Little America by Members of the Byrd Antarctic Expedition." *Mon. Wea. Rev. Wash.*, 59: 117–118 (1931).

55. Harrison, H. T., "Antarctic Meteorology." *Mon. Wea. Rev. Wash.*, 59: 70–73 (1931). (Observing program of 1928–29 expedition.)

56. Hitchings, M. G., "Campbell Island—A Subantarctic Meteorological Station." *Weather*, 4: 389–392 (1949).

57. Kerr, I. S., *Mean Values of Heights of Tropopause, 300, 500, and 700 Millibar Surfaces at Auckland, Hokitika, and Taieri.* New Zealand Meteor. Off. Circular Note No. 43, 1947.

58. Lockhart, E. E., and Court, A., "Oxygen Deficiency in Antarctic Air." *Mon. Wea. Rev. Wash.*, 70: 93–96 (1942).

59. Loewe, F., "Zum Klima der Ross-Barriere." *Meteor. Z.*, 54: 28–30 (1937). (Advance data from [50].)

*60. —— and Radok, U., "A Meridional Aerological Cross Section in the Southwest Pacific." *J. Meteor.*, 7: 58–65 (1950).

61. Loewe, F., "Foehn Effects near the Balleny Islands." *Weather*, 5: 152–154 (1950).

*62. Palmer, C. E., "Upper Winds at Little America." *Trans. roy. Soc. N. Z.*, Pt. 4, 72: 311–323 (1942). (Reprinted as New Zealand Meteor. Off. Note No. 26, 1943.)

63. Pennnorf, R., "Die mittleren Temperaturverhältnisse der freien Atmosphäre am Rande des antarktischen Kontinents." *Meteor. Z.*, 60: 201–204 (1943). (Based on [50].)

*64. Ramage, C. S., *The Atmospheric Circulation of the Ross Sea Area.* New Zealand Meteor. Off. Prof. Note No. 2, 14 pp., 1944.

65. Rough, J., "La variation diurne de la température dans l'antarctique." *C. R. Acad. Sci., Paris*, 212: 94–95 (1941). (Based on [50].)

66. Tikhomirov, E., "Meteorologicheskie usloviia Rainoa Moria Rossa." (Meteorological conditions in the Ross Sea Region.) *Problemy Arktiki*, No. 12, pp. 19–28 (1939). (Based on [50].)

D. Australian and South Indian Ocean Sectors

67. Anonymous, "Synoptic Weather Maps for the Antarctic." *Nature, Lond.*, 161: 709 (1948). (Review of [74] and [75].)

68. Bureau, R., Douguet, M., et Wehrlé, P., "Radiosondages dans les mers australes." *C. R. Acad. Sci., Paris*, 208: 1419–1420 (1939). (Eight ascents at Kerguelen in February 1939.)

69. Campbell, S., "Australian Aims in the Antarctic." *Polar Rec.*, 5: 317–323 (1949). (Weather stations established on Heard and Macquarie Islands.)

70. De Monts, "Le front polaire dans l'Océan Indien." *Ann. Phys. Globe France d'Outre mer*, 2: 93–95 (1935); "La météorologie des Mascareignes commandées par les perturbations du front polaire." *Ibid.*, 3: 51–62 (1936).

71. Douguet, M., "L'Expédition antarctique du *Commandant Charcot.*" *Rev. marit.*, No. 41, pp. 1129–1142 (1949); "Observations du *Commandant Charcot* pendant l'expédition antarctique." *Ibid.*, No. 42, pp. 1313–1317 (1949).

72. G. C. S., "Review." *Quart. J. R. meteor. Soc.*, 74: 229–231 (1948). (Review of [75].)

*73. Gibbs, W. J., "A Period of Analysis for the Southern Ocean and the Implications in Southern Hemisphere Circulation." *Wea. Devel. Res. Bull., Aust.*, 12: 5–42 (1949).

*74. Kidson, E., "Discussions of Observations at Adélie Land, Queen Mary Land and Macquarie Island." *Aust. Antarctic Exped. 1911–1914, Sci. Rep.*, Ser. B, Vol. 6, 121 pp. (1946).

*75. —— "Daily Weather Charts Extending from Australia and New Zealand to the Antarctic Continent." *Ibid.*, Vol. 7, 405 pp. (1947).

76. Lamb, H. H., "Antarctic Discoveries." *Quart. J. R. meteor. Soc.*, 73: 192 (1947).

77. —— "A Meteorologist's Experiences on a Floating Whaling Factory." *Mar. Obs.*, 17: 75–83 (1947).

78. —— "Strange Moonrise in the Antarctic." *Mar. Obs.*, 17: 98–100 (1947).

79. —— "A Meteorologist in the Antarctic." *Meteor. Mag.*, 76: 231–234, 247–251 (1947).

80. —— "Australasian Antarctic Expedition, 1911–14." *Meteor. Mag.*, 77: 108–111 (1948). (Review of [74] and [75].)

*81. —— "Topography and Weather in the Antarctic." *Geogr. J.*, 111: 48–66 (1948).

82. —— "On the General Circulation of the Atmosphere in Middle Latitudes: Southern and Northern Hemispheres Compared." *Bull. Amer. meteor. Soc.*, 29: 391–394 (1948).

*83. —— "Scientific Results of the *Balaena* Expedition 1946–47." *Meteor. Mag.*, 78: 104–112 (1949).

84. —— "The Meteorological Results of the *Balaena* Expedition 1946–47." *Mar. Obs.*, (in press).

85. Langford, J. C., "A Study of Summer Cyclo-genesis and a Cold Outbreak." *Wea. Devel. Res. Bull., Aust.*, 11: 47–61 (1948). (Analyzes March 14–17, 1948 from Heard and Macquarie records.)

86. Loewe, F., "Das Klima von Adelie Land und der Macquarie Insel." *Meteor. Z.*, 52: 57–61 (1935).

*87. —— "Pressure Waves in Adélie Land." *Quart. J. R. meteor. Soc.*, 61: 441–445 (1935); "A Further Note on Antarctic Pressure Waves." *Ibid.*, 71: 344–349 (1945).

*88. Mawson, D., ed., "Records of the Queen Mary Station," "Meteorological Log of the S.Y. *Aurora*," and "Sledge Journey Weather Records." App., "Macquarie Island Weather Notes for 1909–1911." *Aust. Antarctic Exped. 1911–1914, Sci. Rep.*, Ser. B, Vol. 5, 281 pp. (1940).

*89. Palmer, C. E., *Synoptic Analysis over the Southern Oceans.* New Zealand Meteor. Off. Prof. Note No. 1, 1942, 38 pp. (Reprinted as NAVAER 50-1R-7, Aerology Section, Chief of Naval Operations, 1944.)

90. Thomas, M., "Les dépressions dans les Iles Australes." *Ann. Phys. Globe France d'Outre-mer*, 6: 95 (1939). (Noted in *Bull. Amer. meteor. Soc.*, 20: 323 (1939).)

E. African Sector (and East Weddell Sea)

*91. Flohn, H., "Grundzüge der allgemeinen atmosphärischen Zirkulation auf der Südhalbkugel (auf Grund der aerologischen Ergebnisse der deutschen antarktischen Expedition *Schwabenland* 1938–39)." *Arch. Meteor. Geophys. Biokl.*, (A) 2: 17–64 (1950); "Die planetarische Zirkulation der Atmosphäre bis 30 km Höhe." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 12, SS. 156–161 (1950).

92. King, J. A., and van den Boogaard, H. M. E., "Marion Island." *Weather*, 5: 26–30 (1950).

93. Kirk, T. H., "The Weather of the 1945–46 Antarctic Whaling Season." *Weather*, 2: 35–37 (1947).

94. Lange, H., und Regula, H., "Die Arbeiten der Expeditionens Wetterwerte. I. Terminenbeobachtungen, Höhenwindmessungen, Wetterdienst, Sonderuntersuchungen. II. Radiosondenaufstiege" in *Vorbericht über die deutsche antarktische Expedition 1938–39. Ann. Hydrogr., Berl.*, 67 (Supp.): 33–35 (1939).

95. Mosby, H., "The Sea Surface and the Air." *Sci. Res. Norweg. Antarctic Exped.*, No. 10 (1933); "The Waters of the Atlantic Antarctic Ocean." *Ibid.*, No. 11 (1934).

*96. Norske Meteorologiske Institutt, "Meteorological Observations Made on 9 Norwegian Whaling Floating Factories during the International Polar Year 1932–33." *Norwegian publications from the International Polar Year 1932–33*, No. 1 (1933). (Six-hourly observations on ships between 24°W and 134°E, October–April.)

*97. Regula, H., "Einige Ergebnisse der Radiosondenaufstiege der Deutschen antarktischen Expedition 1938–39." *Meteor. Rdsch.*, 2: 229–232 (1949).

98. Reuter, F., "Die Witterungsverhältnisse an der Kerguelen-Station." *Veröff. geophys. Inst. Univ. Lpz.*, (2) 5: 211–339 (1932).

99. Schneider-Carius, K., "Der Inversionstyp der Grundschicht." *Meteor. Rdsch.*, 1: 226–228 (1948). (Uses data of [101].)

100. Tauber, G., "Plavanie v Antarktike v 1947–48." (Voyage to the Antarctic in 1947–48). *Vsesoiuznoe Geograficheskoe Obshchestvo, Izvestiia* (Journal of the All Union Geographic Society) 81: 369–385 (1949). (Routine weather observations to aid in studying whales.)

II. OTHER WORKS CITED

A. Antarctic Meteorology before 1930

101. Barkow, E., "Die Ergebnisse der meteorologischen Beobachtungen der Deutschen antarktischen Expedition 1911–12." *Veröff. meteor. Inst., Univ. Berl.*, Abh. 7, Nr. 6 (1924). (See also Mill, H. R., "Review." *Meteor. Mag.*, 60: 149–150 (1925).)

102. Barlow, E. W., "The Wind Systems of the Arctic and the Antarctic." *Mar. Obs.*, 6: 245–248 (1929).

103. Hepworth, M. W. C., "Wind Systems and Trade Routes between the Cape of Good Hope and Australia." *Quart. J. R. meteor. Soc.*, 17: 21–27 (1891).

104. Hobbs, W. H., *Characteristics of Existing Glaciers.* New York, Macmillan, 1911. *The Glacial Anticyclone.* New York, Macmillan, 1926.

105. Lockyer, W. J. S., *A Discussion of Australian Meteorology.* London, Solar Physics Committee, 1909. *Southern Hemisphere Surface Air Circulation.* London, Solar Physics Committee, 1910.

106. Meinardus, W., "Die mutmassliche mittlere Höhe des antarktischen Kontinents." *Petermanns Mitt.*, 11: 304–309, 355–360 (1909).

107. —— "Tasks and Problems for Meteorological Exploration in the Antarctic." *Mon. Wea. Rev. Wash.*, 42: 223–230 (1914). Translated by Cleveland Abbé from "Aufgaben und Probleme der meteorologischen Forschung in der Antarktis." *Geogr. Z.*, 20: 18–34 (1914).

108. —— "Über den Wasserhaushalt der Antarktis." *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.*, SS. 184–192 (1925); "II. Der Wasserhaushalt der Antarktis in der Eiszeit." *Ibid.*, SS. 137–172 (1928).

109. —— "Die mittlere Höhe und Eisbedeckung der Antarktis." *Ibid.*, SS. 363–367 (1927).

110. Shaw, N., "Preface," in *Meteorology, Part I, National Antarctic Expedition 1901–4.* London, 1908.

111. —— *Manual of Meteorology*, Vol. 2. Cambridge, University Press, 1928, rev. ed., 1936.

112. Simpson, G. C., *Meteorology, Part I, Discussion, British Antarctic Expedition 1910–1913.* Calcutta, 1919.

113. —— *Scott's Polar Journey and the Weather.* (Halley Lecture, 17 May 1923) Oxford, Clarendon Press, 1926.
114. Taylor, G., *Antarctic Adventure and Research.* New York, Appleton, 1930. (See p. 192)

B. Antarctic Geography and Glaciology

115. Byrd, R. E., *Little America.* New York, Putnam, 1930. (See pp. 323–345)
116. —— "Our Navy Explores Antarctica." *Nat. geogr. Mag.*, 92: 429-522 (1947).
117. Croll, J., *Climate and Time.* New York, Appleton, 1897. (See pp. 375–382)
118. David, T. W. E., "Narrative" in Shackleton, E. H., *The Heart of the Antarctic.* London, Heinemann, 1909. (See Vol. 2, pp. 184, 187)
119. Debenham, F., "The Problem of the Great Ross Barrier." *Geogr. J.*, 112: 196–218 (1949).
120. Ellsworth, L., *Beyond Horizons.* New York, Doubleday, 1938. "The First Crossing of Antarctica." *Geogr. J.*, 89: 193–213 (1937). (See also Joerg, W. L. G., "The Topographical Results of Ellsworth's Trans-Antarctic Flight of 1935." *Geogr. Rev.*, 26: 454–462 (1936); "The Cartographical Results. . . ." *Ibid.*, 27: 430–444 (1937).)
121. English, R. A. J., "Preliminary Account of the United States Antarctic Expedition, 1939–1941." *Geogr. Rev.*, 31: 466–478 (1941). (See also various authors in "Reports on Scientific Results of the United States Antarctic Service Expedition, 1939–1941." *Proc. Amer. phil. Soc.*, 89: 1–400 (1945).)
122. —— *Sailing Directions for Antarctica.* H. O. 138, U. S. Navy Hydrographic Office, 1943. (See pp. 181–183)
123. Flint, R. F., *Glacial Geology and the Pleistocene Epoch.* New York, Wiley, 1947.
124. Gould, L. M., *Cold.* New York, Brewer, Warren and Putnam, 1931. (See pp. 150–162)
125. Hayes, J. G., *The Conquest of the South Pole.* New York, Macmillan, 1933.
126. Herrmann, E., "Die geographischen Arbeiten," in *Vorbericht über die deutsche antarktische Expedition 1938–39. Ann. Hydrogr., Berl.*, 67 (Supp.): 23–26 (1939).
127. Ronne, F., "Ronne Antarctic Research Expedition 1946–48." *Geogr. Rev.*, 38: 355–391 (1948). (See p. 375) (See also *Antarctic Conquest.* New York, Putnam, 1949)

C. General Meteorology

128. Brier, G. W., "40-Year Trends in Northern Hemisphere Surface Pressures." *Bull. Amer. meteor. Soc.*, 28: 237–247 (1947).
129. Chackett, K. F., Paneth, F. A., and Wilson, E. J., "Chemical Composition of the Stratosphere at 70 km Height." *Nature, Lond.*, 164: 128–129 (1949).
130. Court, A., "Insolation in the Polar Atmosphere." *J. Franklin Inst.*, 235: 169–178 (1943).
131. Gifford, F., "The Effect of Continentality in the Middle Troposphere over North America." *Bull. Amer. meteor. Soc.*, 29: 194–196 (1948).
132. Glueckauf, E., and Paneth, F. A., "The Helium Content of Atmospheric Air." *Proc. roy. Soc.*, (A) 185: 89–98 (1946).
133. Haurwitz, B., and Austin, J. M., *Climatology.* New York, McGraw, 1944. (See pp. 372–382)
134. Henry, T. J. G., and Armstrong, G. R., *Aerological Data for Northern Canada.* 271 pp., Dept. of Transport, Meteor. Div., Toronto, 1949.
135. Landsberg, H., "Climatology of the Pleistocene." *Bull. geol. Soc. Amer.*, 60: 1437–1442 (1949).
136. Leighly, J., "On Continentality and Glaciation" in *Glaciers and Climate. Geogr. Ann., Stockh.*, Vol. 31, Nos. 1 and 2 pp. 133–145 (1949).
137. Namias, J., "Characteristics of the General Circulation over the Northern Hemisphere during the Abnormal Winter 1946–47." *Mon. Wea. Rev. Wash.*, 75: 145–152 (1947).
138. Simpson, G. C., "Possible Causes of Change in Climate and Their Limitations." *Proc. Linn. Soc. Lond.*, 152: 190–219 (1940). See also "Ice Ages." *Nature, Lond.*, 141: 591–598 (1938), reprinted in *Ann. Rep. Smithson. Instn.*, pp. 289–302 (1938).
139. Willett, H. C., "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6: 34–50 (1949).
140. —— "Significant Correlation between Meridional Heat Transport and the Zonal Character of the General Circulation." *J. Meteor.*, 6: 370 (1949).

ARCTIC METEOROLOGY*

By HERBERT G. DORSEY, Jr.

Arctic Weather Central, Air Weather Service

ARCTIC ATMOSPHERIC CIRCULATION STUDIES

About a century ago theoretical considerations of the probable pressure distribution over the Arctic led to the postulation of a zonal system of westerly winds converging toward a center of low pressure at the North Pole. Shortly afterward an increasing amount of expedition data revealed a trend toward higher pressures and relatively frequent easterlies north of the great oceanic storm belts. These new data enabled Teisserenc de Bort to identify the large-scale features of the pressure distribution as early as 1881. Concurrent studies of the probable radiative cooling over the ice-covered polar basin were summarized by Helmholtz in 1888 in his first paper on atmospheric motion, in which he indicated that the relatively high pressure at sea level should be replaced by a low-pressure system aloft. During this period the role of polar air in the origin of storms was being investigated. A successful culmination to many important contributions was achieved in 1922 by J. Bjerknes and H. Solberg with their paper, "The Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation."

Perhaps worthy of brief comment at this point is the thought that possibly Bjerknes and Solberg's fundamental statement of modern meteorological principles in that paper was assumed by some glaciologists to be an attack on the "Glacial Anticyclones" of Professor William H. Hobbs. His theories, giving Greenland's "Glacial Anticyclones" a dominant role in the general circulation, were first published in 1910 [8] and have been rigidly reiterated up to now [10] despite a steady increase in adverse meteorological evidence[1] [5, 11].

Following the discussions appearing in *Problems of Polar Research*, compiled and published by the American Geographical Society in 1928, one of the earliest and most detailed studies to appear on weather conditions in the Arctic was that by Baur [1], presenting results in the form of monthly average charts for various elements, including temperature, cloudiness, and pressure. Baur showed in 1929 that the polar maximum of pressure was closely related to minimum temperature regions, and traced its movement from eastern Siberia in winter to a position north of the Arctic Archipelago in spring when the subarctic continental regions became relatively warm, with a continued eastward movement toward northeast Greenland and Spitsbergen in early summer. The decreasing central pressure was down to about 1014 mb at the summer minimum. In early autumn the Canadian Arctic Archipelago and the adjoining polar pack ice cool rapidly to attain a secondary maximum in the annual pressure trend, but this is soon exceeded by the Siberian anticyclone and its weaker counterpart over the Yukon Territory. Baur's charts, obtained from station pressures and corrected for large-scale anomalies presumably present in some of the shorter records, lose little in comparison with the indications of modern data.

The theoretical basis for much subsequent research on the heat balance of the earth and the general circulation of its atmosphere was published by V. Bjerknes in 1933 [2]. In the same year, Shaw [18] included in his comprehensive *Manual of Meteorology* charts of sea-level normal pressure and upper-air normal pressure charts (partially from earlier work of Teisserenc de Bort) for heights up to 8 km. However, Shaw's charts were not intended to be relied upon over the north polar region. The next sea-level pressure charts to be prepared for the Arctic were those resulting from the joint studies of Sverdrup, Petersen, and Loewe, published in the Köppen-Geiger *Handbuch der Klimatologie* in 1935 [19]. Their pressure patterns were quite similar to those of Baur.

Upper-air research conducted by the U. S. Weather Bureau was temporarily concentrated on a revision of Shaw's 2- and 4-km charts, after Wexler [23] investigated the formation of polar anticyclones. Many new data were available, and the final charts by Byers and Starr for January [4] gave the first definite indication that the elongated circumpolar vortex had separate closed centers, one over northeastern Siberia and the other over northern Hudson Bay. Subsequently, Namias and Smith [13] prepared monthly normal charts for the 10,000-ft level, then converted these to 700-mb contours. Revisions are made by Namias whenever new data suggest possible improvements in the work of the Extended Forecast Section, U. S. Weather Bureau.

In the great amount of theoretical research on the general circulation drawn upon by Rossby to obtain his schematic model of 1941 [15], and in later work, the north polar region has not been neglected. The principal results contributing to arctic meteorology are those pertaining to displacements of the principal centers of action. Appropriate reference will be made to some extended-range forecasting principles derived by Willett [24] which have significant value for arctic operational planning purposes. A recent contribution by Rossby [16] discusses the release of potential energy involved in the southward displacement of anticyclonic cold domes from high latitudes.

A 1946 publication of the U. S. Weather Bureau [20] contains monthly normal sea-level pressure maps for the Northern Hemisphere. These normals, computed from the noteworthy 40-yr series of daily synoptic charts, are no better over the polar basin than are the

---

* Submitted with the permission of the Air Weather Service. Conclusions and opinions are those of the author.

1. For an appraisal of the controversy over this point see "Some Climatological Problems of the Arctic and Sub-Arctic" by F. K. Hare, pp. 952–964 in this Compendium.

original daily charts. There were little data in high latitudes to restrict most of the original analysts, who seem to have had a fairly uniform tendency toward overemphasis of arctic anticyclones.

Petterssen [14] has good sea-level and upper-air pressure charts for January and July, but meteorologists must refer back to Sverdrup [19] for realistic seasonal pressure patterns, unless the new charts by Dzerdzeevskii [6] become more readily accessible. Dzerdzeevskii analyzed the data obtained by the Russian Polar Expedition, 1937–38, and included the results of previous studies in a new investigation of circulation patterns in the central Arctic. He shows that intense cyclonic and frontal activity was encountered near the Pole, and that the coldest air came from the American Arctic. His monthly average sea-level pressure charts are generally similar to those of Baur and Sverdrup, but fail to account completely for the midwinter secondary minimum in pressure over northern arctic America.

In 1949 an important study of the upper air over northern Canada was published by Henry and Armstrong of the Canadian Meteorological Service [7]. Using the data collected from the new aerological stations set up in the Canadian Arctic under the joint United States–Canadian program, the authors have carried out an exhaustive climatological analysis. Isotherms and contours are shown in map form for the 850-, 700-, 500-, and 300-mb surfaces for January, April, July, and October. The maps cover the greater part of Canada, including the Archipelago. Tabulated upper-air data are given for all stations, not only for the average year, but for each individual year of the record. Finally, there is an excellent discussion of the mean circulation over the Canadian Arctic. The work represents the most ambitious attempt yet published to clarify our picture of the circulation on the American side of the Arctic.

OPPOSED CONCEPTS OF ARCTIC CIRCULATION PATTERNS

It is difficult to believe that cooperating groups of modern scientists, such as meteorologists and explorer-geographers, could have sharply contrasting concepts of the Arctic, yet that rather surprising contention has been repeatedly expressed by Professor Hobbs. His restatement in 1948 of this viewpoint [10] was based on concepts of the atmospheric circulation patterns over two different arctic regions. Meteorologists have rarely published refutations of his contentions, but since his unacceptable theories have received wide publicity it now seems highly desirable to compare the meteorologists' concept of the Arctic with that ascribed to them by his analysis of supposedly contrasting concepts. The arctic data used by meteorologists and climatologists have been gleaned from expedition reports wherever required to supplement those from fixed stations. No conflict exists and no contrast can be found between explorers and meteorologists in this matter, despite Hobbs's statements to the contrary. While it has often been impossible to incorporate the roving explorer's data into current weather maps, these hard-won data have always been considered in thorough investigations of regional characteristics. Indeed, during the last two decades and even today, a major proportion of all field journeys into the Arctic and the Antarctic have had meteorological objectives.

Characteristics of the Central Arctic Basin. Since meteorologists and other scientists familiar with the atmospheric circulation of the Northern Hemisphere know that the continental interior regions of Canada and Siberia supply the polar air for the intense frontogenetic zones off the east coasts of North America and Asia, with Greenland and the arctic ice fields acting as secondary source regions because of the lesser contrasts at high latitudes, there is no clear reason for Hobbs's reference to Bjerknes [10], "... the hypothesis of the North Polar Front, or polar anticyclone, above the Arctic Basin was, at the time of its promulgation, contrary to all that had then been learned concerning its climate...." Over western Europe, Bjerknes' polar "high" is rarely an anticyclonic cell from the North Pole, but is more likely to be developed from an anticyclonic circulation in maritime polar air over the North Atlantic Ocean or in an extreme westward flow of continental polar air from Eurasia.

A remark that Nansen was able to *confirm* the Helmholtz theory would appear to be counter to all of Hobbs's elaboration on mean annual air pressure as observed by explorers in and about the Central Arctic Basin [10], but that is not intended nor is any such implication required to aid in the elimination of imaginary controversies between explorers and meteorologists. While meteorologists prefer to deal with average pressures on a monthly, semimonthly, ten-day, or five-day basis when analyzing atmospheric circulation patterns, they are quite ready to accept Hobbs's emphasis on the fact that the "normal" pressure isobar of "1013 millibars surrounds the Central Polar basin" [10]. Thereby he has shown it to be the area of *relatively* high pressure that Helmholtz anticipated, and with which meteorologists are familiar. This is true because what is "normal" pressure for the temperate zone is relatively "high" pressure north of the almost globe-circling belts of lower pressure associated with the Aleutian and Icelandic centers of cyclonic activity in winter, and with their related but westward displaced centers over northern Siberia and Canada in summer.

On profiles of Northern Hemisphere pressure the latitude of minimum pressure averages about 62°N when taken on an annual basis. Much of the winter contribution to the higher pressure north of 60°N comes from the North American anticyclone in the Yukon territory, and from part of the Siberian anticyclone. Only in early autumn, in late spring, and in summer, when the Arctic Basin is colder than continental areas, is there a relatively strong arctic anticyclone with pressures consistently higher than over polar continental areas. In summer the North Pole itself is more often in a weak trough or ridge situation between "lows" over northern Baffin Bay and the Laptev-Kara Sea areas, or between "highs" cresting over the Spitsbergen-

Fridtjof Nansen Land area and the Canadian Arctic Archipelago.

In the array of evidence supporting his "normal" pressure regime for the North Pole area, it appears that Hobbs would have done better to stress, rather than suppress, the fact that cyclonic storms have been encountered in the Central Arctic Basin. He cited Dzerdzeevskii as to the mean annual pressure of 1013 mb [10], but did not mention the frequent large fluctuations in pressure and temperature observed by the Soviet North Pole station before it drifted between northeast Greenland and Spitsbergen. Dzerdzeevskii noted correctly that these changes, reaching as much as 40 mb in pressure and 29C in temperature, were associated with the passage of intense cyclonic disturbances, frontal systems, and different air masses. Furthermore, popular accounts of the Soviet expedition also mention the stormy conditions experienced while they were still close to the North Pole: "North Pole, June 10th. There was a violent snow storm on the 8th and 9th ... the gusts of wind attained a speed of 60 feet per second (41 m.p.h.)" [3], and on June 29th, "A savage north wind has been raging for more than twenty-four hours. . . . Rain has been pouring down . . . " [3].

During the last few years, cooperative action between the United States, Canada, and Denmark has resulted in the establishment of new meteorological stations in the Canadian Arctic Archipelago and northern Greenland. They are all in locations formerly visited only by explorers, and the annual resupply of each constitutes an undertaking comparable to that expended on an entire expedition of former years. As for the North Pole area, of which Hobbs has said, "For all time it is likely that meteorologists . . . will be unable to gain any knowledge of the weather of the Polar Basin" [9], meteorological reconnaissance flights have been made more often than twice weekly over the American sector of the arctic ice-fields by the U. S. Air Weather Service through the last three years. Meteorologists have used the explorer's hard-won observations and are adding daily increments to their knowledge of the arctic region. Where is the sharp contrast between concepts and methods of explorer-geographers and professional meteorologists investigating the Central Arctic Basin?

RECENT IMPROVEMENTS IN ARCTIC OBSERVATIONAL DATA

New Stations in Northern Arctic America. This region had long been a "blind spot" on North American synoptic weather charts and was a logical focal point of meteorological planning when the end of World War II allowed consideration of necessary improvements in the existing observational network. Canada, the United States, and Denmark agreed upon the desirability of cooperative effort, and plans for several new stations were implemented in 1946. A joint Danish–U. S. Weather Bureau station was established and in full operation at Thule, Greenland, by October 1946. Air transportation was used in the spring of 1947 to install the second of the new stations, operated jointly by Canada and the United States, at the northern end of Eureka Sound, Ellesmere Island, Canada. Four additional stations in the Arctic Archipelago are now maintained by Canadian–United States cooperation, as follows: Resolute Bay, southern Cornwallis Island, since October 1947; Mould Bay, Prince Patrick Island, and Isachsen Peninsula, Ellef Ringnes Island, by air lift since the spring of 1948; and Alert, Ellesmere Island, by air lift since the spring of 1950. Locations in Peary Land, Greenland, have been considered for a possible future station in this new net.

The short period of record for these stations does not permit comprehensive statistical treatment of the new meteorological data, but the analysis of sea-level and upper-air charts over this region has become more accurate with each increment of observations. Furthermore, the predominance of the arctic air mass is such that its local and seasonal characteristics have been approximately determined from a relatively short series of upper-air soundings.

Weather Reconnaissance Flights. When the elimination of the North American meteorological "blind spot" was being planned, the problem of obtaining weather information from the American sector of the arctic pack ice received considerable attention. The consolidated ice fields in the central portion of this area do not drift as rapidly as does the North Pole ice toward the Greenland Sea, so a drifting station might stay there for several years. It was finally agreed that weather reconnaissance flights would furnish a maximum of meteorological data at a relative minimum of hazard for the personnel involved. In March 1947 the Air Weather Service was able to begin occasional reconnaissance flights near the 700-mb level over the region between Alaska and the North Pole. Two flights a week were planned for a trial period, but three or more flights a week near the 500-mb level were made at times during 1948–49.

Expedition Stations. A Danish expedition to northeast Greenland was in that area during 1948–49. Its main base, with radiosonde equipment, was located near the former Danmarkshavn. An outpost station, established and supplied by air lift, wintered at Brönlunds Fjord in Peary Land to obtain the first cold-season scientific data from this part of Greenland. Observations from both these stations appeared intermittently with the Greenland collective reports in 1948–50.

The arctic division of the French National Polar Expeditions, directed by Paul E. Victor, established an observatory on the Greenland Ice Cap in July 1949. A location (70°54′N, 40°42′W) approximating that of Wegener's "Eismitte" at an elevation of about 3000 m was selected in order to obtain data comparable with those [22] of the 1930–31 expedition. Ice-cap meteorology should be further clarified when these new data become available.

ARCTIC CIRCULATION PATTERNS

Average Monthly Sea-Level Pressure Charts. Consideration of the typical properties of arctic air should be facilitated by a prior knowledge of the large-scale air

currents normally flowing toward and across northern arctic America. Seasonal variations in arctic circulation patterns at the gradient wind level will be described in terms of average sea-level pressure charts for the midseason months: July, October, January, and April. These charts represent a personal synthesis of what are conceived to be the most realistic features appearing in the previously discussed charts of Baur, Sverdrup, and Dzerdzeevskii. In describing the seasonal trends in arctic circulation patterns, it was thought more logical to start with the minimum stage of air flow, represented by July conditions, than to break in near the annual maximum of atmospheric motion.

July. The chart of average sea-level pressure for the month of July (Fig. 1) shows the relative weakness of

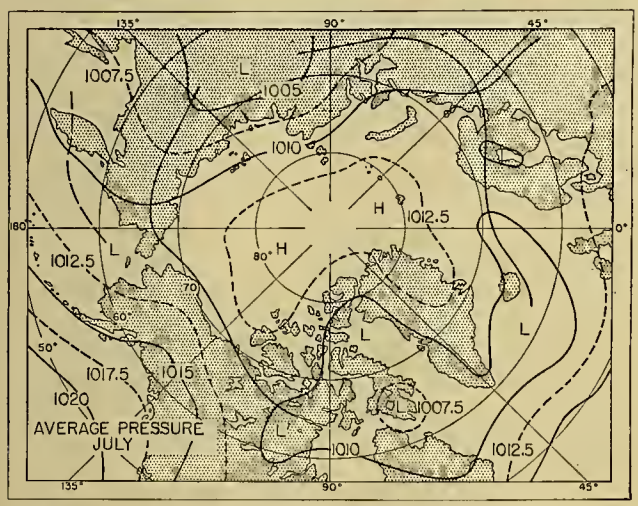


Fig. 1.—Average sea-level pressure (mb) during July.

the circulation patterns in this midsummer period of minimum thermal contrast between the polar and temperate zones. The circumpolar belt of low pressure is at its highest latitude in this month, with a definite extension of cyclonic activity toward the North Pole itself, although a feeble anticyclonic circulation persists over the congested pack-ice area northwest of the Canadian Archipelago and over the northeast Greenland–Fridtjof Nansen Land area. Along the Laptev Sea coast line, the contrast between warm air from the interior of Siberia and cool air from the pack ice is sufficient to maintain persistent cyclonic activity. Weak disturbances, originally associated with poleward thrusts of North Atlantic maritime air masses, may here be reintensified to resume their eastward progression toward the similar but milder thermal gradients along the arctic coast of Canada.

It has been mentioned that the July chart retains the best features derived by prior research of other arctic specialists [1, 6, 19]. The more widely circulated publication, "Normal Weather Maps, Northern Hemisphere, Sea-Level Pressure," [20] has presented a different pattern for July, especially over northern arctic America. With a 1017-mb isobar encircling the well-substantiated col area (average pressure less than 1013 mb) of the North Pole, and with little hint of a semi-permanent cyclonic system over northern Baffin Bay, the "40-yr normal July" could be very misleading. However "abnormal" the recent Julys may have appeared to be in this area, expedition reports indicate that midsummers over a hundred years ago were more dominated by cyclonic than by anticyclonic conditions.

Because of the simple thermal aspects of the July circulation scheme and the annual return to relatively the same percentages of ice-covered or ice-free surfaces and total solar heating, radical departures from the flow patterns described by this critique are not expected. Cyclonic activity develops and travels along the major frontal zones indicated by continuous thin lines on Fig. 1, except in some Julys when the Okhotsk-Bering Sea frontal zone becomes more active. The secondary cyclonic centers may be occupied by individual low-pressure cells, depending on the relative strength of the zonal circulation in the observed July. Frontogenesis in the North Greenland–Ellesmere Island topographic trough is associated with the reintensification of northern Baffin Bay disturbances and their subsequent rapid movement toward the North Pole. The arctic anticyclone usually centers on one or the other of the lobes shown, but not on both.

October. Comparison of the average October pressure patterns (Fig. 2) with those of the Weather Bureau

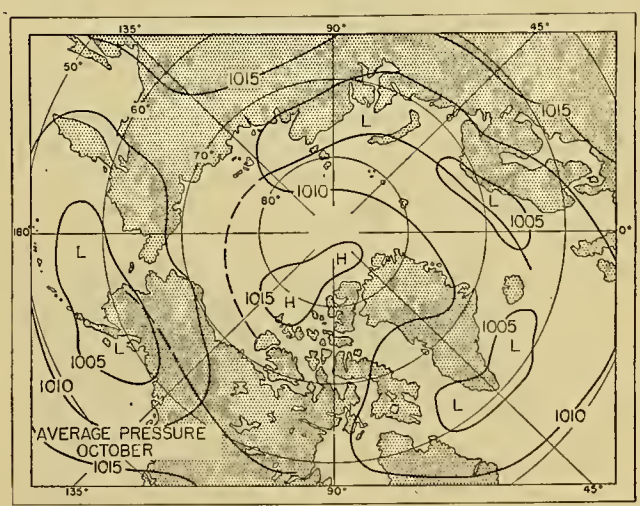


Fig. 2.—Average sea-level pressure (mb) during October.

"normal" map reveals errors of greater importance than for July, considering the increased zonal circulation and the onset of autumn storms. The many years of few data and extrapolated analyses apparently resulted in too great an over-all smoothing and in a loss of knowledge regarding the *terrain-imposed location of cyclonic centers associated with variations in the general circulation.* Again the arctic high is overemphasized, although shown at a *lower* central pressure than July. The important eastward extension of North Atlantic cyclonic activity to the Kara Sea is not definitely shown, nor is there sufficient indication that two cyclonic cells are to be expected in the Icelandic system. The arctic high tends to be centered over North Greenland or west

of the Canadian Archipelago. It rarely spreads over both.

January. Among the many sources of "normal" pressure patterns it is believed that none except the 40-yr series have a January chart showing higher pressure over the Arctic Ocean than in the Canadian Yukon Territory. An occasional January mean map will have an Arctic Ocean high, January 1948 for example, but such a January is known to be relatively rare. By comparison with Fig. 3, the Weather Bureau January map

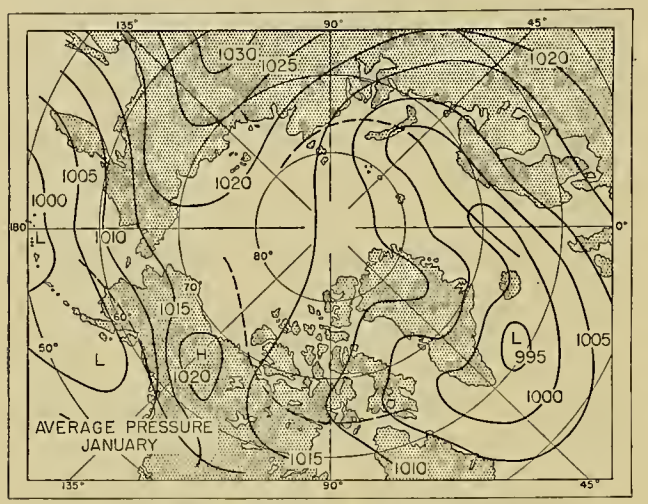


Fig. 3.—Average sea-level pressure (mb) during January.

fails to indicate the true intensity and relative frequency of cyclonic activity in the Kara Sea, nor does it place sufficient emphasis on Greenland's effect as a topographic barrier.

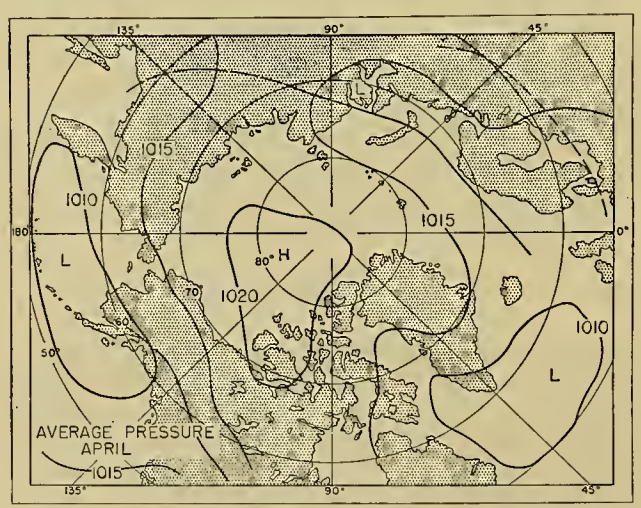


Fig. 4.—Average sea-level pressure (mb) during April.

April. By this time the mid-latitude continental warming has become well established although the Arctic remains cold with predominantly anticyclonic conditions. Pressures over the Canadian Archipelago and northern Greenland reach their annual maximum in April (Fig. 4). Arctic air flow southward over eastern North America is favored by the reduced zonal circulation.

The principal weakness in the Weather Bureau "normal" pressure map for April is found in its failure to indicate the southwestward displacement of the North Atlantic cyclonic activity to an annual minimum latitude. The arctic anticyclonic circulation rarely departs from the single-cell pattern indicated for April, but its center may be located over northeast Greenland or the Canadian Archipelago when large-scale circulation anomalies disrupt the average pressure distribution.

Average 700-mb Contour Patterns. A survey of previous research on the upper-level circulation over the Arctic was presented earlier in this article. The use of mean circulation methods developed by Rossby, Willett, and Namias has emphasized the importance of major anomalies in flow patterns aloft. Extended forecasting for the Arctic and elsewhere requires almost continuous reference to the upper-level "normal" maps published by the U. S. Weather Bureau [12, 13]. As a result of this writer's research, it appears that many of the errors found in the 40-yr sea-level normals have been extrapolated up to the 700-mb charts. Namias [13] used many short-term radiosonde records, especially for high-latitude stations, by applying a correction based on the concurrent departure from normal of the local sea-level pressure and surface temperature. In preparing July and January average 700-mb contour charts for this article, the elimination of errors attributable to incorrect "normal" sea-level pressures has received primary emphasis.

July. The suggested revision of the Namias 700-mb contours for July is shown in Fig. 5. A comparison with

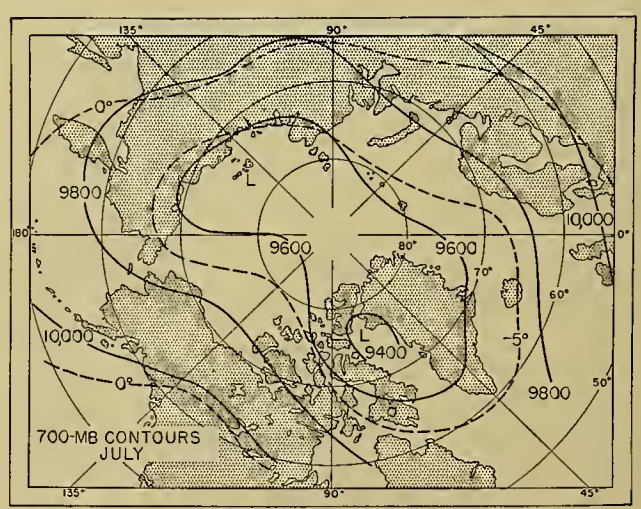


Fig. 5.—Average 700-mb contours (ft) and isotherms (°C) during July.

the corresponding Namias [12] chart shows the reduced heights over northern Baffin Bay and the Laptev Sea that are associated with the average greater than "normal" cyclonic activity mentioned in the earlier criticism of the July "normal" map. Reference has been made to Petterssen's summer charts of 2-km and

4-km isobars [14] and to the new series of Historical Weather Maps, for additional support of the height increases required over the Okhotsk and Bering Seas.

The air flow over northern arctic America, and over the Siberian Arctic, is seen to be predominantly cyclonic at the 700-mb level in July, closely corresponding to the gradient circulation derived from average sea-level pressure. Weak anticyclonic flow is frequently observed over the northeast Greenland–Fridtjof Nansen Land area and western sections of the American Arctic, but is rare and of brief duration over northern Baffin Bay.

January. The major revisions appearing in the new 700-mb chart for January (Fig. 6) are directly related

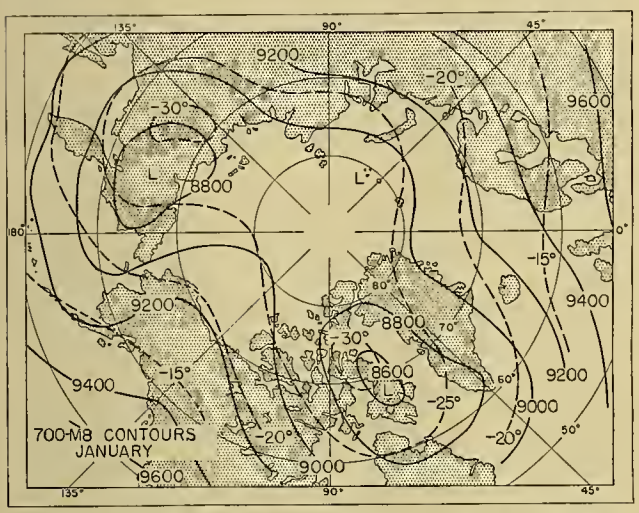


Fig. 6.—Average 700-mb contours (ft) and isotherms (°C) during January.

to the areas of excessive sea-level pressure, as previously noted over northern Baffin Bay and over the Kara Sea, on the Weather Bureau "normal" map. With the introduction of more realistic perturbations in the circumpolar westerlies over these areas came an equally logical intensification and southward displacement of the east Asiatic cyclonic cell. It is suggested that the intensity of the American cell will be exceeded by that of the Asiatic cell only during the anomalous absence, or complete enclosure, of the Barents-Kara perturbation.

The cold-season cyclonic circulation over northern arctic America, being mainly a dynamic system of cold dry air, is relatively cloud free except during periods of reintensification.

SEASONAL VARIATIONS IN ARCTIC AIR

Variations over Northern Arctic America. The foregoing summary of average circulation patterns, and the geographical location of the American Arctic, indicate that all the large-scale air currents will have been subjected to arctic influences before their arrival over this region. The region itself provides the cold source area of arctic air masses affecting central and eastern North America. Variations in the observed properties of arctic air are directly related to the seasonal variations of the underlying surfaces and the air flow over them. The vertical structure of arctic air over the Canadian Archipelago during the different seasons of the year is shown by the average soundings for July, October, January, and April, which are plotted in Fig. 7.

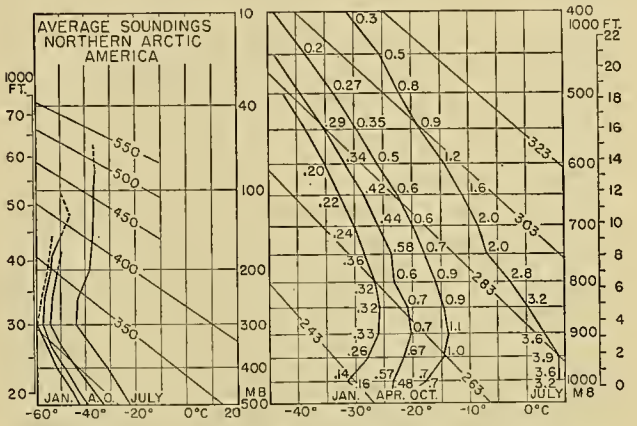


Fig. 7.—Average soundings over northern arctic America.

July. The average sounding for July, based on 28 ascents to the 250-mb level obtained at the Eureka station during July 1947, shows the midsummer structure of arctic air over ice-free interior areas. Because of the predominantly cool and moist underlying surface conditions, the regional air mass is maritime arctic in summer. There is a characteristic surface inversion over the pack ice and cold water, but convective instability in the lower levels is readily produced by land heating after a short period of onshore flow. A moisture content of about 4 g kg^{-1} near the surface is fairly representative of the regional air masses. The cyclonic circulation around northern Baffin Bay and the associated middle-cloud system create a characteristic moisture inversion near the 700-mb level of the average July sounding. During this individual month, the average height of the tropopause was about 9000 m (29,500 ft).

October. In sharp contrast to the relative warmth caused by 24-hr insolation in summer, the autumn conditions of almost no solar heating and predominantly frozen surfaces are relatively ideal for the rapid modification of maritime polar air. The average sounding for October, derived from 27 Eureka ascents to the 250-mb level in October 1947, shows well-developed continental characteristics at this early stage in the transition to the deep, cold, continental arctic air of winter. The average October circulation pattern was closely approximated in 1947. Moderate anticyclonic conditions dominated northern arctic America, and the sounding has no obvious traces of Baffin Bay cyclonic activity. The tropopause was about 8600 m (28,000 ft) high, 400 m lower than in July. Continental arctic air of this type begins to form over the pack ice during early autumn, but it becomes unstable and produces heavy snow showers as it flows southward over unfrozen waterways and warmer land masses in the rear quadrants of autumn storms.

January. The average sounding for January, based on 21 ascents to the 350-mb level above the Resolute

Bay station in 1948, shows the vertical structure of continental arctic air as it flows southward from the Canadian Archipelago in the Baffin Bay cyclonic circulation. The arctic air arrives at Resolute Bay after passing over frozen waterways en route from a colder source region, exhibiting characteristic surface turbulence and instability. Specific humidity values are near the annual minimums of 0.2 to 0.3 g kg^{-1} with little variability throughout the deep and surface-controlled thermal equilibrium below the 700-mb level. The average height of the tropopause is about 8300 m (27,500 ft) in winter.

January 1948 had a number of striking departures from the normal circulation patterns. The existence of an arctic "high" has been mentioned, and that—in conjunction with an active arctic frontal zone across northern Canada—created an unusually strong flow across southern sections of the Archipelago. Thus, the January average sounding probably shows more warmth and moisture than would be observed in a long-term average.

April. Despite the 24-hr insolation in spring over the region investigated, there is little immediate change in the surfaces underlying arctic air masses, and April retains much of winter's coldness. The average sounding, from 26 ascents to the 250-mb level above Resolute Bay in April 1948, is still typical for continental arctic air but shows a distinct moderation from the severe winter characteristics of the regional air mass. The winter temperature contrast between ice fields and land has been eliminated, allowing an increase in surface stability, but the Baffin Bay cyclonic circulation extends back over Resolute Bay above the 800-mb level. Apparently the tropopause height varies between the 7500- and 8500-m levels in April.

The American Arctic Versus the Eurasian Arctic. A brief survey of the Eurasian arctic data has provided a few "typical" soundings which are included in Fig. 8 to illustrate characteristic regional variations in arctic air. Because of its high latitude and great distance from the major frontal zones, northern arctic America produces a distinctive continental type of arctic air.

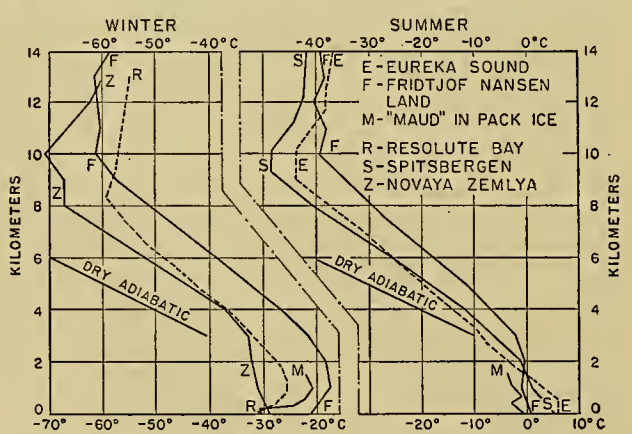


Fig. 8.—Comparative soundings.

With reference again to Fig. 8 for summertime comparisons, it will be seen that the surface temperatures are lowest over the pack ice and highest over the relatively large ice-free land areas of northern arctic America. The coldness and low-level instability of the *Maud* sounding result from the incorporation of June and August data in the summer "average," and from the fact that steady winds and stirred air were necessarily present during successful kite flights. The Fridtjof Nansen Land sounding shows relatively warm maritime polar air, cooled in its lower levels over the arctic ice fields, and brought onshore in a strong return flow over cold water. Maritime arctic air reaches Spitsbergen in a steady flow across relatively warm water. Land heating of the coldest type of maritime arctic air produces convective instability over arctic America, but near icy waterways the lowest levels are stabilized by the sea-breeze circulation. The tropopause heights show maritime influences over Fridtjof Nansen Land, but Eureka and Spitsbergen have the typical, uncomplicated arctic tropopause, with Eureka's lower because of greater insolation aloft and proximity to a "normal" upper-level vortex.

In winter the Arctic Archipelago is colder than the pack ice, but the Baffin Bay circulation prevents the formation of extreme temperature inversions of the pack ice or continental interior type. The Fridtjof Nansen Land sounding shows maritime polar air after a relatively short period of cooling from below, with a maritime polar tropopause being frequently observed. Many of the "winter" soundings at Novaya Zemlya were made in the rear quadrants of intense storms. They show evidence of strong stirring in continental arctic air and a complex tropopause mixture of low arctic and extremely cold maritime types. The Resolute Bay sounding shows the comparatively stable continental arctic air extending all the way to a warm arctic tropopause.

In addition to the temperature-versus-height curves compared in Fig. 8, the annual range of average moisture and stability conditions in Eurasian and American arctic air masses is shown by means of a brief tabular comparison (Tables I and II). The Eurasian data for arctic air in winter and summer were obtained from Petterssen [14]. The physical properties chosen for comparative purposes, specific humidity (q) and potential psuedo-wet-bulb temperature (θ_{sw}), are conservative

Table I. Arctic Air in Winter

Station	p (mb)	H (m)	T (°C)	q (g kg^{-1})	θ_{sw} (°C)
Resolute Bay (all soundings)	1000	75	−31	0.16	−32
	900	800	−26	0.32	−19
	800	1650	−26	0.30	−13
	700	2600	−30	0.25	− 9
	600	3700	−35	0.22	− 4
	500	4950	−43	0.06	− 2
Fridtjof Nansen Land (stirred air)	1000	—	—	—	—
	900	750	−30	0.29	−24
	800	1550	−34	0.21	−20
	700	2450	−38	0.16	−15
	600	3500	−42	0.12	−10
	500	4650	−49	0.07	− 6

indicators of moisture content and stability, and are easily computed from radiosonde data. The tabulated data are not strictly comparable, since the Eurasian soundings were selected to include only fresh outbreaks of arctic air, while only the average January and July soundings were available for the American Arctic. Even so, there was a strong tendency for decreasing θ_{sw}-values between 950 and 850 mb above Eureka Sound in July. Individual ascents would certainly show convective instability similar to that observed along the Barents Sea Coast.

TABLE II. ARCTIC AIR IN SUMMER

Station	p (mb)	H (m)	T (°C)	q (g kg^{-1})	θ_{sw} (°C)
Eureka Sound (all soundings)	1000	30	+ 6	3.6	+ 3
	900	900	+ 1	3.6	+ 5
	800	1800	− 3	2.8	+ 7
	700	2800	− 8	2.0	+ 8
	600	4000	−14	1.2	+10
	500	5400	−22	0.8	+12
Barents Sea Coast (stirred air)	1000	—	+ 9	5.1	+ 6
	900	850	+ 2	3.7	+ 5
	800	1750	− 2	3.5	+ 8
	700	2800	−11	1.8	+ 6
	600	3850	−16	0.8	+ 8
	500	5150	−27	0.3	+ 8

TYPICAL ARCTIC CIRCULATION PATTERNS

Variations in the Average Circulation Patterns. The discussion of the midseason average arctic circulation patterns, based on maps of average pressure for July, October, January, and April, has already indicated the major centers of cyclonic and anticyclonic activity and traced the frontal zones associated with the development and movement of cyclonic storms. A knowledge of the major variations to be expected in the average flow patterns is always a valuable aid in the analysis and prognosis of daily weather situations, especially in the Arctic where observing stations are widely separated and where there is not much past experience to draw upon. Much useful information on anomalous circulation patterns has been gained from research on forecasting by mean circulation methods, but no specific studies on the truly arctic regions have been made.

Since this article is concerned with a region which has not entirely ceased to be a "blind spot" on Northern Hemisphere charts, a general application of mean circulation principles will be described for the principal seasons. A more detailed application would require extensive research along synoptic lines as suggested by Willett [24].

Summer Variations. The strong thermal control of the average summer circulation has been mentioned. Thus, as shown by Schell [17], large-scale anomalies in the semipermanent ice and snow cover would undoubtedly be locally important, but could hardly counteract an opposing anomaly in the general circulation. A strong summer circulation (high index) might at first be considered unfavorable for polar anticyclonic developments, but it is thought that summer conditions are quite probably the reverse of winter. It is the subtropical anticyclones, at unusually high latitudes over the oceans, that are associated with the periods of weak zonal index in summer. At such times, the circumpolar cyclones are forced into the Arctic much more persistently than is normal.

The fact that summer circulation patterns tend to persist and change but slowly should in itself be of considerable guidance to forecasters in the high latitude arctic regions—provided they have at least one daily set of surface and upper charts with a reasonable synoptic coverage.

Late spring and early fall disturbances may be strong, but the cyclonic systems of the true summer period are relatively weak unless abnormally cool conditions prevail over the Arctic Ocean area. As suggested by the average July chart, disturbances coming from Siberia are often reintensified along the Canadian Arctic coast line. Their subsequent movement depends mainly on the position and intensity of the Baffin Bay upper-level cyclonic circulation. Normally no summertime disturbances approach arctic America from the direction of the North Pole, though northern Baffin Bay cyclonic systems frequently move poleward up the North Greenland-Ellesmere trough. Some part of almost every disturbance reaching the vicinity of Hudson Bay from the west or south will eventually affect the northern Baffin Bay section. Summer occurrences of other seasonal patterns will be noted.

Autumn Variations. Early autumn is a season of strong thermal contrast between the Arctic, which is already producing a winter type of air mass, and the more slowly cooling subarctic continental mainland areas with their summer-type maritime polar air. The average circulation pattern is one of strong zonal westerlies, representing a great increase over average summer conditions, but usually there is an anomalous return to a summer type after the first onset of autumn. Continental mainland areas warm up again and, with the high latitude Arctic cooling rapidly, the stage is set for a strong intensification of weak disturbances crossing coastal sections of Siberia and/or Canada. With specific reference to arctic America, deepening over the Mackenzie or Coppermine valleys completes the genesis of extreme types of autumn storms, which will curve into the Archipelago or northern Baffin Bay and dominate the circulation for possibly a week as they slowly fill. Weaker variations of the above anomalous type account for a high percentage of summer disturbances affecting this region. As this is also true for autumn and early winter, the so-called "autumn type" is very important to northern arctic America.

Winter Variations. Since the cold season includes November and March, minor variations of the average January circulation pattern are common throughout the winter. Strong zonal westerlies are characteristic of the average pattern. Circulation anomalies of the high-index type frequently result in a pattern similar to that of October, with the required strengthening of pressure gradients. Except for occasional high-latitude cyclonic eddy systems, northern arctic America is usually cut

off from the subpolar cyclonic disturbances at such times and has a preponderance of fine weather. On the other hand, the meridional flow patterns of low index may be similar to an accentuated pattern of the April type, with intense low-latitude storms recurving into arctic areas. Much of the cold season storminess affecting northern arctic America is associated with this type of development which will occasionally equal or exceed the intensity of the autumn-type circulation anomaly.

Spring Variations. The continuation of winter cold in the Arctic, while the continents are warming, results in exceptionally favorable conditions for anticyclonic developments over high-latitude arctic regions. April's average circulation pattern is typical of the best spring weather, but anomalies in the general circulation may cause the April pattern to appear in March and again in May, with an April pattern resembling January's strong circulation. An April anomaly of the above type does not seriously affect arctic America, except rarely when an Atlantic storm makes the polar circuit and approaches the Canadian Archipelago from the northwest. Variations of the winter low-index type are the more frequent and more serious circulation anomalies affecting arctic areas especially in early spring.

ARCTIC METEOROLOGICAL PROBLEMS

Observational Data. Because our knowledge is severely limited as to the relationship between the arctic tropospheric and stratospheric circulations themselves, and because our understanding of the more general relationship between the arctic and the hemispheric circulation is by no means complete, it is difficult to visualize or suggest what might eventually constitute a representative coverage of arctic meteorological data. The problem of observational coverage was easier to resolve in 1945 when the synoptic reporting network for the North American Arctic was entirely inadequate. Since then the situation in that area has been greatly improved by the establishment of the new stations and by the weather reconnaissance described in the third section of this article.

At this time the existing arctic weather station network (and the availability of data from these stations) is probably adequate for the preparation of routine synoptic analyses. However, the existing situation is not entirely satisfactory with respect to the availability of upper-air research data from the Eurasian Arctic. To solve this problem it will be necessary to achieve a higher degree of cooperation than apparently exists in the international exchange of current arctic meteorological data. Granting the relative adequacy of arctic tropospheric data, trends in current and planned meteorological research indicate that arctic soundings to altitudes well above 120,000 ft will be desirable when similar data from mid-latitude soundings to the ozonosphere are available. These high-level observational data will be required in connection with the continuation and expansion of studies on the general atmospheric circulation, on its large-scale short- and long-period fluctuations, and on the possible linkage between solar activity and tropospheric meteorological phenomena.

Synoptic Studies. Because of the short period for which relatively adequate data have been available, there are several aspects of arctic synoptic meteorology which have not received thorough consideration. Valuable contributions to existing information should result from studies on the connections and interactions between the state of the general circulation and semipermanent systems, such as, the Aleutian low, Icelandic low, Siberian high, polar high, and North American high.

There can be little doubt as to the desirability of undertaking further studies on the general circulation as regards that of the Arctic in relation to the tropospheric jet stream of middle latitudes.

It also is suggested that another worth-while objective for future investigations would be the comparatively unknown nature and extent of control imposed on the arctic circulation by contrasts between continental and oceanic regions in the Arctic and in adjacent temperate areas.

Basic Research. Investigations of a more fundamental nature than the suggested and expected investigations of synoptic relationships between arctic flow patterns and the hemispheric circulation will be undertaken because of the simple fact that the Arctic must necessarily be involved in much basic research on the dynamics of the earth's atmosphere. An associated product of such studies will be the accomplishment of essential research on the interactions between arctic and middle-latitude regions with respect to transfer of mass, momentum, angular momentum or vorticity, entropy, heat content and moisture content, and the propagation of horizontal oscillations or perturbations between different latitudinal zones. Attainment of these research objectives would concurrently require solutions to the problem of heat balance and exchange by various causes, such as radiation, conduction, convection, advection, and latent heat changes.

REFERENCES

1. Baur, F., "Das Klima der bisher erforschten Teile der Arktis." *Arktis*, 2:77–89, 110–120 (1929).
2. Bjerknes, V., and others, *Physikalische Hydrodynamik*. Berlin, J. Springer, 1933. (See Chaps. 14 and 15)
3. Brontman, L., *On the Top of the World*. London, Gollancz 1938. (See pp. 244, 255)
4. Byers, H., and Starr, V., "The Circulation of the Atmosphere in High Latitudes during Winter." *Mon. Wea. Rev. Wash.*, Supp. No. 47 (1941).
5. Dorsey, H. G., Jr., "Some Meteorological Aspects of the Greenland Ice Cap." *J. Meteor.*, 2:135–142 (1945).
6. Dzerdzeevskii, B. L., *Tsirkuliatsionnye skhemy v troposfere Tsentral'noi Arktiki*. Moskva, Izdatel'stvo Akademii Nauk (SSSR), 1945.
7. Henry, T. J. G., and Armstrong, G. R., *Acrological Data for Northern Canada*, 271 pp. Dept. of Transport, Meteor. Div., Toronto, May 1949.
8. Hobbs, W. H., "Characteristics of the Inland-Ice of the Arctic Regions." *Proc. Amer. phil. Soc.*, 49:57–129 (1910).
9. —— "The Greenland Glacial Anticyclone." *J. Meteor.*, 2:143–153 (1945).

10. —— "The Climate of the Arctic as Viewed by the Explorer and the Meteorologist." *Science*, 108:193–201 (1948).
11. Matthes, F. E., "The Glacial Anticyclone Theory Examined in the Light of Recent Meteorological Data from Greenland." *Trans. Amer. geophys. Un.*, 27:324–341 (1946).
12. Namias, J., *Extended Forecasting by Mean Circulation Methods*. U. S. Weather Bureau, Washington, D. C., 1947.
13. —— and Smith, K., *Normal Distribution of Pressure at the 10,000 Foot Level over the Northern Hemisphere*. U. S. Weather Bureau, Washington, D. C., 1944. (See pp. 2, 5)
14. Petterssen, S., *Weather Analysis and Forecasting*. New York, McGraw, 1940. (See pp. 149–186)
15. Rossby, C.-G., "The Scientific Basis of Modern Meteorology" in *Climate and Man: Yearbook of Agriculture*. Washington, D. C., U. S. Govt. Printing Office, 1941. (See pp. 599–655)
16. —— "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6:163–180 (1949).
17. Schell, I. I., "Polar Ice as a Factor in Seasonal Weather." *Mon. Wea. Rev. Wash.*, Supp. No. 39, pp. 27–51 (1940).
18. Shaw, Sir Napier, *Manual of Meteorology*, Vol. 2, *Comparative Meteorology*. Cambridge, University Press, 1936.
19. Sverdrup, H. U., Petersen, H., und Loewe, F., "Klima des kanadischen Archipels und Grönlands," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. II, Teil K. Berlin, Gebr. Bornträger, 1935.
20. U. S. Weather Bureau, *Normal Weather Maps, Northern Hemisphere, Sea-Level Pressure*. Washington, D. C., U. S. Weather Bureau, 1946.
21. Wade, F. A., "Wartime Investigation of the Greenland Ice Cap and Its Possibilities." *Geogr. Rev.*, 36:452–473 (1946).
22. Wegener, K., *Wissenschaftliche Ergebnisse der deutschen Grönland Expedition Alfred Wegener, 1929 und 1930–31*. Bd. IV, Teil I. Leipzig, Brockhans, 1933.
23. Wexler, H., "Formation of Polar Anticyclones." *Mon. Wea. Rev. Wash.*, 65:229–236 (1937).
24. Willett, H. C., "Report on an Experiment in Five-Day Weather Forecasting." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 8, No. 3 (1940). (See pp. 36–45)

SOME CLIMATOLOGICAL PROBLEMS OF THE ARCTIC AND SUB-ARCTIC

By F. KENNETH HARE

McGill University

INTRODUCTION

The Arctic is pioneer territory for the climatologist. For generations, students of the atmosphere have depended for their views of the arctic circulation upon hypotheses rather than facts. Because it has lain beyond the reach of large-scale observation, the Arctic has been the happy hunting ground of partisan theorists. One sometimes suspects that the compilers of world maps of pressure, temperature, and precipitation distribution have breathed a sigh of relief when they reached the Arctic, for here at last was a region where statistics were rarely troublesome. For every painstaking study of the calibre of Sverdrup's work on the *Maud*, or Georgi's at Eismitte, there have been tens of superficial interpretations.

Until World War II the initiative in the climatological study of the Arctic rested with the Scandinavians, and to some extent (chiefly since 1930) with the Soviet Union. The growth of air transport and the unstable international situation have altered the North American viewpoint, and both Canada and the United States are now belatedly committed to a large-scale meteorological invasion of high latitudes. That it has taken so long to awaken our interest in this vital field, and then only under the grimmest necessity, is no tribute to our scientific initiative.

Throughout the Arctic, lack of observational data is an acute problem. Except along the subarctic margin, long observational series are extremely rare. Enormous areas, like the permanent pack ice of the Arctic Ocean, are entirely unknown. Though the situation is improving rapidly, it will be some time before thoroughly standardised climatological normals will be available for enough stations to allow the preparation of reasonably accurate distribution maps across the Arctic. Even then the pack-ice belt will remain the largest observational gap in the Northern Hemisphere.

It is not proposed in this article to attempt a general account of the physical climatology of the Arctic. A brief review of the more important literature will be presented, but thereafter attention will be directed towards certain specialised fields in which scholarly research is possible with existing materials. For the moment these seem to have more to offer than broad syntheses.

The terms "Arctic" and "sub-Arctic" require definition. The Arctic is usually regarded as consisting of the tundra lands beyond the poleward limit of tree growth. The isotherm of 50F (10C) for the warmest month happens to coincide very roughly with the limit, and is often taken as the arctic boundary. At sea, those areas heavily affected by coastal and pack-ice are properly regarded as arctic. The sub-Arctic has no accepted connotation. The author usually applies it to the Boreal forest region of the Northern Hemisphere, that is, the great ring of coniferous forest stretching from Alaska to Labrador and from Norway to eastern Siberia. Neither definition, however, is rigorously adhered to in the following account.

Previous General Literature. Apart from certain brief reviews like that of Nordenskjöld [45], no general climatology of the arctic regions has been attempted. The available literature is thus regional in scope and must be discussed on this basis.

Over the Arctic Ocean, existing knowledge still depends very largely on the published results of the *Fram* [44] and *Maud* [54] expeditions, and to some extent from less accessible Russian sources. The best general summary remains that of Sverdrup [55]. Though published fifteen years ago, Sverdrup's distribution maps, and the data published in the *Maud* results, are the usual sources today of climatic maps of the inner Arctic.

In North America, the most comprehensive accounts are those of Ward, Brooks, Connor, and Fitton for Alaska [63], Sverdrup for the Canadian Archipelago [55], and Connor for the Canadian mainland [8]. All these are parts of the Köppen-Geiger *Handbuch der Klimatologie*, published in the middle 1930's, and are now considerably out of date. Since then various summary accounts have appeared in official papers, but few have been given wide circulation. The Canadian Department of Transport published a pamphlet on the meteorology of the Canadian Arctic in 1944 [6], the scope being actually climatological. The only other general work of consequence was a report by Döll [10] on the climate of the Labrador coast, as exemplified by the records of the Moravian Missions, which have maintained climatological diaries for many years.

Since 1945 the newly accumulated records from this region have allowed more intensive studies. Though as yet unpublished, detailed works are available on various parts of the Canadian North, many of them in thesis form. Among these we may list studies by Montgomery [42] on coastal Labrador, Hare [19] on the eastern Canadian Arctic and sub-Arctic, Wonders [68] on the Canadian Archipelago, and Currie [9] on the Keewatin-Mackenzie district.

The Greenland region was discussed by Petersen (coastal region) [55] and Loewe[1] (inland ice) [55] in the Köppen-Geiger series. A recent publication of the Greenland administration [46] brings the picture up to 1939, after which little is available.

Northern Russia and Siberia are adequately covered by certain Russian publications. A simple treatment is given by Borisov [2], who has been considerably

1. F. Loewe was an observer with the Wegener Eismitte party.

influenced by the idea of "dynamic climatology" put forward by Bergeron. Borisov's study is illustrated by numerous distribution maps, some of them drawn from the U. S. S. R. *Great Soviet World Atlas* [61], and all of them emerging from official sources. An invaluable supplement is the exhaustive study of Tikhomirov [58] on the Arctic and sub-Arctic of Soviet Asia. Tikhomirov presents tabulated summaries of the chief observational data from most of the new Siberian stations set up since 1930 by the resurgent Soviet states. A general review of Russian climates from the point of view of the normal circulation has been given by Alisov [1].

To these general studies must be added a large volume of local or specialised work scattered through the literature, almost entirely from Scandinavian sources. When reading either the general studies or the more specialised papers, one should bear in mind the scattered nature of the observations upon which they are based. The peninsula of Labrador-Ungava may be taken as an instance. This great land mass has an area of over 500,000 square miles; from north to south the peninsula extends over 14° of latitude, the same span as that separating New York and Palm Beach, Florida, and in longitude it extends over 22°, a distance equivalent to that separating New York from Minneapolis. In area, the peninsula is thus equivalent to the whole of United States territory lying east of the Mississippi and south of latitude 40°. Yet in the whole of this area there were no inland climatological stations before 1915, and only one from 1915–37 (Mistassini Post). Today there are ten, corresponding to less than one per state in the southeastern United States as defined above.

It is plain, then, that the time for comprehensive climatological synthesis is far distant. Yet there are many specialised fields in which the climatologist can work, and in which much valuable material has already been published. The author has selected for discussion below those special fields of research that seem to him most profitable with our present limited resources. They are:

1. The ecological climatology of the Arctic and sub-Arctic, a field that tends to be neglected by the climatologist; and
2. The climatological relation of sea ice, a vast and disorderly field in which much research is now progressing.

In addition, some attention is given below to the highly significant question of Greenland's icecap climate, a topic much bedevilled by controversy.

Inevitably this selection omits many subjects considered important by others. Prominent among these will be the physical characteristics of arctic snow cover. Restrictions in space and the author's experience must excuse these omissions.

ECOLOGICAL CLIMATOLOGY

The relation of climate to natural vegetation is one of the most significant branches of modern climatology. Vegetation and soils are mirrors of the normal climate of a region; the concepts of climax vegetation and of the great world soil groups both depend on the idea that climate is the ultimate ecological control. There is fairly general agreement that the type of fully developed natural vegetation and the mature soil profile are faithful climatic indicators. Soil and vegetation maps divided into regions ought therefore to give us a useful method of defining rational climatic regions. Most systems of climatic classification, like those of Köppen [36] and Thornthwaite [56, 57], have been based upon this assumption.

The ecological approach to climatology is especially valuable in the Arctic and sub-Arctic, for these are by definition major world regions in which climate restricts growth and economic potential through the agency of excessive cold. The natural vegetation of the cold northern lands falls into two major formations:

1. The *tundra*, the truly arctic landscape with no tree growth, where the dominant plants are the sedges, rushes, mosses and lichens, with some small shrubs and bushes. Very little work has been done on the climatic relations of the tundra in North America, and the Russian tundra is still being mapped.
2. The *Boreal forest*, which is the term applied to the great forest formation that everywhere borders the tundra along the latter's southern flank. The tundra and Boreal forest formations are separated by a transition zone in which both occur, the so-called *forest-tundra*, or *forest-tundra ecotone*.[2] The arctic tree line is the extreme northern limit of tree growth and may in general be regarded as the northern edge of the forest-tundra. As so defined, the line is the absolute northern limit, not of the forest, but of patchy and often stunted tree growth, extensive islands or "peninsulas" of tundra associations lying to the south of it.

Tundra associations occur in places far south of the true arctic limit. These anomalous areas occur in two types of environment: (1) on high ground, where the effect of altitude lowers the summer temperatures to the point at which only tundra vegetation can be supported (this is the so-called *alpine tundra*, and one also speaks of the *alpine tree line*); and (2) on certain exposed coasts in high latitudes, as for example in Labrador, Iceland, and parts of Arctic Scandinavia (to these anomalies the term "maritime tundra" may be applied).

The Nature of Climatic Control. It has long been assumed that temperature and the length of the growing season are the climatic elements controlling the distribution of climax vegetation in the north. It is a demonstrable fact that interformational boundaries tend to follow the general trend of the mean isotherms across continental masses. This correlation of the great vegetation regions with thermal climate has been used by such climatologists as Köppen and Thornthwaite to define climatic boundaries, which will be discussed later. Thornthwaite explicitly states [56, p. 648] that in the cold climates restriction of growth by cold far

2. The term *ecotone* is used ecologically to indicate a zone of transition from one major plant community to another. See Weaver and Clements [64, p. 104].

outweighs the effect of scanty precipitation. In his view, precipitation is everywhere sufficient to meet the very limited demands of plant growth throughout the Boreal and Arctic regions.

This common assumption rests upon very little serious research or experiment. The Scandinavian school of ecologists appear to have been pioneers in this field. Several workers have studied the effects of climate upon radial and vertical growth of common coniferous species, such as the Scotch pine (*Pinus sylvestris* L.) (Hustich [29, 30]) and European spruce (*Picea abies* (L.) Karst.) (Eidem, quoted by Hustich [30]). Though the climatic relations of individual species are not strictly comparable with the relations of a vegetation formation, they offer a valuable indication of the mechanism of climatic control near the poleward limits of tree growth, and are hence of outstanding interest.

The most recent review of this work comes from Hustich [30], who has carried out important field surveys in Labrador and Arctic Scandinavia, and who has thus had an opportunity to study the problem in differing environments. He found that both vertical and radial growth in northern conifers near their northward limit were closely dependent on midsummer temperature. His results are of such interest that they deserve a comprehensive summary. In a stand of Scotch pine at Utsjoki (Lapland, lat. 69°30′N) he found the following correlations:

1. Radial growth (*i.e.*, increase in diameter due to addition of annual ring of woody tissue) was highly correlated with July mean temperature for the same year at nearby Inari (68°57′N, 26°49′E, 153 m (502 ft) above sea level).

2. Vertical growth (*i.e.*, increase in height) was correlated with July mean temperature *for the previous year*, though less clearly than in the case of radial growth.

He also showed that variations in radial growth had followed quite closely secular variation of July temperature since 1890. In another series near Sodankylä [29] he obtained a correlation coefficient of $+0.54 \pm 0.12$ between July temperature and radial growth. In this series the corresponding coefficient between July rainfall and radial growth was -0.24 ± 0.16. All other attempts to show correlation between rainfall and growth proved equally abortive. Accordingly, he concluded that ". . . in the northern part of the temperate zone the correlation between temperature and growth is stronger than the correlation between precipitation and growth." He ascribes this independence of growth on summer rainfall to the maintenance of soil moisture content by melting snow.

Erlandsson [12] obtained even more convincing evidence from Karesuanda in northern Sweden. He was able to correlate radial growth in Scotch pine with mean temperatures over the period 1879–1931. His results are presented in Table I.

This table emphasises the importance of July temperature, rather than that of the summer as a whole. Apparently the critical season is extremely short in these high latitudes. Like Hustich, Erlandsson concluded that precipitation was a negligible factor in determining growth.

Table I. Correlation Coefficients Between Climatic Factors and Radial Growth of Scotch Pine at Karesuanda, Sweden, 1879–1931
(*After Erlandsson* [12])

Climatic factor	Correlation coefficient
June, mean temperature	$+0.29 \pm 0.13$
July, mean temperature	$+0.70 \pm 0.07$
August, mean temperature	$+0.24 \pm 0.13$
Rainfall, June–September	-0.19 ± 0.13

These important results depend upon the well-established technique of tree-ring analysis; the width of the annual ring of xylem is carefully measured, variations from year to year in thickness being assumed to indicate variations in climatic stimulus to growth.

Hustich [30] points out that the northern conifers, though they grow in a region of great climatic hazard, have a remarkable freedom from "suppressed" or "missing" annual rings, such as are constantly referred to by the Arizona school of tree-ring workers. In lower latitudes, where temperatures are high enough to support growth for a large part of the year, summer drought becomes an all-important factor. Along the forest-steppe margin, occasional drought years may virtually suppress radial growth, and the tree trunk therefore contains only a stunted, incomplete, or vestigial "ring" as a record. This is true of all parts of the world in which occasional droughts occur, even, for example, in equable southern England [33]. Precipitation is a notoriously variable element; hence in any region in which the maintenance of soil moisture through a long growing season depends on summer rainfall, the annual-ring spectra of trees will contain marked variations and "suppressed" rings.

If precipitation is unimportant as a factor influencing growth in northern regions, the annual-ring spectra should show only small variations in annual growth, for July temperature, the apparent prime control, is nowhere very variable. Hustich's observation that suppressed rings are absent from northern Scandinavian conifers confirms this deduction. Hustich's finding is in turn confirmed by several investigations in North America. Thus Giddings, accustomed to work in the dry western part of the United States, apparently did not notice suppressed rings in the Alaskan [14] and Mackenzie regions [15]. Marr, in an important monograph on the forest-tundra ecotone of western Labrador-Ungava [38, pp. 142, 143], was even more specific. Referring to white spruce (*Picea glauca* (Moench) Voss), he writes ". . . annual growth has been so uniform at Gulf Hazard during the last 230 years, that only 41 growth-rings are sufficiently below average width to permit visual detection of the deviation."

There is thus ample support for the view that summer temperatures, and especially July temperatures, are the major factor determining radial growth (and probably vertical growth and volume increment too)

in the coniferous trees that dominate the Boreal forest, at least in its northern parts. This does not mean that precipitation is unnecessary to the growth, but that the combined snow melt of May and July and the precipitation of June, July, August, and September are everywhere enough to support the restricted growth these subarctic temperatures can sustain.

This view has been challenged by Sanderson [50], who found that drought is a marked restricting factor in the Boreal forest of western Canada. By using Thornthwaite's new system of climatic classification, she was able to show that widespread moisture deficiencies occur throughout the part of the Canadian Boreal forest lying west of Ontario; the entire region is dry sub-humid or semi-arid, according to her results. She maintained that the forest is stunted by this summer drought, or may even be replaced by scrub or prairie [48]. Soil acidity is much lower than one might expect for the typical Boreal forest podsols, *p*H determinations varying from 7.0 to 7.6 in northern Saskatchewan, and from 6.3 to 7.8 at Fort Simpson on the Mackenzie. These are values ranging from mildly acid to mildly alkaline, far different from the strongly acid reactions of soils from Quebec, Labrador, and the Finnish plateau.

It is probable that the truth lies somewhere between these two opposed views. Hustich, Erlandsson, and the other workers of the European school have been accustomed to the Boreal forest in its wetter phase—the Finnish and Lapp plateaus resembling Labrador and Northern Ontario in climate—and may have been misled by sheer lack of exposure to the drier types of microthermal climate. On the other hand Giddings' tree-ring work on the Mackenzie [15] confirms Hustich's view rather than Sanderson's: he makes no mention of stunted rings due to drought years. Sanderson is also open to attack on the grounds that Thornthwaite's scheme contains sweeping assumptions about available soil moisture that may not apply in permafrost areas, or in areas of glacially disturbed drainage. As Clark [7] says, her view that drought restricts growth in the Canadian northwest is ". . . a somewhat startling conclusion for the many who have spent summers there with almost perpetually wet feet! This apparent anomaly is explained as the result of poor drainage rather than of humid climate . . . ."

It appears quite probable that there are certain areas of the Boreal forest where summer drought is effective in the manner Sanderson suggests, not only in Canada but in Siberia: salty or alkaline marshes and prairies occur in the Yakutsk basin of the Lena Valley, just as they do in the Peace River Valley [53]. But it also seems likely that *once enough moisture is provided during the growing season*, as it is in most areas, further moisture has little or no influence on the forest: it cannot induce greater luxuriance, nor can variations in available moisture above this quantity have any effect on the annual-ring record. In all but a few areas the effective climatic control within the Boreal forest appears to be thermal; the regional subdivisions of the forest are hence correlated with isopleths of temperature or functions of temperature, and ignore the precipitation distribution.

Zonal Divisions of Arctic and Sub-Arctic Vegetation. With this background, we can now examine the climatic correlations of the major divisions of natural vegetation.

The Arctic Tree Line, the Northern Limit of Tree Growth. This line is often extremely sinuous, and is hard to map, but its course is now tolerably well known round the world. As it divides the truly Arctic tundras on the north from the Boreal forest formation, it is an extremely important boundary.

Early attempts to find the poleward limits of tree growth stressed July temperature. Supan [52] observed that the isotherm of 10C (50F) for the warmest month followed the arctic tree line closely. Later Köppen [36] adopted this isotherm as the division between tundra (Ekistothermal) and microthermal climates. The general correspondence of the two lines can be seen on world maps, though in detail they may diverge considerably. The 10C isotherm has enjoyed a prestige among biologists and geographers far exceeding its merit.

Vahl [62] differed from Köppen in assigning some significance to winter temperatures. Forests are observed to flourish in certain maritime regions right up to or even beyond the 10C isotherm. Thus Tierra del Fuego, at the southern tip of South America, is forested, though the warmest month is only 8–9.8C (46–50F). Here the mean daily temperatures remain above 0C throughout the winter, and it is evident that the mildness of the cooler season is the permissive factor. In continental interiors, on the other hand, the tree line fails to reach the 10C isotherm for the warmest month. This isotherm passes just north of Baker Lake in Keewatin, for example, some 200 mi north of the tree line. Presumably the much greater winter cold of the continents is the operative factor. In short, the colder the winter, the warmer the summer has to be to support tree growth.

Nordenskjöld [45] tried to express this idea in a rough-and-ready substitute for the Köppen-Supan line. He suggested that the tree line was most closely followed by the isotherm

$$W = 9 - 0.1k, \tag{1}$$

where W is the mean temperature of the warmest month and k is the mean temperature of the coldest month, both in centigrade degrees. In Fahrenheit, this identity becomes

$$W = 51.4 - 0.1k. \tag{2}$$

The "Nordenskjöld line," as it is often called, fits the arctic tree line better than any other purely climatic isopleth. Figure 1 shows its course.

The attempts so far discussed are hit-and-miss affairs with no rational basis. No adequate discussion of tree-line climates is known to the present author. Thornthwaite has proposed a climatic classification supposedly "rational" in that it depends upon a reasoned consideration of the soil moisture cycle and the water needs of

plants in growth [57]. He divides the world into thermal regions defined by reference to what he calls "potential evapotranspiration."[3] The latter is a function of the temperature, whose value for a month is given by

$$E = 1.6\ (10T/I)^a \text{ cm}, \qquad (3)$$

where T is the centigrade mean air temperature, I is a heat index dependent on the annual course of mean temperature and is equal to the sum of the twelve monthly values of $(T/5)^{1.514}$, and a is a constant dependent on I. The twelve monthly values of E are added (after they have been adjusted for variations in length of day and month) to give an annual value of potential evapotranspiration.

FIG. 1.—The extent of the Arctic. Solid line—isotherm of 10C (50F) for warmest month. Dashed line (with capital *N*'s)—the Nordenskjöld arctic boundary (see text). Alternating dots and dashes—southern limit of general permafrost. Dotted line—southern limit of permafrost in patches. The shaded area is tundra or icecap. Alpine tundra is not shown, except along the Urals (for reference) and again in northeastern Siberia, where it is included to indicate the isolated area of wooded country in the Anadyr Valley.

This value forms the basis of Thornthwaite's thermal zones. Along the tundra/microthermal boundary, E should be 28.5 cm (11.2 in.), according to Thornthwaite's divisions. In practice, this theoretical line

3. This function is defined as the amount of water that would evaporate from a plant-covered soil if there were no restriction in water supply in the root zone. It is considered independent of the floral composition of the vegetation.

lies well north of the tree line in North America, as well as in most parts of Siberia. The northernmost trees in Keewatin do not get far beyond the line $E = 33.0$ cm. In Labrador-Ungava they locally attain $E = 29.5$ cm, but again tend to halt at the 33-cm value.

Attempts to find climatic equivalents for the tree line assume that the latter is in fact "climax," that is, has reached the northernmost limit of feasible tree growth. In many areas this is not the case. The trees have not yet reoccupied all the climatically suitable lands from which they were driven by the Wisconsin glaciation and its attendant cold. Thus Marr has shown that active northward migration of the tree line is in progress on the east coast of Hudson Bay [38]. Griggs has similarly described the spread of the Pacific and Boreal forests into the Alaskan maritime tundra [17], notably on Kodiak Island and the Alaskan peninsula. It seems quite evident that the tree line has not yet attained its equilibrium position in many areas; furthermore the rapid fluctuation characteristic of recent subarctic climates makes it doubtful whether any such equilibrium can be struck at all. In the circumstances, it is unlikely that any rational climatic equivalent can be found for the tree line. Nordenskjöld's line, arbitrary though it is, comes closer to the present position than any "rational" attempt. A glance at Fig. 1 will convince the reader of this close fit.

The "maritime tundra" already briefly discussed presents many interesting problems. Along many coastal belts of the sub-Arctic, treeless tundras, usually with a high proportion of typically arctic species, descend far south of the normal poleward limit of trees. Excellent examples are the Aleutian-Alaskan peninsula region, the Labrador coast, parts of Iceland, coastal Fennoscandia, and certain localities on the coast of the Maritime Provinces of Siberia, facing the Okhotsk and Bering Seas. Similar coastal treeless formations occur still farther south, as in the moss-barrens of south-coastal Newfoundland, the coasts of Ireland and Scotland, and the Kuril Islands. These regions are different, however, in that floristically the vegetation has many nonarctic affinities; furthermore, human agencies may have destroyed pre-existing forests, except in Newfoundland.

The best-studied case is that of coastal Labrador [28, 66], where maritime tundra forms a narrow belt extending as far south as the Strait of Belle Isle. Though rarely more than a few miles wide, the strip grades inland into the Boreal forest through a well-marked tree line and a zone of stunted brushwood like the *krummholz* of the Vorarlberg and the thickets of Mount Washington. It was long assumed that this tundra was the result of the chilling of summer temperatures by drifting ice in the Labrador current. It has been shown, however, that the "thermal efficiency" of the coastal climate is adequate to support a poor forest or woodland cover [19, pp. 237, 238; 19*a*, p. 631]. The absence of trees is ascribed by Wenner to excessive transpiration (and hence physiological drought) induced by the strong coastal winds. Whether the explanation holds or not, it is certainly true that the forest finds it more difficult to invade the coastal sub-Arctic than the interior.

There are very few studies of the montane tree line in subarctic regions comparable with those on high, isolated summits in temperate and tropical latitudes.

The Forest-Tundra Ecotone. This zone is a belt of variable width in which treeless and woodland associations intermingle. In North America it has been delimited in Labrador by Hustich [31], its southern boundary being thermally adjusted to the isopleths of

$$E = 35\text{–}37 \text{ cm}$$

potential evapotranspiration [19*a*, p. 630]. Elsewhere its boundary has yet to be adequately mapped. In Siberia its extent is well known west of the Yenisei, where again its southern edge is at about $E = 35$ cm. It is clear that in these regions the ecotone is a belt whose position and width are thermally determined, thus conforming to the expectations raised by the work of Hustich, Erlandsson, and others discussed above.

The Boreal Forest Itself. This zone can be divided into two broad subdivisions that have been recognized both in North America and in the Soviet Union. On the northern flank there is a belt of open woodland, characterised by widely spaced spruce (*Picea* spp.), pine (*Pinus* spp.), or larch (*Larix* spp.) set in a rich lichen floor, the dominant genus of lichen being *Cladonia* ("reindeer" or "caribou moss"). On the southern flank lies a broad belt of true forest, with close stands of spruce, fir, pine, larch, and some hardwoods. The ground is in deep shadow and is usually carpeted with mosses and shade-loving herbs. The two belts are fairly distinct, meeting along a mappable boundary; they go by various names, but the term "Boreal woodland" and "main Boreal forest" will be used for them here.

The boundary between them is indicated for Canada in the official publication "Native Trees of Canada" [5], the Boreal woodland being indicated as "transitional" to the Arctic to the north. Hustich, however, distinguishes (1) a forest-tundra ecotone, already discussed, (2) a "taiga" belt, or Boreal woodland of the lichen-floored type, and (3) a southern Boreal forest, equivalent to the main Boreal forest defined above [31]. Hustich's divisions correspond structurally with those widely recognised for the Boreal forest of Finland and Soviet Russia, and are hence of great interest [53].

In Labrador-Ungava, it is at once clear that the woodland/forest boundary coincides very closely with the isopleth of potential evapotranspiration $E = 43$ cm [19*a*, p. 630], which is that selected by Thornthwaite to divide the microthermal climates into warmer and cooler halves. Though a similar investigation of the Russian climate has not been carried out, Arkhangelsk, about 100 mi south of the boundary, has $E = 44.5$ cm, and Ust-Sylma, on the Pechora River some 50 mi within the Boreal woodland has $E = 40.3$ cm. Furthermore, it is clear that all the belts of the Boreal forest west of the Yenisei extend zonally along the isopleths of potential evapotranspiration, as they do in Labrador-Ungava [31]. Obviously the woodland/close-forest boundary is

a thermally controlled divide in the same position climatically in both Labrador and Russia.

The Southern Limit of the Boreal Forest. This zone is less clearly defined climatically and is variable in ecological character. In Alberta, Saskatchewan, and Manitoba, the forest passes south into prairie (or steppe) lands through the so-called parklands, or *forest-steppe ecotone.* A similar forest-steppe boundary is observed in Siberia from the Urals to the Yenisei. In eastern North America the Boreal forest passes southwards into a mixed hardwood-softwood formation, the *Great Lakes–St. Lawrence forest* of Halliday [18], or the *Lake forest* of Weaver and Clements [64, pp. 496–500]. A similar passage is observed in European Russia along a line extending from Estonia through Yaroslavl to the Urals near Sverdlovsk. Both in America and Russia, the mixed forest is distinct floristically from the Boreal forest to the north and the deciduous forest to the south, though it has some elements in common with both.

In eastern North America, certain species typical of the Great Lakes-St. Lawrence forest penetrate the true Boreal forest, reaching an east-west line with potential evapotranspiration about 48 cm. South of the isopleth $E = 51$ cm, the dominants are everywhere the white and red pine (*Pinus Strobus* L., *Pinus resinosa* Ait.), yellow birch (*Betula lutea* Michx. f.), and various maples; the spruces (*Picea* spp.), balsam fir (*Abies balsamea* (L.) Mill.), and Jack pine (*Pinus Banksiana* Lamb.), typical of the Boreal forest, are subsidiary elements. Here again we are faced with a thermally-correlated boundary; the southern limit of the Boreal forest follows an isopleth of Thornthwaite's thermal efficiency function.

Once more the same value appears, on a casual inspection, to hold for the Russian forests. Moscow, 150 mi south of the Boreal-mixed forest line at Yaroslavl, has a value of $E = 54.2$ cm, which in North America would put it well into the mixed belt. On the other hand, Perm (Molotov), 75 mi north of the line, has $E = 50.8$ cm. These values are obviously very close to those applicable in eastern North America. What is even more surprising is that the inner (forest) edge of the forest steppe *both in the Canadian Prairies and in western Siberia* closely corresponds to the same value of potential evapotranspiration.

Gathering all these threads of evidence, we can draw these conclusions, all of which remain tentative because of the fragmentary evidence employed:

1. Through most parts of the Boreal region of the earth, the temperature of midsummer is the main factor determining variations in growth of the common tree species.

2. Variations in summer precipitation have little or no effect on growth, except near the forest-steppe margin.

3. The Boreal-forest formation shows a broad division into latitudinal zones: a forest-tundra ecotone, next a woodland belt, then a close-forest region. The arctic tree line and all the interzonal boundaries are highly correlated with Thornthwaite's thermal efficiency function, the potential evapotranspiration (E). The values of E applying along these lines are much the same in the Soviet Union west of the Yenisei River.

Obviously these conclusions need to be checked and rechecked. There are probable sources of error at every stage of the analysis. For the ecologist, there is the task of more adequately mapping the zonal boundaries of the Boreal forest. For the climatologist, there is the responsibility of collecting and analysing more and more climatic data. Thornthwaite's stimulating climatic classification, used above to illustrate the dependence of the Boreal forest on thermal efficiency, needs to be scrupulously examined as to its applicability to the cold climates. Sanderson's work in the Mackenzie Valley [51] is an important contribution in this particular. Above all, the type of analysis here attempted needs to be applied to the mountainous lands of Alaska and the Yukon, and of Siberia east of the Yenisei. Here the montane effect must vastly complicate the correlations quite easily established in Labrador and the western Soviet Union.

Permafrost Distribution. A widespread condition in the cold lands is that of permanently frozen soil, christened "permafrost" by Muller [43]. This condition is achieved when the annual wave of summer heating fails to descend to the base of the layer of frozen soil. The mechanical properties of permafrost are the business of the engineer, but its distribution is of interest to climatologists. Figure 1 shows its approximate extent in the Northern Hemisphere. The data used in its preparation come from studies by Muller [43] and Jenness [32].

Few attempts have been made to correlate permafrost with climate. Sumgin (quoted by Muller [43, p. 4]) states that Russian estimates give 0C (32F) annual mean temperature as the outer limit of patchy permafrost, and that geographically continuous permafrost occurs north of the annual mean isotherms of −3C to −6C (26–21F). Thomson (quoted by Jenness [32]) suggests that the −5C (23F) isotherm comes closest in Canada. All these are approximate estimates, and it is quite obvious that such quantities as depth and duration of snow cover must also affect the distribution. In detail, the occurrence of permafrost is affected by soil drainage, slope, exposure, the presence of large bodies of water, and the geological history of the region. The climatological correlations of permafrost urgently await study.

THE DISTRIBUTION OF SEA ICE

It was earlier suggested (p. 952) that the term "Arctic" might well refer at sea to those areas perennially or annually affected by sea ice, either landfast or pack. The presence or absence of sea ice is of prime importance to both the meteorologist and the climatologist, who are confronted with questions of two sorts: (1) the effect of ice formation on air-mass structure, and hence upon climate; and (2) the climatological requirements for the formation of ice, and for its drift into other waters. The literature on these questions is abundant but scattered; moreover much of it comes from

Russian sources and is hence inaccessible. The review that follows is as inadequate as the subject is important. All that can be attempted is a review of the fields in which research is at present active.

The physics of sea-ice formation, of its drift, and of the deformation induced in it by atmospheric disturbance lies beyond the scope of this article. A general summary of present knowledge was recently prepared by Gordon and Woodsworth [16], and a review of Russian work in the field has been published by Zubov [69]. In North America most research in this field is contained in various publications of the U. S. Hydrographic Office [60].

Much time has been spent by research climatologists on the influence of arctic sea ice on the climate of the middle latitudes. Particular attention has been paid to the Labrador and East Greenland Currents, which are the chief routes by which heavy arctic pack enters lower latitudes [21]. This is a question of vital importance in the study of climatic change, but has less direct reference to the arctic climates themselves. In Greenland and along the Labrador Sea, however, the relationship is direct, and will be discussed later.

Hudson Bay: A Concrete Example. Until 1945, it was generally assumed that Hudson Bay remained unfrozen throughout the winter, despite the grim cold of nearby land areas. This assumption had the support of Canadian, British, and American official papers. Many authorities entered into considerable detail in describing the narrow belt of fast ice along the shores, but insisted that the central area rarely or never froze from shore to shore. The coastal fast ice was variously stated as being 5 to 60 mi wide [59].

This early view was first challenged when military aircraft began flying across the northwestern flanks of the Bay during World War II [49]. In March 1948, Lamont [37] flew directly across the central area of the Bay, photographing an unbroken pack covering the whole area. Since then regular reconnaissance flights have been made across all parts of the Bay by the Royal Canadian Air Force, with scientific observers on all flights and with complete photographic records. These flights have established that the Bay freezes annually in the late fall and is completely frozen from January until June, except for (1) a coastal lead, separating fast ice from the central pack, and (2) small, temporary leads or fissures set up in the central parts by storms, tides, and other sources of stress. These new facts have been summarised by Montgomery [20, Part II]. Associated with the reconnaissance program, there has been some active research, to be described below.

Burbidge [3] set out to examine the changes wrought in continental polar air masses in their travel across the Bay. By plotting trajectories at the main aerological levels, he was able to trace changes in the air before it arrived at the east coast aerological station of Port Harrison (58°27′N, 78°08′W). His results can be summarised as follows:

1. In the fall, before the freeze-up, great heating and dampening occur in the airstreams as they cross the warm open water. This effect rises to a maximum in late November and early December, when the surface temperature rises an average of over 20F (11C) in the crossing from west to east. Thereafter, the consolidation of sea ice cuts off the source of heat, and heating is negligible from mid-January on; apparently the snow-covered ice surface entirely insulates the water below. Curve (*f*) in Fig. 2 shows the amount of this

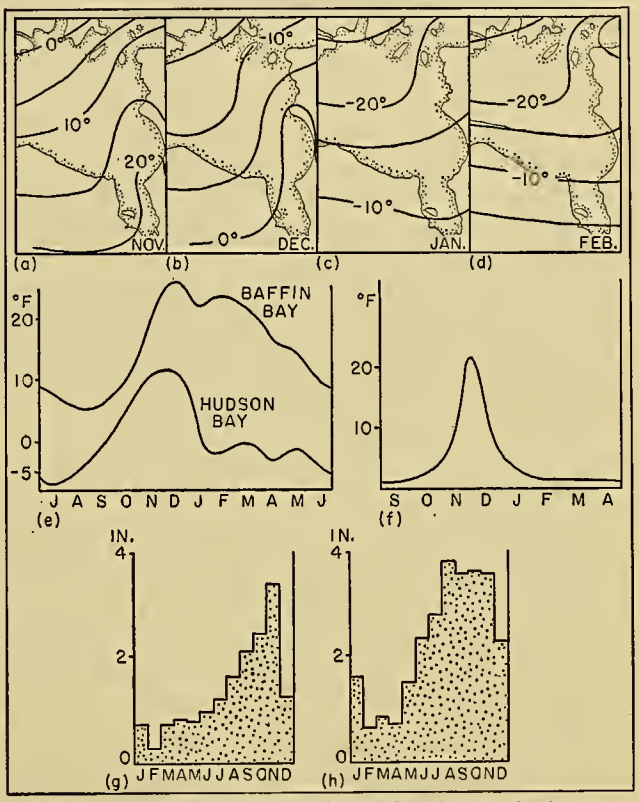


Fig. 2.—A summary of the evidence bearing on the freeze-up of Hudson Bay. Maps (*a–d*)—mean air temperature, 1930–48, November–February. Curve (*e*)—temperature differences (upper curve) between west and east coast of Baffin Bay at latitude 70°N, showing year-round mildness of Greenland coast; (lower curve) between west and east coast of Hudson Bay, showing greater warmth of east coast *during fall months only*. Curve (*f*)—Burbidge's modification curve for *cP* airstreams (see text). Curves (*g*) and (*h*)—mean monthly precipitation for Port Harrison (*g*) and Great Whale River (*h*) on east coast of Hudson Bay.

warming effect month by month. The rapid decline after December is very striking.

2. This heating leads to the development of an unstable layer at low levels, with lapse rates close to the dry adiabatic. By late November, the peak time, the average depth of the unstable layer is 7000 ft, and may attain 10,000 ft. Dense cumuliform cloud and frequent snow flurries occur within the layer, and total snowfalls along the Quebec coast are heavy (over 30 in. in November). With the freeze-up, the unstable layer disappears or becomes very shallow (less than 1000 ft) as do the cloud and snow. In Fig. 2, diagrams (*g*) and (*h*) represent mean monthly precipitation at Port Harrison and Great Whale River, respectively, on the east coast of the Bay. In both cases heavy autumnal snow gives way to light falls in the new year.

It is noteworthy that Burbidge obtained these results before reconnaissance had proved them beyond all doubt, thus demonstrating the value of careful meteorological analysis. His findings are quite what one would expect, but it was not until 1949 that they were proved or even suspected by many.

A similar result was obtained by Hare and Montgomery in a climatological analysis [20]. They showed that conspicuous gulfs of warmth characterise the winter mean air temperature maps in areas where open water persists. Three such gulfs were noted in the North American Arctic:

1. Along the east shore of Baffin Bay, where considerable open water persists in the warm West Greenland Current as far north as latitude 70°N.

2. Along Hudson Strait, where high tidal range and exposure to storms keep the pack ice well broken throughout winter.

3. Over eastern Hudson Bay. This gulf was visible, however, only until December, after which it vanished. The authors therefore assumed with Burbidge that the Bay was entirely frozen over by the new year. Diagrams (*a*)–(*d*) in Fig. 2 show mean temperature over the Bay from November to February. The conspicuous gulf of warmth on the November (*a*) and December (*b*) maps vanishes in January.

The reconnaissance flights in the winter of 1949–50 showed that in that year the freeze-up began along the northern and western flanks of the Bay; by November 22, 1949, continuous though thin ice extended as far south as Portland Promontory on the east coast and beyond Churchill on the west. The last area to freeze (in late December) was south and east of the Belcher Islands.

Other Seas in the American Arctic. In several other areas problems comparable with those of Hudson Bay are awaiting solution by similar methods. By direct reconnaissance, by meteorological inference, or by the observation of climatological anomalies, it will shortly be possible to achieve greater precision in our ideas about their winter ice cover.

Outstanding among these problem areas is Baffin Bay, and its narrow link with the Labrador Sea, Davis Strait. Its winter ice cover comprises three components: (1) fast ice forming along the shores and in the fjords of Baffin Land and of Greenland north of Holsteinsborg; (2) the central pack ice, composed of heavy arctic ice from Lancaster, Jones, and Smith Sounds, together with locally formed floes; the whole mass drifts southwards on the Canadian Current, ultimately discharging into the Labrador Current, which in turn may carry it to the Grand Banks late in the season; and (3) the Storis, a drift of heavy arctic ice from the East Greenland Current; rounding Cape Farewell in January, it advances northwards in the West Greenland Current, reaching farthest north (about 63°–64°N) between April and June.

It has already been mentioned that open water tends to persist through most of the winter between the central pack and the fast ice of Greenland (see p. 959). This break extends at least as far north as 70°N at most times. Furthermore, the central pack rarely approaches the Greenland coast south of Holsteinsborg, and probably never makes lasting contact with the Storis. The lane of open water up West Greenland, then, must extend without a break from the Labrador Sea. Small wonder that the climate of the West Greenland coast is so remarkably mild in winter. Disko Island, for example, is some 25F (14C) warmer than the Baffin Land coast. See curve (*e*) in Fig. 2

However confidently these claims are made, one should remember the extent to which they depend on inference. The open water off the Greenland coast has never been precisely mapped. The warming effect of subsidence from the icecap to the east, seemingly required by the mean winter pressure distribution, may also contribute to the warmth of the West Greenland coast. The tendency for warm maritime polar air to travel northwards along the Bay may be yet another factor. A general study of the circulation over Baffin Bay, with an associated survey of ice conditions by aerial reconnaissance, has much to offer the arctic climatologist.

An area of special interest lies at the northwest corner of the Bay near the entrance to Smith, Jones, and Lancaster Sounds. This area is the celebrated "North Water," a region free of ice and with quite warm surface waters during the navigation season. It is believed to remain partially unfrozen during the winter [34], and it is well known that Lancaster Sound never freezes solidly from shore to shore. Stations on Devon Island (Dundas Harbour), Baffin (Arctic Bay), and Bellot Strait (Fort Ross) plainly show the warming influence of open water throughout the winter [20, Part I]. The North Water must be mapped by aerial reconnaissance to determine both its position and its extent far more adequately.

The Scope for Research. Research into the climatological significance of sea ice has much to offer, being still in its infancy as far as North America is concerned. The obvious fields for investigation can be summarised as follows:

1. The modification of air masses over ice-covered seas is a fertile field for the dynamic meteorologist. Since Sverdrup's careful work on the *Maud* [54], very little has been done on this important subject. Burbidge's analysis over Hudson Bay suggests that a thorough understanding of the mechanics of modification may ultimately make it possible to predict both the extent and surface characteristics of offshore ice by shore-based aerological stations. Quantitative studies like those for *cP* air overland by Wexler [67] and for the transformation of *cP* to *mP* by Burke [4] and Klein [35] are urgently needed for the flow of intensely cold arctic air masses over unfrozen sea surfaces.

2. The climatological background of the freeze-up of the open sea areas also requires investigation, obviously in conjunction with the foregoing problem. This is a task requiring the co-operation of oceanographers. Hudson Bay, as a virtually enclosed body of water,

ringed by aerological stations and passing annually through the cycle from completely frozen to completely ice-free, seems an obvious laboratory for such a study.

3. The characteristics of coastal ice, especially in harbours, is another topic worthy of study by the climatologist. At present the relation between climate and freeze-up, breakup, and thickness of coastal ice is only qualitatively understood. In 1943 the United States Air Weather Service began serious observation of both fresh water and sea ice, as well as of snow cover, at all its Canadian, Alaskan, and Greenland stations. It is to be hoped that these records will ultimately become available to research workers.

THE PROBLEM OF GREENLAND

Most problems of arctic climatology are hemispheric: they are widespread in character, and are as much problems in Siberia, for example, as they are in Alaska. Greenland presents us with an entirely distinct set of questions, to which highly controversial answers are sometimes given. Since many of these problems are critically important to students of Pleistocene glaciation and to glaciologists as well as to climatologists, they demand a brief review here.

The central icecap of Greenland, occupying as it does the great bulk of the land mass, has remained throughout the past hundred years a virtual blank on the weather map. Our knowledge of its climate rests upon the reports of the handful of explorers who have crossed it, and on the two or three brief periods of observation by icecap meteorological stations. In place of real knowledge, we have had a generation or more of heated controversy concerning the theory of the so-called "glacial anticyclone." This controversy has recently [26] been likened to a debate between explorers (who have seen for themselves) and absentee meteorologists (who persist in ignoring the evidence provided by the explorers). This is a grave oversimplification, but it does contain the seeds of the truth; the controversy wages over different interpretations of the evidence, that of the explorers—which is essentially inexpert—and that of the meteorologists, many of whom have had practical experience in Greenland.

Nearly all the explorers who have crossed or penetrated the icecap refer [47] to a remarkable wind and weather régime. They speak of strong and monotonously regular downslope winds blowing out from the central core, and pouring out through the coastal fjords to the surrounding seas. Peary [47] described the flow as being analogous to the flow of water down a slope. In a sense these stories represent very large scale parallels of the *Gletscherwinde* of the Alps.

In 1910 the well-known geologist, W. H. Hobbs, presented in its earliest form the theory of the glacial anticyclone [22], by which he explained not only the Greenland climate but also the general characteristics of Pleistocene icecap climates. With added experience he published a greatly expanded version of this theory in his book *The Glacial Anticyclones* [23] and his views have remained substantially unchanged ever since [25]. Though he has been hotly attacked, Hobbs has maintained his position with great tenacity; the "glacial anticyclone" has become an established part of the legend of Pleistocene geology, and has coloured the thinking of arctic specialists for a generation.

We may summarise the main features of Hobbs's theory as follows:

1. An ice surface is a good reflector and a good radiator; hence the air resting on the icecap should be subject to intense radiative cooling which cannot be counteracted by absorption of insolation.

2. The cooling must lead first to an accumulation of very cold air in the central regions of the cap, and hence to rising pressure. Finally, the accumulated mass of cold air must break out to the coast in a spectacular downslope surge or "stroph" of the glacial anticyclone. To replace the outflowing air, general subsidence begins, and at very high levels there are radially inflowing winds. The subsiding air, however, is warmed thermodynamically, and eventually destroys the cold of the central regions. In a few days the "stroph" is finished, and the icecap settles down to a fresh period of refrigeration, after which the cycle is repeated.

3. In effect, this view postulates that a fixed, permanent anticyclone lies over the icecap; in it the normal clockwise circulation of the anticyclone is replaced by a gigantic downhill, gravity-impelled divergence of the chilled surface air. The presence of such an anticyclone over the cap must plainly inhibit the passage of cyclones, and it is a fundamental tenet of Hobbs's theory that cyclones do not penetrate more than a few miles inland. He has published several papers [27] designed to prove this.

4. The alimentation of the icecap, however, presents a problem, for if cyclones do not cross it, from where does the snow come to make good the large annual loss by ablation? In Hobbs's view, hoarfrost, not snow, feeds the Greenland ice, as it did the Pleistocene sheets. The radiative cooling of the subsiding air at the ice surface causes the condensation of huge quantities of hoarfrost, enough to make good all losses by ablation.

5. Hobbs's final point—and in some ways the most pregnant—is that the "strophs" of the glacial anticyclone set off the main frontal cyclones of the Atlantic; the icecap acts, in fact, as the "North Pole of the winds," to use his own picturesque phrase. At all seasons the air over the icecap is the coldest in the Arctic, and hence must replace the polar cap itself as the source of true arctic outbreaks.

All aspects of this remarkable theory have been hotly debated for more than a generation. It is not too much to claim that Hobbs's work was directly responsible for the bulk of the exploration of the Greenland interior carried out after World War I. The University of Michigan sent several expeditions to the west coast between 1926 and 1933 under the leadership of Hobbs himself and L. Gould [40]. These expeditions set up stations both on the coast and on the icecap, carrying out an elaborate series of surface and upper-air observations. In 1930–31 Alfred Wegener organised a re-

markable excursion which maintained stations on both flanks as well as in mid-ice (Eismitte) for a calendar year. The Wegener results provide by far the most important evidences yet at hand concerning the icecap climate. In the same year an official British party established an icecap station on the east slope, rather to the south of the Wegener station. Thereafter there was a period of inactivity until the United States established stations during 1944. In 1949 a well-equipped French party led by P. E. Victor reoccupied the Eismitte site, resuming the Wegener observations on an enlarged scale. The station is still in operation (June, 1950) and efforts are being made to keep it open for another year.

As a result of these studies the consensus of meteorologists is that the glacial anticyclone is in large measure based on a confusion of fact with deduction. To summarise the present position:

1. The radial wind system, the basis for Hobbs's views, is more complex than early accounts allowed. Stations established on the slopes [40, 41] reveal that the surface wind most commonly flows downslope, but by no means invariably so. At the Eismitte station, where, according to Hobbs, the wind should have been light, variable, or easterly, the winds were actually strong [65]. All stations showed periods of upslope flow, and considerable variability was general. The University of Michigan station on the west slope indicated that the outblowing winds might extend to 4 km (13,000 ft) or above [40, Part I].

The general view today is that the radial wind system is a katabatic circulation; the cooled air flows down the slope under gravity, eventually reaching sea level along the numerous fjords with which the coast abounds. General subsidence in the central regions of Greenland must occur to feed the katabatic flow. In this sense the term "anticyclone" is faintly justified, since anticyclones are regions of subsiding air. In every other sense, however, the term is a complete misnomer.

Several writers have expressed the opinion that the katabatic winds blow only during periods of feeble general circulation over the icecap. Dorsey [11], who operated the U. S. Icecap station during 1944, states that the katabatic winds are absent if there is an appreciable upslope component in the regional circulation. Much the same conclusion is drawn by Mirrlees [41]. Both these writers produce evidence that exceptionally strong flow down the east slope occurs in the rear quadrants of cyclones in the vicinity of Denmark Strait or Iceland; in other words, excessively strong "strophs" of icecap air require the joint action of the katabatic flow and a parallel regional flow.

2. The alimentation of the cap by hoarfrost has been similarly disputed. Matthes [39], in a closely argued critique of Hobbs's views, points out that the air subsiding into the central regions of the Greenland "anticyclone" can contain only negligible amounts of moisture; furthermore, it must have been rendered relatively drier by the adiabatic warming postulated by Hobbs.

The condensation of rime (liquid cloud or cloud droplets frozen onto solid surfaces) is a well-known phenomenon in moist air masses. Mount Washington Observatory affords numerous striking and photogenic examples every winter. It is very possible that rime accretion is appreciable on the lower slopes of the Greenland cap. Hobbs, however, has specifically ruled that moist maritime air does not penetrate more than a few miles inland, and rime cannot possibly form from the subsiding air of the icecap.

On the other hand, there is abundant evidence that there is frequent heavy snow on the icecap. Georgi [13, pp. 52–87], observer at Eismitte, was almost overwhelmed by the heavy snow of August 1930, when he established the station; Courtauld, hero of the celebrated six-month vigil on the icecap of the British Arctic Air Route Expedition, was actually buried, and had to be rescued through the ventilating shaft of his quarters [41]. Moreover the periods of snow fell during spells of upslope winds, and were preceded by cloud sequences similar to those observed at lower levels during frontal passages. It is impossible to believe that experienced observers like Georgi and Courtauld could have confused snow and hoarfrost.

Most of these observers concluded that cyclones could and did cross the icecap, and that the snow was normal cyclonic snow. Dorsey [11], however, felt that it was unrealistic to talk of sea-level pressure systems crossing a 10,000-ft barrier; he suggested, with convincing evidence, that it is the higher-level perturbation that overlies a surface cyclone that crosses the icecap. This is the mechanism, for example, by which Pacific cyclones appear to enter central North America across the Western Cordillera. Dorsey argued that it was the height of the barrier that blocked the free movement of cyclones, and not the presence of a fixed anticyclone. A similar blocking action is exerted by many mountain systems such as the Alps, the Pyrenees, and the Caucasus.

An inspection of the remarkable observational diagram drawn up by Georgi [24] makes it impossible to doubt the force of these objections. The record shows the rapid fluctuations of pressure, temperature, and wind velocity typical of a disturbed cyclonic climate. To quote Georgi [13, p. 19]: "... the wind régime over the icecap is much more complicated than the simple model of the glacial anticyclone allows." To a reader of Georgi's diary this seems a conservative comment.

3. Dorsey [11] has also pointed out the basic fallacy in Hobbs's argument that the Greenland cap is the source of the cold waves which touch off the Northern Hemisphere's chief cyclonic storms. It is true, he admits, that the icecap air is the coldest in the hemisphere on a year-round basis. Potentially, however, it is much warmer than the air over, for example, Arctic Canada, and it loses its coldness when it descends to sea level. Actually, icecap air enters the circulation of the Atlantic very much warmer than the typical continental polar air masses of Canadian provenance.

In sum, then, the idealised model of a glacial anticyclone appears quite inadequate to account for the observed facts; the theory has elements of the truth,

but in general goes too far in deducing mechanisms that do not exist. Only a refusal to examine evidence has made it possible for Hobbs to maintain his position unchanged for so many years.

A mere rejection of Hobbs's views does not remove the need for further investigation of the icecap climate. The cap is the only existing continental glacier where complete scientific exploration seems within the bounds of possibility with available means of transportation. As such, it offers us enormous rewards. Nowhere else on earth can the climatology of glaciation be so effectively studied.

So far, the attention of climatologists has chiefly been upon establishing the character of the circulation over the cap. The work of assessing the distribution of fresh snowfall and its redistribution to the margins by the strongly divergent winds has hardly begun. It is imperative that the ablation cycle of the cap be studied, not merely along the route to Eismitte, but also round the little known northern and northeastern flanks. Though such work is normally reckoned the proper field of the glaciologist, the climatological significance of this problem is obvious.

CONCLUSION

The review of arctic climatological problems presented here is by itself an inadequate document. Within this Compendium there are several other papers that also discuss problems of climate that affect high latitudes, and these should be read alongside the present account. The vast, intricate, and controversial subject of climatic change, for example, has been left in other hands. Nowhere on earth have secular changes in climate been more conspicuous than in the Arctic, especially along its Atlantic flank, and the reader should seek out the facts of these changes in this volume and elsewhere.

It seems likely that the next ten years will see a great lifting of the obscurity now extending across the Arctic. The author earnestly hopes that this chapter will become out-of-date more rapidly than any other in the Compendium.

REFERENCES

1. Alisov, B. P., *Klimaticheskiie Oblasti i Raŭony S.S.S.R.* Moskva, Geografgiz, 1947.
2. Borisov, A. A., *Klimaty S.S.S.R.* Moskva, Uchpedgiz, 1948.
3. Burbidge, F. E., *The Modification of Continental Polar Air over Hudson Bay and Eastern Canada.* M. Sc. Thesis, McGill University, Montreal, 1949.
4. Burke, C. J., "Transformation of Polar Continental Air to Polar Maritime Air." *J. Meteor.*, 2:94–112 (1945).
5. Canada, Dept. of Mines and Resources, "Native Trees of Canada." *Dom. Forest Service Bull. 61*, 4th ed. (1949).
6. Canada, Dept. of Transport, *Meteorology of the Canadian Arctic.* Toronto, Meteor. Div., 1944.
7. Clark, A. H., "Contributions to Geographical Knowledge of Canada since 1945." *Geogr. Rev.*, 40:285–308 (1950). (See p. 294)
8. Connor, A. J., "Climates of Canada," Pt. 4 of Ward, R. de C., Brooks, C. F., and Connor, A. J., "The Climates of North America," *Handbuch der Klimatologie*, Bd. II, Teil J., W. Köppen und R. Geiger, Hsgbr. Berlin, Gebr. Borntträger, 1936.
9. Currie, B. W., Unpublished Mss., Dept. of Physics, University of Saskatchewan, Saskatoon, Sask., 1949.
10. Döll, L., "Klima und Wetter an der Küste von Labrador." *Aus. d. Arch. dtsch. Seew.*, 57 (2):1–21 (1937).
11. Dorsey, H. G., Jr., "Some Meteorological Aspects of the Greenland Ice Cap." *J. Meteor.*, 2:135–142 (1945).
12. Erlandsson, S., *Dendrochronological Studies*, Data 23. Uppsala, 1936.
13. Georgi, J., *Im Eis vergraben.* Münich, Müller, 1933.
14. Giddings, J. L., Jr., "Dendrochronology in Northern Alaska." *Bull. Univ. Ariz.*, Vol. 12, No. 4 (1941).
15. —— "Mackenzie River Delta Chronology." *Tree-Ring Bull.*, 13:26–29 (1947).
16. Gordon, A. R., and Woodsworth, W. C., *Some Interrelationships of Snow and Ice Conditions and Weather in the Arctic.* Paper presented at the 30th Annual Meeting of the Amer. Meteor. Soc., St. Louis, Jan. 5, 1950.
17. Griggs, R. F., "The Edge of the Forest in Alaska and the Reasons for Its Position." *Ecology*, 15:80–96 (1934).
18. Halliday, W. E. D., "A Forest Classification for Canada." *Dom. Forest Service Bull. 89* (1937).
19. Hare, F. K., *The Climate of the Eastern Canadian Arctic and Sub-Arctic, and Its Influence on Accessibility.* Ph.D. Thesis, University of Montreal, 1950.

19*a*. ——"Climate and Zonal Divisions of the Boreal Forest Formation in Eastern Canada." *Geogr. Rev.*, 40:615–635 (1950).

20. —— and Montgomery, M. R., "Ice, Open Water and Winter Climate in the Eastern Arctic of North America." Part I (F. K. Hare), "Distribution of Winter Temperature over the Eastern Arctic and Sub-Arctic." *Arctic*, 2:79–89 (1949); Part II (M. R. Montgomery), "The Pattern of Winter Ice." *Arctic*, 2:149–164 (1949).
21. Hepworth, M. W. C., "The Effect of the Labrador Current upon the Surface Temperature of the North Atlantic, and the Latter upon Air Temperature over the British Isles." *Geophys. Mem.*, 1:1–10 (1912) and 1:211–220 (1914).
22. Hobbs, W. H., "Characteristics of the Inland-Ice of the Arctic Regions." *Proc. Amer. phil. Soc.*, 49:57–129 (1910).
23. —— *The Glacial Anticyclones.* New York, Macmillan, 1926.
24. —— "Rhythm in the Greenland Glacial Anticyclone." *Trans. Amer. geophys. Un.*, 25:491–494 (1944). (In this article the celebrated Georgi diagram is most accessible to North American readers.)
25. —— "The Greenland Glacial Anticyclone." *J. Meteor.*, 2:143–153 (1945).
26. —— "The Climate of the Arctic as Viewed by the Explorer and the Meteorologist." *Science*, 108:193–201 (1948).
27. —— and Belknap, R. L., "Errors in the Synoptic Weather Charts Which Cover the Greenland Region." *Trans. Amer. geophys. Un.*, 25:482–490 (1944).
28. Hustich, I., "Notes on the Coniferous Forest and Tree Limit on the East Coast of Newfoundland-Labrador." *Acta geogr. Helsingf.*, 7:1–64 (1939).
29. —— "The Scotch Pine in Northernmost Finland and Its Dependence on the Climate in the Last Decades." *Acta bot. fenn.*, 42, 75 pp. (1948).
30. —— "On the Correlation between Growth and the Recent Climatic Fluctuation" in *Glaciers and Climate*, C. M. Mannerfelt, ed., pp. 90–105. Stockholm, Generalstabens Litografiska Anstalt, 1949. (Also published as *Geogr. Ann., Stockh.*, Nos. 1–2 (1949).)

31. —— "On the Forest Geography of the Labrador Peninsula." *Acta geogr. Helsingf.*, 10:1–63 (1949).
32. Jenness, J. L., "Permafrost in Canada." *Arctic*, 2:13–27 (1949).
33. Jones, E. W., "Comments on a Paper by I. Hustich." *Nature*, 160:478 (1947).
34. Kiilerich, A., "Nordvandet. Forsøg paa en Forklaring af det isfri Havomraade i Smith Sund." *Geogr. Tidsskr.*, 36:53–61 (1933).
35. Klein, W. H., "Modification of Polar Air over Water." *J. Meteor.*, 3:100–101 (1946). (See also Reps. 884 and 971, U.S.A.A.F. Weather Div., 1945)
36. Köppen, W., "Versuch einer Klassification der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt." *Geogr. Z.*, 6:593–611, 657–679 (1900). (See also "Das geographische System der Klimate," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. I, Teil C. Berlin, Gebr. Borntträger, 1936.)
37. Lamont, A. H., "Ice Conditions over Hudson Bay and Related Weather Phenomena." *Bull. Amer. meteor. Soc.*, 30:288–289 (1949).
38. Marr, J. W., "Ecology of the Forest-Tundra Ecotone on the East Coast of Hudson Bay." *Ecol. Monogr.*, 18:117–144 (1948).
39. Matthes, F. E., "The Glacial Anticyclone Theory Examined in the Light of Recent Meteorological Data from Greenland, I." *Trans. Amer. geophys. Un.*, 27:324–341 (1946).
40. Michigan, University of, *Reports of Greenland Expeditions, 1926–1933*. Part I, *Aerology*, S. P. Fergusson, ed.; Part II, *Meteorology, Physiography and Botany*, W. H. Hobbs, ed. Ann Arbor, University of Michigan Press, 1941.
41. Mirrlees, S. T. A., "Meteorological Results of the British Arctic Air Route Expedition, 1930–31." *Geophys. Mem.*, Vol. 7, No. 61 (1934).
42. Montgomery, M. R., *The Climate of Labrador and Its Effect on Human Settlement*. M. A. Thesis, McGill University, Montreal, 1949.
43. Muller, S. W., *Permafrost*. Ann Arbor, Edwards, 1947.
44. Nansen, F., ed., *The Norwegian North Polar Expedition 1893–1896: Scientific Results*, 6 Vols. London, Longmans, 1906.
45. Nordenskjöld, O., and Mecking, L., *The Geography of the Polar Regions*. New York, American Geographical Society Special Publ. No. 8, 1928. (See p. 73 and also Chap. on Climate)
46. Oldendow, K., "Sammendrag af Statistiske Oplysninger om Grønland, I." *Beretn. Vedrorende Grønlands Styrelse*, I (1942).
47. Peary, R. E., "Journeys in North Greenland." *Geogr. J.*, 11:213–240 (1898).
48. Raup, H. M., "Phytogeographic Studies in the Athabaska–Great Slave Region." *J. Arnold Arbor.* 27:1–85 (1946). (See p. 69)
49. Robinson, J. L., *The Canadian Eastern Arctic*. Ph.D. Dissertation, Clark University, 1946.
50. Sanderson, M., "Drought in the Canadian Northwest." *Geogr. Rev.*, 38:289–299 (1948).
51. —— "Experimental Measurements of Evapotranspiration at Norman Wells, N. W. T." (in preparation).
52. Supan, A., "Die Temperaturzonen der Erde." *Petermanns Mitt.*, 25:349–358 (1879).
53. Suslov, S. P., *Fizicheskaya Geografiya S.S.S.R.* Moskva, Uchpedgiz, 1947. (Chart of vegetation in folder.)
54. Sverdrup, H. U., *Meteorology, The Norwegian North Polar Expedition with the* Maud, *1918–25*, II, Pt. I, Discussion. Bergen, 1935.
55. —— Petersen, H., und Loewe, F., "Klima des Kanadischen Archipels und Grönlands," *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Hsgbr., Bd. II, Teil K. Berlin, Gebr. Borntträger, 1935.
56. Thornthwaite, C. W., "The Climates of North America according to a New Classification." *Geogr. Rev.*, 21:633–655 (1931).
57. —— "An Approach toward a Rational Classification of Climate." *Geogr. Rev.*, 38:55–94 (1948).
58. Tikhomirov, E., *Climatological Handbook of the Soviet Section of the Arctic*. Moscow, 1940.
59. U. S. Hydrographic Office, *Ice Atlas of the Northern Hemisphere*. Washington, D. C., 1946. (See Charts 1–12)
60. —— Articles on reverse of various issues of monthly *Pilot Chart of the North Atlantic Ocean*, as follows: "Arctic Ice," June, 1947; "Arctic Ice and Its Drift into the North Atlantic Ocean," 9th ed., Supp. to Chart of May, 1949; "Physical Properties of Sea Ice," July, 1946.
61. U. S. S. R. *Bol'shoi sovetskii atlas mira.* (*Great Soviet World Atlas.*) Moskva, 1937–1939. (Vol. I, V. E. Motylev, ed., 1937. Reviewed by G. B. Cressey in *Geogr. Rev.*, 28:527–528 (1938). Vol. II, S. A. Kutaf'ev, ed., 1939. This great, two-volume work is very inaccessible. It contains a mine of climatological material.)
62. Vahl, M., "Zones et biochores géographiques." *Overs. danske Vidensk. Selsk. Forh.*, pp. 269–317 (1911).
63. Ward, R. de C., Brooks, C. F., and Connor, A. J., in collaboration with Fitton, E. M., "The Climates of Alaska," Pt. 3 of "The Climates of North America," *Handbuch der Klimatologie*, Bd. II, Teil J, W. Köppen und R. Geiger, Hsgbr., 5 Bde., Berlin, Gebr. Borntträger, 1936.
64. Weaver, J. E., and Clements, F. E., *Plant Ecology*, 2nd ed. New York, McGraw, 1938.
65. Wegener, K., *Wissenschaftliche Ergebnisse der deutschen Grönland Expedition Alfred Wegener, 1929 und 1930–31*, Bd. 4, Teil I. Leipzig, Brockhaus, 1933.
66. Wenner, C.-G., "Pollen Diagrams from Labrador." *Geogr. Ann., Stockh.*, 29:137–373 (1947).
67. Wexler, H., "Cooling in the Lower Atmosphere and the Structure of Continental Polar Air." *Mon. Wea. Rev. Wash.*, 64:122–136 (1936).
68. Wonders, W., *Weather and Climate of the Canadian Arctic Archipelago*. Ph.D. Thesis, University of Toronto, 1950.
69. Zubov, N. N., *L'dy Arktiki*, Moskva, Izd. Glavsevmorputi, 1945. (See review by D. B. Shimkin, *Arctic*, 2:65–69 (1949))

CLIMATOLOGY

Climate—The Synthesis of Weather *by C. S. Durst* 967

Applied Climatology *by Helmut E. Landsberg and Woodrow C. Jacobs* 976

Microclimatology *by Rudolf Geiger* 993

Geological and Historical Aspects of Climatic Change *by C. E. P. Brooks* 1004

Climatic Implications of Glacier Research *by Richard Foster Flint* 1019

Tree-Ring Indices of Rainfall, Temperature, and iver Flow *by Edmund Schulman* 1024

CLIMATE—THE SYNTHESIS OF WEATHER

By C. S. DURST

Meteorological Office, Air Ministry, London

Introduction

Looking back over the last half century, one sees how greatly the approach of scientists to the meteorological problem has changed. The nineteenth-century meteorologist was in most cases an amateur; even the professional forecasters were not educated in a scientific approach and learnt their trade by experience and observation.

It was Shaw who first saw the possibilities of a professional meteorological service and it was Shaw's researches reported in the *Life History of Surface Air Currents* which started the great wave of scientific meteorological thought on synoptic matters which has transformed meteorology into a science. If, however, one looks at our progress dispassionately, it is apparent that the greatest advance in this scientific attitude has been in those branches which deal with the forecasting of weather and the physical problems which are associated with it; one can trace the stimulus in that direction back to the illumination of the problem by V. and J. Bjerknes in the second decade of this century.

The emphasis which has been placed, in the scientific mind, on forecasting has been reinforced by the insistent demands which have arisen from aviators for help in their day-to-day flights, in much the same way that the necessity of knowledge of the climate of the oceans stimulated Maury and FitzRoy in the nineteenth century to the organisation of the charting of the winds of the sea areas in the very earliest days of organised meteorology. One must admit that the scientific approach has not permeated so deeply into some other branches of meteorology as into forecasting. In spite of the great work of systematisation done by Köppen and Hann and their persistent groping for the physical causes, there has not generally been an insistence on knowledge of the physical reasoning which must underlie climatology. In its application to hydrology, in its use in agriculture, particularly in climatological application of rainfall data to the needs of man, there has been a woeful tendency to the use of the bones of bare statistics and mean values without the flesh of physical understanding. But it is not hydrology and rainfall that are to be discussed in this article. A more general proposition is put forward: that climatology, as at present practised, is primarily a statistical study without the basis of physical understanding which is essential to progress. As I see them, the essential needs of climatology are in the first place a reorientation of the expression of climate and of the teaching of climate, and secondly, the explanation of climate as a physical and dynamical phenomenon. The latter needs to show the process by which one season succeeds another.

The Expression of the Climate of the World

Twenty years ago, climatology was assumed to consist of the meticulous amassing of the monthly means and extremes of elements, the mapping of normals, and the description of climate in terms of those normals and extremes. Yearbooks were published and the mass of data increased, but the meanings of those data were not examined. Of the attempts made to summarise and systematise them the best, at any rate in English, was that due to Kendrew [17] whose *Climate of the Continents*, after more than a quarter of a century, is still the most lucid description of the normal distribution of the elements over the land masses of the world. In reading it, however, one has a certain feeling of suspended animation, of unreality because everything is static; monsoonal wind arrows, indicating speeds of 10 or 20 mph on the average, change the pattern of flow not a whit from month to month. Of this Kendrew was himself well aware and in the preface to a later work entitled *Climate* [18] he says:

> . . . between the Tropics and the Poles the weather is so variable . . . that it is difficult to form a conception of the climate unless it be the idea of something very changeable. Certainly no picture of climate is at all true unless it is painted in all the colours of the constant variation of weather and the changes of season which are the really prominent features, and it is inadequate to express merely the mean condition of any element of climate.

The treatment that Kendrew then gave is by elements in contrast to the *Climate of the Continents*, which as its name implies discusses climatology on a regional basis. Very recently Kendrew has revised his *Climate* and renamed it *Climatology* [19]. In this revision he has taken the various elements and discussed the physical causes which lead to their changes and then has endeavoured to synthesise the picture into the description of the climate of mountain and plateau and the climate of typical regions—the Sudan, the Mediterranean, and the Westerlies. Here, one is aware that the elements obey physical laws and some of the tools are provided with which to work out for oneself the picture of the climate of any locality in which one may be interested; however the completeness which the perfect climatology should express is still lacking.

Mention must also be made of Miller's *Climatology* [25], which originally treated the subject in the conventional manner, but into which, in the most recent edition, is inserted a chapter on the more modern ideas of air masses, though the more statistical approach is retained in the rest of the volume. More recently Haurwitz and Austin [14] have combined in one volume what Kendrew did in two. They logically treat first

the elements severally, and then in a second part discuss the various regions of the earth.

Undoubtedly as a general introduction to climatic studies some such textbooks as those mentioned above are an essential on which to build, but there remains still an indefinite desire for a new approach; perhaps this desire cannot be satisfied in any work which embraces the whole world in its purview and we must be content to treat the regions in turn.

The Demands of Aviation Climatology

Fifteen or twenty years ago a new demand arose. Aviation was spreading. The new machine could be persuaded to cover longer and longer distances. Weather was all-important, but the machine flew out of the weather map—and indeed we were ignorant of how we could use a weather map in the tropics even if we had observations with which to make one. The aviator clamoured for guidance for his flight and none could be given but the data of the climatologist. He was shown temperature and rainfall maps and the data accumulated in Knox's *Climate of the Continent of Africa* [20]. The rude aviator said that was not what he wanted. He did not greatly care if it were hot or cold. He wanted to know about the winds, and even so he was not flying on the ground. He was more interested in cloud than rain. He did not mind getting wet; what he wanted was to see where he was going. The International Commission on Air Navigation had indeed, in its wisdom, recommended that frequency summaries should be made of upper winds, of visibility, and of cloud heights. But as yet there were no frequencies for Africa and there would not be any for many years. But the rude aviator said he was flying to the Cape forthwith and he could not fly on Knox.

Clearly a new type of climatology was needed—descriptive rather than statistical—a translation by the meteorologist of means and summaries into something that the aviator could comprehend. In this, synoptic training came to the help of the climatologist. He was forced back on the physical processes underlying Knox's data and on their basis described what he expected the clouds and the winds would be. In this, the idea of air-mass characteristics was ever-present on the lines developed by J. Bjerknes and applied to the world at large by Bergeron [7], but in those days (the twenties) the method had been little applied in detail and the tropical air masses were even less comprehended than they are today. For the African Continent, with which our rude aviator was concerned, much help was obtained from the analysis by Brooks and Mirrlees [11]. In those days, when the writing of descriptive climatology was so little developed, some passages in Braak's "Climate of the Netherlands Indies" [9] were an inspiration, particularly the one in which he described the scene from Mount Pangerango and the changes of the clouds in their regular daily variation. This recalled the same sequence visible from the Malayan Mountains; but, and here is the necessary caution, it was remembered that the sequence may break and for some days give place to another and different order of events. Why? The facile answer is "a change in air mass," but the physical basis has not yet been explained.

The Description of Climate by Air Masses

During the latter part of the second decade of this century, air-mass analysis came to the fore under the lead of Bergeron [7] and Schinze [26]. The conception was that there are definite regions in which the air tends to become horizontally homogeneous. These regions are those where the surface of the earth is uniform, whether land or sea, and where winds are light for a sufficiently long time for the properties of the atmosphere to reach equilibrium. These regions were defined as *source regions*. As air drifted from a source region over some other surface, modification took place and the structure and characteristics of the air mass changed with time.

In 1930 Bergeron [7] propounded the use of air masses as a method of describing climate that would give a picture less static than that given by monthly normals.

In the same year Linke and Dinies [23] formulated a system of designation of air masses for central Europe, which Dinies [12] used in 1932 to assess the frequency with which Germany lay under the influence of different air masses, denominated in eight classes, depending on their origin and their track. This paper was, it would seem, the first serious attempt to present the climate of a region as an entity built up from the characteristics of the air masses. Average values were obtained for temperature and humidity at the surface in each type of air. Frequencies of those air masses, given in numerous tables, are summarized in Table I.

Dinies' investigation was focussed on Frankfurt am Main. The results were of considerable interest, showing up, as would be expected, the strong contrast between air of tropical and polar origins, not only in temperature but also in humidity and cloudiness. The bitter polar continental air of winter stands out in strong contrast to the moist and mild tropical maritime air, a contrast stronger indeed than between any of the summer air masses. For comparison with Table I it may be mentioned that Belasco [6] has given a mean winter temperature of −1C for arctic air which had arrived at Kew Observatory via Russia and Germany. On the other hand he gives the mean value of temperature in tropical air in winter as 10C.

Dinies also surveyed the frequency with which the air masses penetrated to the several parts of Germany. He showed this by means of maps of the average number of days on which the different air masses were over various places in Germany. These maps are on an annual basis and they show the not unexpected effect of penetration of maritime air from the west and southwest and the countersurging of continental air from the east, as well as the sweep of the polar and polar continental air from the northeast.

The object of Dinies' treatment was to present climatology in a much more dynamic picture than had been possible in the past when the climatologist dealt

primarily in the "sterile mean." In this he achieved a considerable measure of success as may be judged even from the brief summary in Table I. His treatment certainly did present the month or season as one of change and conflict. More recently the description of the climate of Germany has been carried a stage further by Flohn [13] who has discussed the regional climatology and has pointed out the effects of the major topographic features on local weather in the different air masses.

Dinies' example has been followed by others, notably by Landsberg [21] who examined the air masses of central Pennsylvania, by Batschurina and his collaborators [5] who reviewed the air masses of Western Russia, by Schamp [31] who applied a classification to the mass properties of America, adopting a somewhat different classification from that of Bergeron in order to suit the local circumstances.

In 1940 Petterssen [29] summarized and reviewed the various assessments of the thermal structure of different air masses and presented original maps of the Northern Hemisphere, setting out the positions of the different sources in winter and summer.

Still more recently Belasco [6] has made elaborate assessments of the temperature and humidity structure of the air of various types reaching the British Isles along a large number of tracks. Perhaps the most interesting feature of this paper is the comparison which he makes of the structure of the air over the Azores

Table I. Frequencies and Characteristics of Air Masses at Frankfurt am Main 1924 to 1929

	Polar	Tropical	Maritime	Continental	Polar Maritime	Polar Continental	Tropical Maritime	Tropical Continental	Ill-defined	Mixed
Frequency of air masses										
Winter (1924–30)%	4.6	2.2	30.2	20.6	11.4	2.6	8.2	1.3	7.0	11.9
Summer (1924–30)%	1.8	3.4	34.0	6.6	32.4	0.8	4.6	1.0	8.6	6.8
February										
Mean temp. (°C)	1.8	3.5	5.8	−2.8	0.8	−15.8	8.3	4.5	—	—
Max. temp. (°C)	6.2	10.4	7.8	4.2	3.6	−10.4	11.1	10.5	—	—
Min. temp. (°C)	−2.3	1.4	2.4	−4.0	−1.3	−19.0	6.2	−0.7	—	—
Vapour pressure (mb)	5.1	6.7	7.5	3.5	5.2	1.6	9.1	5.6	—	—
No. of occasions	5	2	41	27	10	4	18	?	—	—
Winter										
Overcast days (%)	5	44	72	32	52	8	81	—	—	—
Clear days (%)	53	22	0	37	5	38	0	—	—	—
July										
Mean temp. (°C)	—	22.5	18.9	23.8	17.3	—	21.4	—	—	—
Max temp. (°C)	—	28.6	23.8	30.4	21.8	—	29.9	—	—	—
Min. temp. (°C)	—	16.6	14.4	17.0	13.1	—	15.6	—	—	—
Vapour pressure (mb)	—	18.1	15.5	18.5	12.0	—	19.5	—	—	—
No. of occasions	0	2	56	17	29	0	6	0	—	—
Summer										
Overcast days (%)	13	13	36	4	39	0	44	—	—	—
Clear days (%)	13	25	8	23	5	75	6	—	—	—

Greek Islands, and by Roy [30] who has more recently attempted to apply air-mass analysis to the Indian Peninsula. In this careful study Roy has not only classified the air masses, using the data of modern synoptic charts and aerological ascents, but in addition he has described concisely the climatological features of the air masses of the various seasons in regard not only to their temperatures, humidities, and thermal structure, but also in regard to the incidence of fog, and the characteristics of cloud. Roy suggests that there should be compiled in India detailed air-mass climatological tables for a number of representative stations as a routine to form the basis of the discussion of the air-mass climatology of the subcontinent.

An extension of the examination of air masses was made by Palmén [27] by forming mean values of temperature and humidity in the upper air in the several air masses, using data for the British Isles. In this he used essentially Bergeron's classification of air masses [7]. Willett [33] at about the same time listed the air- and Iceland with that which reaches the British Isles after travelling from those directions. In Tables II and III are set out the comparisons of mean temperatures in winter and summer.

A caution must be given that in Belasco's work the air trajectories were not taken back to the source but only for a distance of about 1500 miles from the British Isles. Hence there is, for example, no certainty concerning the actual source region of the trajectories which passed over the Denmark Strait on their course to the British Isles. Table II however does give some idea of the manner in which modification takes place in the temperature of the air when it passes over long stretches of water and at the same time travels into a latitude where the radiation balance is reached at a different temperature level.

All these studies of the upper air in relation to air-mass classification were intended primarily to further the task of the synoptic forecaster. To a great extent, however, they can be made to serve the climatologist in

the manner which Bergeron, Linke, and Dinies envisaged, though it must be remembered that they are typical values and not means or normals.

Table II. Comparison of Temperatures over the Azores with Those of Air Reaching Britain from the Azores

	Summer					Winter				
Height (mb)	900	800	700	600	500	900	800	700	600	500
Tropical air arriving over Azores (°C)	16	13	8	1	−7	10	7	1	−6	−15
Tropical air arriving over Britain from the Azores (°C)	14	11	5	−2	−11	8	4	−2	−9	−18
Maritime polar air arriving over Britain from the Azores (°C)	11	4	−2	−9	−19	4	−3	−9	−17	−27

Table III. Comparison of Temperatures over Iceland with Those of Air Reaching Britain from Iceland

	Summer					Winter				
Height (mb)	900	800	700	600	500	900	800	700	600	500
Polar air arriving over Iceland (°C)	1	−6	−12	−18	−28	−13	−19	−26	−33	−43
Polar air reaching Britain from Denmark Strait (°C)	8	1	−4	−11	−20	−2	−8	−15	−23	−32
Polar air reaching Britain from Jan Mayen (°C)	5	−1	−7	−16	−26	−6	−12	−20	−29	−39

Difficulties of Air-Mass Climatic Description

The primary difficulties in using air masses in climatological work are (1) the difficulty of a definition of air masses which can have world-wide application, (2) the difficulty of dealing with frontal weather, and (3) the complications introduced by vertical motion in anticyclones and wedges.

To some extent the first difficulty is also encountered by the synoptic meteorologist, but in practice the forecaster for a particular region concentrates on the air-mass types which are most likely to occur in that region. It is of far less importance to him that the definitions in the more remote parts of his chart are not precisely in accordance with the physical conditions there, since before the air from those remote regions has entered into his working area, it will have become modified to a great extent. In his original formulation Bergeron [7] set out a comparatively simple classification, but as is pointed out by Willett [33] this classification was made primarily from the study of the air masses which affected northwestern Europe and though it was possible to apply it on world-wide scale, it did not suit the local conditions of the North American Continent. To a very considerable extent the difficulties in adopting a world-wide classification are difficulties of scale. The broad view which encompasses the world will fail if attention is fixed on one continent alone, because of necessity the detail considered will be greater. In the same way the classification suitable to the continent will be less applicable to a smaller unit, state or country. It would be logical to apply the same reasoning right down to microclimatic studies, since the essence of air-mass analysis is the vertical structure, and it is possible to conceive a discussion of microclimate by means of micro-air-mass analysis. Indeed that is in fact the method which leads to the insistence on the influence of atmospheric stability in the more recent studies of microclimate (*e.g.*, Pasquill's discussion of the diffusion of water vapor and heat near the ground [28]).

In the second place a difficulty, which has no counterpart for the synoptic meteorologist, arises in the use of air masses in climatology, namely the difficulty of the expression of the effects of frontal passages. It may be comparatively simple to give a generalised picture of the weather which is likely to occur in the cores of the air masses, even when they are suffering marked modifications, but the generalisation of the weather at frontal zones in all their manifold diversities is most complex. A smoothing of the elements, which is the result of the climatology of the mean, masks much that is essential, but how should frontal zones be shown and classified in air-mass climatology? The plotting of the normal position of deformation fields and fronts such as is shown by Bergeron [7] over the Pacific reveals much more, particularly in the tropics, but it is not an ideal generalisation of the arctic front nor of the polar fronts in higher latitudes, because of the wide range through which those fronts are likely to oscillate.

The third difficulty is the greatest of all. Air masses are not only transformed by the surfaces over which they move and the radiation which they receive; they are also profoundly influenced by subsidence and at present we are not in a position to estimate the magnitudes of those effects. All that seems possible, therefore, is to show how air masses have in the past been affected by moving over different surfaces with a proviso that no account is taken of the presence or absence of subsidence.

The Use of Synoptic Maps in Climatology

From what has been said above it is clear that for certain purposes the climatologist's task and that of the synoptic forecaster were becoming increasingly dependent on the same technique. Indeed the synoptic forecaster from his memory of his maps could write

down the climate of a particular locality with a facility which might not be available to a statistical climatologist. It is true the forecaster had to contend with difficulties not least of which was that he knew too much. To him, intent on his daily synoptic charts with their infinite diversity, it was not easy to generalise without so much detail as to mar the climatological picture. But it could be done, and it was possible thereby to convey a far more graphic picture than would be possible from the consideration of any number of statistics and normal charts.

With the coming of World War II the need for descriptive climatic information became more pressing, and this technique for the use of synoptic charts and daily observations was developed in Britain in a series of manifolded reports on the climate of various regions. In these the attempt was made to describe the effects which were likely to be produced in the neighbourhood of different topographic features when certain air streams impinged on them, the object being to convey to the reader the idea of climate as a synthesis of the daily weather as opposed to the rather static presentation given by a conventional write-up of the normal maps.

One great advantage of this method was that the climatic description was readily assimilable for forecasters, so that a forecaster arriving in a new region could have a broad outline of his forecasting problem before him from the outset.

An example of this presentation is the description of the climate of Southeast Asia by Hare [1]. From the examination of whatever weather maps could be obtained or constructed—and in this the *Daily Synoptic Series, Historical Weather Maps* [32] was of the greatest value—Hare isolated certain air flow patterns to which the weather map tended to return if it were disturbed, and he identified certain positions of the quasi-permanent fronts to which they too returned. Thus he was able to describe the climate associated with the flow pattern, particularly in the case of the winter monsoon, in some detail, and moreover to emphasise what departures could be expected with changing synoptic situations and how those situations once more reverted to the standard pattern. At the same time he drew attention to the characteristics of the air masses, using a local modification due to Huang [15]. The known high frequency with which the fully developed winter monsoon held sway served to guide the reader as to what climate to expect, and the simplification led him to a concept of change within a repeating pattern. It is true that the simplification, like all simplifications, masked some degree of detail, but in climatology some detail must inevitably be lost and the generalisation of weather into this climatic representation gave something that was easily assimilated.

The season of the summer monsoon was less easily generalised, but here too on the basis of the synoptic charts it was possible to establish a broad outline of the repeating patterns of the air streams. The transition periods of spring and autumn were treated in a similar way.

After Hare's work had been produced, Lu's "Winter Frontology of China" [24] was published in the *Bulletin of the American Meteorological Society*. Lu's account was on a more elaborate scale, but was essentially a statement of climate in terms of air masses and fronts. It bore out Hare's analysis to a high degree and the comparison is interesting in that it shows how much can be achieved in this type of representation by one who has no greater acquaintance with the country of which he is writing than is afforded by a series of weather maps.

At different dates from 1939 onwards British reports on similar lines were prepared for most of the theatres of war. They were manifolded, but the number of copies was limited.

To impress on the reader the variability of the climate, some of these reports included diagrams which

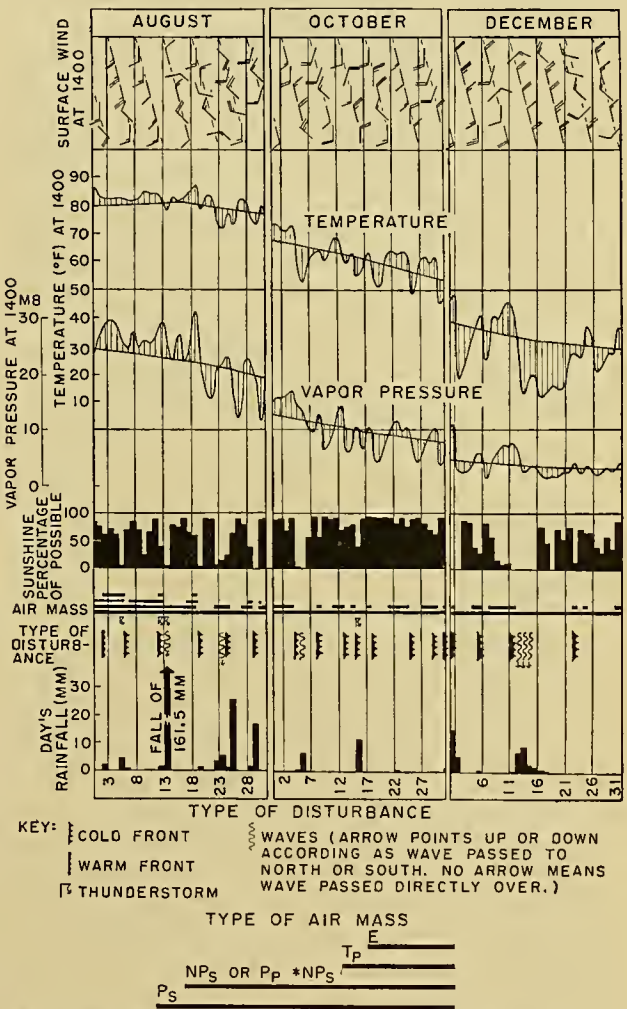


Fig. 1—Daily weather at Dairen, Manchuria, for August, October, and December, 1929.

gave the daily observations of a year ranged out on a single sheet of paper. An extract from one such diagram is shown in Fig. 1 in which the months August, October, and December, 1929 are illustrated for Dairen in Man-

churia. It is not possible to reproduce conveniently more than this extract and, as would be expected, the complete specimen year conveys a far better idea of the normal change, but even from these three months one receives the impression of how the elements vary from one day to another. In this diagram, wind, temperature, and humidity are illustrated as well as sunshine, pressure, and rain. In addition, the air-mass classification and the fronts which lie in the neighbourhood of the station are shown.

These studies were essentially climatic and contained, in addition to the descriptive matter, tables of normals to serve as the inevitable background. Here it may be emphasised that, however the climatological data are discussed, accurate and well-arranged general tables of normals must be published if only to serve as a yardstick against which to measure the variability.

By way of contrast between the description according to conventional climatology and that which can be built up from the study of synoptic situations, it is of interest to compare the description of the winter climate of China as given in Kendrew's *Climate of the Continents* [17] with that given by Hare [1] and Lu [24]. Admittedly it may be objected that readers for whom Kendrew wrote were less erudite than Lu's audience or even than Hare's, but if one type of description is read and then the other, it will at once be apparent which is the truer account.

In making this contrast between the conventional approach and that of the synoptic meteorologist, I do not wish to veil the value of the mass of information which conventional climatology has provided. It has indeed formed the basic background from which it seems there is now scope for advance in a more detailed description of climate, a description based on the physical causes which in the aggregate produce that which is called climate.

Since World War II, the procedure which has been described above in regard to the use of synoptic charts has provided the basis for several reports such as *Aviation Meteorology of South America* [4] for which J. S. Sawyer was responsible, and *Aviation Meteorology of the Azores* [3] by L. Jacobs and R. M. Murray.

This method of approach was the natural outcome of the need for a similar type of information in other countries. In Germany it would seem that methods of air-mass climatology were more or less dropped during the war, at any rate in their more detailed application. In a number of climatic reports the Germans made use of representative types of weather, illustrating them with typical weather maps for individual days, but there does not seem to have been the systematisation which might have been expected. The Germans in their reports seem to have relied largely on a great mass of statistics suitably arranged for aviation and related needs.

In America the stress on securing information as a basis for planning during World War II led to a rather similar approach under the lead of W. C. Jacobs [16], who assembled large masses of data on punch cards for the region in which he was interested. By sorting for one element (*e.g.*, air flow) he was able to draw up a picture of the probable weather with which each type of flow was associated. However, with the advantages of the punch-card system it was possible to break down the problem still further in whatever direction was most advantageous, to calculate, for example, the chance that low cloud would occur with a particular general wind flow and the chance that the low cloud would affect one region when another was clear.

From the practical point of view these summaries could be linked effectively with the weather map, for the forecaster could tell at a glance which particular type of the numerous arrangements of the data was most like the chart with which he was dealing. However, the presentation of the data was essentially statistical; it masked rather than emphasised the physical processes underlying the significant weather. Nevertheless, as a practical tool in climatology, the punch card, wisely used, will undoubtedly play its part in the final solution of the climatic problem.

The Russian approach to the problem of climatic representation is shown in a recent publication by Alisov [2]. In his foreword, Alisov is emphatic that the study of climatology must not end with the collection of data, but that it must be discussed on the basis of the physical process by which the climatological phenomena have been developed. He then proceeds to discuss the general climatology of Soviet Russia as a whole on the basis of the origin of the air masses and their modifications in transport. The remainder of his extensive book is devoted to a more detailed description of the climates of the different regions, but in each case the description is based on an integrated knowledge of the synoptic charts rather than on normal values.

Climate as a Physical and Dynamical Problem

The climatologist will not be content with the mere recording of observations nor with their description, even if that description is based on the physical processes in the atmosphere. It will be his natural desire to seek a physical explanation of climatic change, the change from one month to the next, as well as the changes which occur from one period of years to another. The climatologist has for long been intensely interested in long-range forecasting—seasonal foreshadowing as Sir Gilbert Walker termed it. But the methods by which the mechanism of climate is investigated have been sadly lacking in a reasoned approach.

The methods most frequently adopted are (1) the determination of correlation coefficients between a meteorological element in one place and the same element (or another) in some other place, perhaps with a time lag, (2) the association of the pattern of a mean distribution of a particular element in a particular month with the pattern of the mean distribution of some other element in the same or a different month, and (3) the association of maps of monthly pressure or temperature anomalies with the anomalies of the same element in the succeeding month. One should add to these (4) the

relation of the circumpolar pressure (or wind) distribution averaged over a short period (five days) with subsequent developments of the pressure pattern.

In the first three of these methods the choice of element and the position on the globe are purely at the caprice of the investigator. They are in fact almost entirely blind speculation—hit or miss, if miss, try again. It may be said that, in the past, many of the discoveries in physics were based on such procedure, but in a subject such as climatology covering so wide a geographical field and with so many physical variables the chances for success by such methods are very small. The fourth method based on the movement of long-wave patterns has a dynamical basis and as such is in a different category. The movement of these waves and the meridional flux of air profoundly influence spells of weather, therefore in some manner these long waves and in particular the blocking anticyclones must be taken into account in any concept of physical and dynamic climatology. Nevertheless the path which should be pursued in the construction of such a climatology is not clear.

Following the line of thought in the previous part of this article, we are led to consider the climatological map of any element as the representation of an integration of the values of that element derived from trajectories which have their origins in certain source regions. Now the appropriate value of the element on any occasion is influenced firstly by the state in the source regions and secondly by any modification of that state introduced during its travel, whether by diffusion or radiation or change in pressure or by any other cause.

Clearly we may consider the map for a single month in the same way, and we may review the change in that map by consideration of the change in the source regions or the change in the modifications introduced in the travel of the air masses, as well as by the frequency with which the air travels from the different sources. Such an approach to the problem of climatology has the merit that it endeavours to link the mechanism of climatic change with physical properties, and it is clear that if any advance were possible from this approach, the horizon of usefulness would extend greatly.

The source regions of the Northern Hemisphere are enumerated in considerable detail by Petterssen [29]. They are generally recognised as being regions in which the surface of the earth has a homogeneous covering; there is the additional proviso that the air motion must be sluggish in order that the air mass may acquire uniformity of character. It is known from the synoptic charts that these source regions are liable to considerable movement from one period to another, some indeed may be swept entirely away at times. Their boundaries in any case are rather indeterminate, yet the very fact of the use that is made of the air-mass characteristics in synoptic meteorology emphasises that there is a recognisable continuity in them and gives hope that the examination of the variation with time of the character of source regions is a practicable step. That such an examination has not been made before now, so far as the present writer knows, is probably because a number of the source regions, and those the most important ones, are in locations that are by no means readily accessible, particularly to upper-air observation. However, this difficulty is becoming less with the extension of information.

Now that the drawing of synoptic charts of the whole hemisphere is a practicable process, not only daily but two or even more times each day, there would seem to be an urgent need for an analysis of the variations from day to day in the position, the intensity, and the growth (or decay) of the major sources. Admittedly any such examination is liable to a measure of subjectivity, since the definitions of the source regions are elastic, but even with that limitation such an analysis would teach the meteorologist much.

The next problem is that of the modification of air masses in translation over various surfaces.

Belasco's calculations of the modification of polar air travelling from Iceland to Britain and of tropical air travelling from the Azores are given in Tables II and III. From these figures it would seem that in summer the air in travelling from Iceland to Britain (say 1000 miles) is warmed by about 5C or 6C all the way up from 900 mb to 500 mb, whereas in winter it is warmed by about 9C at 900 mb, but by only about 7C at 500 mb. On the other hand tropical air, in travelling from the Azores (say 1200 miles), is cooled by about 3C all the way up in winter, but by only 1C in summer at 900 mb and by about 3C at the higher levels. In these calculations no allowance is made for change in direction of the wind streams at the greater heights, but the figures would seem to show the magnitudes of the changes.

We still do not know to what extent these changes in temperature arise from convection of heat from the surface, from radiation, or from vertical movements; fields of deformation will have to be considered and the consequential effects on frontogenesis as suggested by Bergeron [7]. However, with the more complete modern maps of the upper air a survey of the changes of temperature (and possibly of humidity) along the tracks of the air masses becomes possible. By these means we may be able to set out more exactly the manner in which the air masses tend to be changed in relation to the surfaces over which they move, and the atmospheric processes affecting them.

After these basic data have been determined, an examination could be made of the life history of the air masses which combined to form the monthly mean values for individual years. A summary of such life histories in a suitable form would undoubtedly enlarge our knowledge.

Severe winters, for example, are in many places chiefly due to the advection of air from intensely cold sources, but we do not know whether they are greatly affected by the intensity of the source, or whether the severity is the result merely of the persistence of the flow; on the other hand, the source may have moved

nearer to the climatic station in those winters, or possibly the track of air has been consistently more direct.

In this way, with much initial labour no doubt, there is hope of getting an insight into the mechanism of climate. It seems very probable that we should learn what factors exert the greatest control, as, for example, in the case cited above, whether proximity of the cold source or persistence of flow carries the major weight in severe winters. If we were able to do this, it might then be possible to resort to statistical methods as a means of foreshadowing the change of the climatic mean from one month to the next, for if we had a better conception of the mechanism of climate, we should be in a far more favourable position for selecting the right form that should be taken by the regression equation.

Some Basic Requirements for the Discussion of Climate

From this review of the methods adopted in the last twenty-five years to express climate certain principles seem to have emerged.

It is clear that scale is the ruling factor in the treatment of climate; nevertheless, the treatment at each level of the scale has to be based on stability or instability characteristics, not necessarily in the rigorous classification of the air-mass climatology school. In some way, descriptions ranging from a world-wide review to a local microclimatic study have to be permeated with the sense of the facilities for vertical transport in the air. It may be that the shorthand expressions used in air-mass analysis are the best; it may be that some better method of describing stability could be devised. The reason for these necessities is patent in the fact that the phenomena of rain and cloud are so closely associated with the stability of the atmosphere; clouds control the amount of heat reaching the earth or escaping from it. Temperature lapse rate too is intimately connected with turbulence and consequently with heat transport from the surface.

The treatment of a world-wide climatology would admit a scale such as that designed by Bergeron in his original exposition, but no greater elaboration. The scale for treatment of an area such as a state or country would need, as Willett and Linke have agreed, an elaboration of detail that would be overwhelmingly cumbrous for the world as a whole. And so on down to the microclimatic studies which, in the case of agriculture, might well need to consider the modifications in the vertical structure of the micro-air-mass due to variations in the surfaces of neighbouring fields.

Perhaps the next lesson to be learnt is the necessity for the synoptic worker and the climatologist to become accustomed to thinking in each other's terms; indeed, to thinking that there should be no demarkation between them, that they should use each other's tools. There would seem to be good ground for maintaining that climatological literature should be written, in part at least, by synoptic meteorologists, though of course this in no way means the exclusion of tables of climatic normals; they are essential as a basis to be used for reference.

There is a great need for detailed climatological literature, based on synoptic information, to be prepared for all parts of the world. One of the advances in climatology which is most likely to take place in the near future is the publication of papers dealing with this type of synoptic climatology.

For the understanding of climate it seems that we need to know much more about the quasi-permanent sources of air masses. In comparison with our knowledge of the structure of depressions, the comprehension of the processes in anticyclones is very small; but anticyclones may be the key of the climatic mechanism because they are the source regions of the air. Perhaps then one of the most interesting problems is that of the process by which air flows into the anticyclone and the reason why it does so. A solution of this problem would throw light on the mechanism of climate.

In regard to sources we need to know more about the time taken for stagnant air to become homogeneous, and what influence radiation has on the establishment of homogeneity. Also we should have information on the movements and the variation in intensity of sources from year to year and whether that intensity is associated with extraterrestrial phenomena.

It may be objected that all the problems just mentioned are those of the synoptic forecaster. Perhaps this is the crux of the whole matter. The physical problem of synoptic forecasting and climatology is the same, for climate is but the synthesis of weather.

REFERENCES

1. Air Ministry, *Meteorological Report on South China.* Aviation Meteorological Report, No. 29, M. O. M. 365/29. London, 1945. (Out of print.)
2. Alisov, B. P., *Klimaticheskie Oblasti i Raŭony S.S.S.R.* Moskva, Geografgiz, 1947.
3. *Aviation Meteorology of the Azores.* Meteorological Report No. 2. London, H. M. Stationery Office, 1949.
4. *Aviation Meteorology of South America.* Meteorological Report No. 1. London, H. M. Stationery Office, 1949.
5. Batschurina, A. A., Blumina, L. I., and Petrowa, L. I., "Die Einteilung der Eigenschaften der troposphärischen Luftmassen am Norden des europäischen Teils des U.S.S.R. im Sommer." *Zh. Geofiz. Meteor.*, Vol. 6, Nos. 2–3, pp. 201 ff. (1936).
6. Belasco, J. E., "Characteristics of Air Masses over the British Isles." *Geophys. Mem.*, (in press).
7. Bergeron, T., "Richtlinien einer dynamischen Klimatologie." *Meteor. Z.*, 47:246–262 (1930).
8. ——"Über die dreidimensional verknüpfende Wetteranalyse." *Geofys. Publ.*, Vol. 5, No. 6 (1928).
9. Braak, C., "The Climate of the Netherlands Indies." *Meded. ned. meteor. Inst.*, Vol. 2, Pt. 2 (1928).
10. Brooks, C. E. P., "The Distribution of Thunderstorms over the Globe." *Geophys. Mem.*, Vol. 3, No. 24 (1925).
11. —— and Mirrlees, S. T. A., "A Study of the Atmospheric Circulation over Tropical Africa." *Geophys. Mem.*, No. 55, 15 pp. (1932).
12. Dinies, E., "Luftkörper-Klimatologie." *Aus d. Arch. dtsch. Seew.*, Bd. 50, Nr. 6 (1932).
13. Flohn, H., "Witterung und Klima in Deutschland." *Forsch. dtsch. Landesk.*, Bd. 41 (1942).
14. Haurwitz, B., and Austin, J. M., *Climatology.* New York, McGraw, 1944.

15. Huang, H.-c., "Air Masses of North China." *Mem. nat. Res. Inst. Meteor., Chungking*, Vol. 13, No. 3 (1940).
16. Jacobs, W. C., "Synoptic Climatology." *Bull. Amer. meteor. Soc.*, 27:306–311 (1946).
17. Kendrew, W. G., *The Climate of the Continents*, 3rd ed. Oxford, Clarendon Press, 1937.
18. —— *Climate*, 2nd ed. New York, Oxford, 1938.
19. —— *Climatology*. Oxford, University Press, 1949.
20. Knox, A., *The Climate of the Continent of Africa*. Cambridge, University Press, 1911.
21. Landsberg, H., "Air Mass Climatology for Central Pennsylvania." *Beitr. Geophys.*, 51:278–285 (1937).
22. Linke, F., "Luftmassen oder Luftkörper?" *Biokl. Beibl.*, 3:97–103 (1936).
23. —— und Dinies, E., "Über Luftkörperbestimmungen." *Z. angew. Meteor.*, 47:1–5 (1930).
24. Lu, A., "Winter Frontology of China." *Bull. Amer. meteor. Soc.*, 26:309–314 (1945).
25. Miller, A. A., *Climatology*, 4th ed. London, Methuen, 1943. Also, 3rd ed., New York, Dutton, 1943.
26. Moese, O., und Schinze, G., "Zur Analyse von Neubildungen." *Ann. Hydrogr., Berl.*, 57:76–81 (1929).
27. Palmén, E., "Aerologische Untersuchungen der atmosphärischen Störungen." *Mitt. meteor. Zent-Anst. Helsingf.*, No. 25 (1933).
28. Pasquill, F., "Eddy Diffusion of Water Vapour and Heat near the Ground." *Proc. roy. Soc.*, (A) 198:116–140 (1949).
29. Petterssen, S., *Weather Analysis and Forecasting*. New York, McGraw, 1940.
30. Roy, A. K., *Air Masses of India*. India Meteor. Dept. Tech. Note No. 16, 1946.
31. Schamp, H., "Luftkörperklimatologie des griechischen Mittelmeergebietes." *Frankfurt. geogr. Hft.*, 13 (1939).
32. U. S. Weather Bureau, *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere, Sea Level*. Washington, D. C., 1943, 1944.
33. Willett, H. C., "American Air Mass Properties." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 2, No. 2 (1934).

APPLIED CLIMATOLOGY

By HELMUT E. LANDSBERG

Research and Development Board, Washington, D. C.

and WOODROW C. JACOBS

Headquarters, Air Weather Service, Washington, D. C.

Introduction

Definition. If we consider climate as the statistical collective of individual conditions of weather, we can define *applied climatology* as the scientific analysis of this collective in the light of a useful application for an operational purpose. The term *weather* includes all atmospheric events and the observation of individual weather elements as single series synoptically or otherwise combined. The term *operational* is also broadly interpreted as any useful endeavor, such as industrial, manufacturing, agricultural, or technological pursuits.

Historical Note. This type of climatological work reflects an old and historical approach to the subject. The earliest scientific climatic observations, such as the ones sponsored by the Societas Meteorologica Palatina, were made for the sake of knowledge. However, the networks of climatological stations inaugurated upon the urging of Thomas Jefferson, Surgeon General Joseph Lovell, Alexander von Humboldt, and others were directed toward the practical ends of land utilization, agriculture, and health. This pioneering work took place considerably before meteorological networks were established to provide data for weather forecasting. Meteorological observations later became subservient to the aims of synoptic meteorology and remained so for many decades. Climatology became an adjunct devoted to a mere statistical summarization of the data. Only lately has there been a resurgence of the idea that the accumulated climatic data have a high practical value, far beyond their use for the descriptive *average* picture of the atmosphere used in comparative geography, a use which has dominated the field of climatology.

Prefatory Note. The following discussion does not attempt to treat the subject of applied climatology exhaustively. Within the allotted space the current state of the art can, in many instances, be given only by references to pertinent papers. Most stress will be laid on the inadequacies and on the lines that might be explored profitably by future research work.

Analysis of Requirements for Climatic Data in Various Activities

The objective is to make climatic information useful for other pursuits. Therefore, the particular climatic data desired in each instance must be established. It is a fallacy to assume that the usual climatological observations, summarized in the classical format now standard at most weather offices, can immediately be applied to the practical problems at hand. Misuse of the readily available climatic data has led to many misconceptions and faulty interpretations, and in some cases has even discredited the science of climatology.

First of all, climatic data should be interpreted by the expert climatologist. His position in relation to the user of this material is analogous to that of a physician or a lawyer or an architect in relation to his client. This makes it imperative that the climatologist should not only know the weather data but should also *diagnose the case* to which the data are to be applied. Usually, a thorough analysis of the operations for which the climatic data are needed has to be made. Two questions have to be answered:

1. Which climatic elements affect the operation?
2. How can the available climatic material be interpreted for the particular purpose at hand?

In practice it is essential to go through a thought process which is roughly sketched in the two questionnaires given below. These lists of questions are not prepared for any specific case and must be altered to fit each individual problem.

Questionnaire I: Analysis of Operation

(*Climatic Anamnesis*)

1. Class of enterprise (business, industry, agriculture, etc.)?
2. What operational subdivisions exist?
3. Class of problem?
   - *a.* To lead to a design.
   - *b.* To affect a procedure.
   - *c.* To make an operational decision as to "where" or "when."
4. How does the prospective user of the weather information define his *weather* problem?
5. How does he define his *operational* problem?
6. Will the climatic or weather information the user believes he needs furnish a solution to the operational problem?
7. Have solutions previously been obtained for analogous operational problems?
8. Can the operational problem be redefined in more realistic climatological terms?
   - *a.* What periodic fluctuations (diurnal, seasonal, etc.) exist?
   - *b.* What are the critical climatic and weather limitations on the operation?

c. Is it of importance to stay below critical limits or are there known or desirable optimum conditions?

9. Are operational figures (production, yield, etc.) available for correlation purposes? For how many years (cases)?

QUESTIONNAIRE II: ANALYSIS OF CLIMATIC DATA

(*Climatic Diagnosis*)

1. Class of climatological technique required (see Table III)?
2. What types of data are desirable?
   a. Elements (pressure, temperature, winds, etc.)?
   b. Are combinations of elements required? If so, what combinations? Are their relationships known?
   c. What length of record is needed for an adequate solution?
3. What climatic data are available?
   a. At the spot; in the area; over the region?
   b. In what form are the data? Are existing summarizations suitable?
   c. What length of record is available?
   d. Are climatic data available for the same periods covered by the operational information? (See Questionnaire I, Item 9).
4. What are the local peculiarities in climate that may affect the usefulness of existing climatic data?
5. What are the physical or statistical limitations imposed on the data and their interpretation?
6. Does theoretical information (or a suitable technique) exist that can serve to supplement, evaluate, or be substituted for, the inadequate observational materials?
7. Are solutions available for analogous climatological problems?
8. Is there a need to elaborate a new theory of interrelations?
9. What form of presentation of the conclusions is desirable?
   a. Graphs.
   b. Tables.
   c. Formulas.
   d. Alignment diagrams.

It is very essential that both parties involved in the climatological analysis of an operation, the climatologist and the *operator*, collaborate. They should understand each other's problems. As far as possible, this understanding should be reached before the job is begun. Frequent consultations during the progress of the actual work are indispensable.

Data for Applied Climatology

The amount of climatic material in the files of the various weather services is staggering. The number of surface observations is running into at least nine digits. Much of this material has undergone some elementary summarization. We estimate that mean monthly temperature and precipitation data are available for not less than 25,000 localities in the world. Some one hundred million synoptic observations (including data on pressure, temperature, cloudiness, current weather, etc., as expressed in the international code) are now available on punched cards. Upper-air data are also accumulating on punched cards at a rapid rate, although techniques allowing their use in the three-dimensional sense remain to be developed. The major depository of these punched cards is the joint U. S. Weather Bureau—Air Force—Navy punched-card library at New Orleans, Louisiana. This material has a great deal of flexibility. Summarizations and frequency distributions can readily be obtained from the raw material at high speeds [2, 37, 68]. There are, however, important limitations inherent in the sorting and tabulating machines [31]. The climatologist has to realize this. The essence of these limitations is that one cannot get more out of a machine than is put into it. Also, the machine will do no thinking.

Spot observations of meteorological data are useful for a variety of problems in applied climatology, but a great many cases require simultaneous observations at several points or over areas. In many cases it is necessary to associate the observations in time and space with the help of synoptic weather maps. For Northern Hemisphere areas the best source material is the series of daily surface weather maps, extending back to 1899 [84]. Since the end of 1945, the available issues include also five hundred millibar charts and the synoptic telegrams on which the charts are based. Some of this synoptic material has been summarized into mean charts (for a detailed bibliography see [87]). Pressure values and storm tracks for ocean areas have also been placed on punched cards.

It cannot be our task to list here all the multifold sources of climatological material. Only a very few standard works of an international character and some of the most important sources in the United States have been singled out [90].

Much of this material, excellent as it may be, restricts itself to only a few climatic elements. Pressuret temperature, and precipitation amounts are most com, mon. Some sources include elements such as frosoccurrence, which were thought to be of agricultural importance. On the whole, however, the standard climatic summarizations leave much to be desired as yet. Most of them are *mean value* climatologies. Frequently, the series from which they are obtained are inhomogeneous. Another shortcoming is the lack of observations other than the type desired for the purposes of synoptic meteorology.

The climatologist can present a long list of observational desiderata for applied purposes. For many of these the proper observational techniques are inadequate or even nonexistent. Table I lists some such additional observations which would be most welcome in applied climatology. They are listed in two categories. One comprises those where current techniques will yield useful data, and the present difficulty is that data are obtained at too few places; the other lists the elements for which new techniques of observation are needed.

Irrespective of how many stations there are or will be and how many elements are observed at each, we will never have complete and homogeneous coverage of the whole surface of the earth. Even in the densest networks there remain blind spots. One of the important questions is therefore, How can we interpolate between stations? The synoptic reconstruction of an observation which is lacking, or the bridging of an observational gap, is an important technique. Given a series of good synoptic charts, the meteorological analyst can make a guess about the probable values for a given locality. This is not much different from making a forecast, except that the whole sequence of events is synoptically known. However, the reliability of such guesses has seldom been put to a controlled test. Hence the limitations and errors of the method are usually not available to the climatologist. Such information can be obtained by applying objective tests to representative analyses.

The question of how to interpolate between stations also enters into the problem of the representativeness of individual climatic stations for various elements. It can, for example, be assumed that a mean pressure value and a frequency distribution of pressures for a station are valid for distances of many miles. The synoptic method can also be trusted to yield readily corrections to these values over distances of more than a hundred miles. It is doubtful if such extrapolations can be made for other elements. Temperature adjustments have occasionally been attempted by making corrections for topography, but it is doubtful if any reliable extrapolations can be made for temperature, precipitation, and surface winds except, perhaps, over the simplest terrain forms such as the ocean. The reasons for this are (1) that present methods of measuring these elements constitute very poor sampling, and (2) that our present knowledge of terrain influence on temperature, wind, and precipitation is inadequate.

Table I. Some Climatic Elements Requiring (*A*) Additional Observations or (*B*) Better Observations

(*A*) *Insufficient coverage with existing equipment (extent).*	
Solar radiation:	Total intensity, spectral distribution, various exposure angles, illumination, vertical variation in intensity, cloud and surface albedos.
Precipitation:	Weight of ice deposits, weight of snow accumulations, dew, horizontal patterns of intensity (simultaneous), extent of snow cover, depth of snow, character of snow cover, vertical variations in precipitation intensity.
Other:	Soil temperatures, depth of penetration of frost.
	Water temperatures.
	Frequency of lightning discharges to ground.
	Cooling power.
	Suspended particles (dust, condensation nuclei).
	Soil moisture.
	Soil temperatures (vertical distribution).
	Soil and surface temperatures (horizontal distribution).
	Extent and character of sea, lake, and river ice.
(*B*) *Inadequate techniques or instruments.*	
Evaporation.	
Corrosive substances in air.	
Composition of rain water.	
Drop sizes.	
Radiation losses.	
Soil conditions (state of ground).	
Slant visibility.	
Vertical component of wind.	
Cloud thickness and layering.	
Rainfall over oceanic areas.	
Evaporation from water and soil surfaces; rates of transpiration.	
Foliage temperatures.	
Cloud density; areal extent of cloud cover as a function of altitude of observer above or below cloud deck; cloud temperatures.	
Vertical temperature, moisture, and wind structure near ground.	
Rate of heat transfer between soil or water surface and atmosphere (instantaneous rates).	

A daily task in applied climatology is the substitution of an observed element for one that is desired but has remained unobserved. Such substitution can often legitimately be made because of known theoretical relations or correlations which have been established from statistics. For example, it is certainly possible and permissible to obtain wind vectors in the free atmosphere from a set of upper-level pressure charts. To use shelter temperatures for an estimate of frost at the soil surface is less reliable, but still within the capabilities of the experienced meteorologist. In many areas it is possible to translate mean cloudiness data into frequencies of clear and cloudy days [53]. On the other hand, very little faith can be placed in an interpretation of precipitation data in terms of limitations to flying conditions (ceiling and visibilities), although that has been tried on occasion. In this field there is a great need for establishing correlations among various meteorological elements at one point and of spatial correlations between one and the same element.

In many practical cases the climatologist is confronted with the question: How long a period of records is needed to give an answer to a problem within an appropriate safety factor? There is no universal rule. In the past there has been a tendency to require decades of records. Many times this proves to be unnecessary in practice. Often suitable statistical techniques, evolved from the theory of sampling, can give quite satisfactory results on the basis of a small universe. The question reduces to this: When does a frequency distribution become essentially stable, so that adding another year of observations would not add significantly to the result?

Experience has shown that this time limit varies from element to element, season to season, and region to region. Preliminary investigations have shown that the orders of magnitude given in Table II may serve as a guide.

The data given in Table II are very tentative. They are based on a few pilot tests [5]. The studies leading to this type of information should be expanded to verify the statements and to cover additional elements.

We frequently encounter the following situation: A short record is available from a locality which is of interest. A long-record station is also located in the same region. The problem is to reduce the short to the long record. This question is dealt with in many stand-

ard climatological texts [19]. The procedure is reasonably straightforward for stations in proximity. But how does the reliability decrease with distance? Further, how reliable are the techniques of constant differences

Table II. Approximate Number of Years Needed to Obtain a Stable Frequency Distribution

Climatic element	Islands	Shore	Plains	Mountains
	Extratropical regions			
Temperature	10	15	15	25
Humidity	3	6	5	10
Cloudiness	4	4	8	12
Visibility	5	5	5	8
Precipitation amounts	25	30	40	50
	Tropical regions			
Temperature	5	8	10	15
Humidity	1	2	3	6
Cloudiness	2	3	4	6
Visibility	3	3	4	6
Precipitation amounts	30	40	40	50

or constant ratios if the two stations have completely different topography or exposure? Finally, suppose the short-record station has temperature and precipitation records but the practical problem requires knowledge of humidity, and such observations are available from the long-record station. How far can comparative techniques be applied? In the past, investigators have usually relied on the synoptic technique but have bypassed statistical procedures. This leaves open a fruitful field for future research.

A strikingly parallel question arises when there are no records of any kind available for the point at which information is desired. Major undertakings warrant the establishment of a temporary climatic station. The data thus collected can serve for comparison. How long should such an auxiliary record be maintained? In case of microclimatic investigations a few clear nights may suffice in order to establish, for example, the spots in an orchard which are most likely to be affected by frost. But the problem has never been studied on a broad scale.

Closely related is the general climatic question of the density of station networks. Should the weather services maintain for climatological purposes only a few *climatic bench marks* and shift other stations at frequent intervals in order to get a quicker coverage of a large territory? And what is the optimal distance for records of the different elements? [15, 74].

There is a need for the development of better statistical techniques specifically adapted to climatological problems. The classical methods developed for genetics or quality control may cover some problems, but satisfactory as they may be for the original purpose for which they were designed, they are not necessarily universally useful. For one thing, the basic assumption of *independence of observations* underlying most of them is almost never fulfilled for meteorological events. Interdependence in time and space, such as persistence, is almost always present in our data. Some work along this line has been done, especially as applied to singularities [6, 7, 12]. Many statistical tests are valid only for the so-called *normal* frequency distributions. Quite a few climatic values are not normally distributed and even mathematical transformations cannot always produce statistical normality. More adequate techniques are gradually being developed. Attention has been given to theories of extreme values [39] and to the multimodal distributions often found in climatological work [21]. There is a great need for the formulation of new criteria of the *significance* applicable to climatic data. A great deal of the modern statistical theory of time series has not yet found entrance into climatological work [65].

Classes of Problems in Applied Climatology

There are four major categories of problems in applied climatology. The major objectives of these are in the fields of (1) designs and specifications, (2) location and operation of a facility or equipment, (3) planning of an operation, and (4) relations between climatic and biological processes.

In the first class of problems we are usually confronted with a question about most frequent or extreme conditions. Sometimes the whole frequency spectrum is involved. Almost always long series of observations are required. In practice, wherever a good series of data is available, this group comprises the simplest type of problem. Even so, the climatic factor is usually very important because it often determines a major capital investment and the end product has to endure for a considerable length of time.

The second class of problems, dealing with location and operation of a facility or equipment, requires generally a greater degree of sophistication in the climatological analysis. The practical question is usually the choice of optimal conditions among several possibilities. The climatic factors are normally only one group out of many others which have equal or greater importance. The analysis will therefore have to illuminate the various advantages and disadvantages which arise from the climatic environment. Frequency distributions, extremes, interrelations of various climatic factors, and their relative importance must be presented to the client.

At the present time perhaps the greatest demand upon the climatologist is for an interpretation of climatic factors that enter into the third class of problems, the planning of an operation. This demand grew out of wartime experiences when weather conditions affected every military operation. In the absence of reliable long-range forecasts the chances had to be assessed on the basis of climatic risks. The peacetime applications are as numerous as the military ones. The basic climatic problems of strategic bombing, for example, are not much different from those of long-distance commercial airline operations; the climatic conditions affecting the logistics for maintaining an army in the field are essentially the same as the industrial distribution problems of peacetime; the weather influences on soil trafficability for military

vehicles are those also encountered in large-scale farming or construction jobs which require tractors, graders, bulldozers, and plows.

In all cases the climatic information is needed for decisions on *where*, *when*, and *how* to perform the job best. In simple terms, the climatological problem is one of giving odds based on past performance of the climatic factors. These odds cannot be given as a single value. One reason is that, in spite of many attempts, the climate cannot be reduced to a single figure. It is always a mixture of elements, some favorable, some unfavorable for the operation. Another reason is the natural variability of climate. The various elements always cover a range. A pessimistic planner will prefer to figure on the worst, an optimistic planner will gamble on the best, the conservative will play the middle ground of the *most likely* event. The prudent planner will want to know *all* the various contingencies.

The climatological analysis for operational planning purposes will almost invariably turn out to be a composite frequency analysis based on the synoptic material. Mean-value climatology is most out of place in this field, even though still used by some analysts. The procedures of synoptic climatology are by far superior [45].

The knottiest problems are encountered in the last major class listed above, in which the aim is to relate climatic factors to biological questions. The requirement is to establish the degree of influence between two very involved complexes of variables. We know, for example, that plant growth and crop yields are dependent upon complex accumulations of temperature, intensity of radiation in certain spectral regions, and soil moisture derived from precipitation. The degree to which each of these climatic elements influences growth and yield is variable within the life cycle of the particular plant. There are various critical times when minimum and optimum requirements exist for heat, light, and water. The climatic data as recorded are not immediately useful. They have to be transposed into intensity-duration frequencies with variable weights. Often only statistical stratification of past records will make it possible to determine the degree of dependency, and multiple correlations will be called for. Moreover, each species of plant has its own peculiar optimal and marginal climatic environment. The studies of this phase of the ecology, while of tremendous practical importance, are at present still in a primitive stage. Most of the time, only the standard meteorological mean values of temperature and total precipitation have been used for analysis. Recently there have been some refinements with the use of cumulative degree days [85], but not nearly enough has been accomplished. The fallacy of using ordinary weather-station data, sometimes from miles distant, has been pointed out [88], but much remains to be done.

The problem of agricultural climatology becomes even more complicated because of the secondary influences of weather conditions upon plant pests and diseases. Their growth and spread are dependent upon environmental factors favorable to them.[1] These nearly always include temperature, humidity, and wind. This second ecological complex is superimposed upon the one directly influencing the plant. It becomes most important for the planning of spraying and dusting operations. In turn, the chemicals and aerosols involved in these protective operations are subjected to the weather elements. The economy and effectiveness of the procedures for fighting fungi, insects, and bacteria are often directly dependent upon such factors as temperature, atmospheric turbulence, and precipitation intensity.

The study of climatic factors in relation to animal life has been a rather neglected field. Some work has been done on livestock [30] and a few investigations have been directed toward weather problems in relation to insects [86].

A whole science of its own is concerned with the influence of climate on human comfort and health. A great deal of information has been accumulated on the reaction of the healthy human body to heat stress. Most of the pertinent physiological studies have been carried on in the laboratory and specially designed climatic chambers. Investigations have been made concerning the influence of several variables, separately and jointly: pressure, temperature, humidity, air movement, radiation [1, 22, 49, 50]. Some instrumentation, supposedly simulating the reaction of man to the total environmental stress, has been designed. This started with the simple katathermometer of Hill [42], became more refined in the Davos frigorimeter [78], and developed into the present frigorimeter of Büttner and Pfleiderer [17, pp. 100–103] and the construction of a *copper man* [32]. These devices record heat requirements and overheating. Some notable field work has also been done, especially in the tropics [23]. A great many devices have been introduced to express the climatic stress in terms of readily measured and available parameters. The most widespread use has been made of the so-called *cooling power*, essentially based on dry-bulb or wet-bulb temperatures and wind speed. Skin temperature, effective temperature, and dew point have been used to assess the climatic heat and cold stress [17, pp. 90–95, 122–126].

It was realized that the reaction of a test individual could not reflect the wide variety of human responses, therefore group tests have also been made. They have covered comfort conditions [4], efficiency of workers [44], mental activity and output, and problems of human reproduction [64]. Admittedly, some of these studies could stand scrutinizing repetition to ascertain the validity of the results claimed. Even so, the conclusions have been used for practical purposes, such as assessing the suitability of various regions of the world for settlement [70]. For this purpose, the existing climatological data had to be translated into the physiologically effective factors established by experimental work. As the need becomes greater for space to ac-

1. Consult "Aerobiology" by W. C. Jacobs, pp. 1103–1111 in this Compendium.

commodate the ever-increasing population of the world, this type of investigation remains one of the most pressing tasks of applied climatology.

Perhaps the greatest challenge to the ingenuity of climatologists, however, is the problem of using climatic data in the field of human pathology—the question of the influence of climate on disease. There is little doubt that the climatic environment contributes to the spread of certain diseases and can also help in the treatment of diseases. The simplest of the many problems in this field is the selection of health resorts for the fatigued yet not completely diseased body. The results of much of the work discussed in the preceding paragraph have led to the establishment of climatic requirements for resort places. No final agreement has been reached on what constitutes a climatic resort, although good proposals have been advanced [61]. Climatic prerequisites for resorts for persons afflicted by various diseases which respond to climatotherapy raise even more complicated questions. Beneficial and adverse environments have been cited for a large group of pathological conditions. Among them are the respiratory ailments such as bronchitis, asthma, tuberculosis, and diseases of the circulatory system. As long as indications and contra-indications are not agreed upon by the physicians, the climatologist can only occasionally lend a helping hand, although some exceedingly fine climatological studies have been prepared for climatotherapeutic purposes [25–27, 52].

Even though the health of the population in general is improving and better drugs and other therapeutical devices, including climatic chambers in hospitals, are rapidly being developed, the problem will remain alive. Progress in health conditions has led to a spectacular increase of old people. The aged body cannot compensate for the climatic stresses as well as the young one can. Gerontologists have repeatedly pointed out that life expectancy increases in areas with favorable climates. What constitutes a *favorable* climate in this sense is at present ill-defined, although certain regions of the world have literally developed into *retirement eldorados.*

Climatic influences on the incidence of certain diseases have been claimed ever since Hippocrates. Modern medical science has accepted some of these claims and has rejected others. But in many cases it has not applied the refined tools of modern physical science to establish undisputed facts. The basic idea of climatic influence on many diseases is generally taken for granted; the details of the mechanism of this influence are ill-understood, matters of opinion, or simply ignored. It is uncertain which diseases are influenced by climatic conditions and which are the specific climatic elements involved in each case. It is further uncertain whether these influences are mainly on the causative agents, bacteria and virus, or on the disease vectors, insects and other carriers, or on the human body. It is quite likely that a concerted effort by teams of physicians, statisticians, and climatologists could solve many of the riddles in this field. (Pertinent literature is extensive. See [3, 13, 67, 73].)

Techniques of Applied Climatology

In the solution of applied climatological problems the technique will vary with the degree of complexity. Three major parameters enter into every problem:

1. Climate, composed of the various weather elements.
2. Space, consisting of the surface and upper layers of the atmosphere.
3. Time, comprising the series and sequence of weather observations.

Each of these parameters can appear in the problem either as a simple or as a complex element. In other words, the climatic factor can involve one weather element or many. Space can mean a single point, several points, or an area. Time may enter only in the restricted sense that the observations cover an interval of time or the problem may pose specific time limitations in terms of cumulative effects or conditioning of subsequent events by preceding situations. These considerations lead to various combinations of the three parameters (Table III).

Table III. Combination of Parameters in Climatic Problems

Space	Time	Climate
Single point	Simple series	Single element Multiple element
	Complex relation	Single element Multiple element
Multiple point or area	Simple series	Single element Multiple element
	Complex relation	Single element Multiple element

Each of the combinations listed in Table III can be illustrated by examples. These can serve to elucidate the problem and show what techniques have been employed to furnish the answers to practical questions. For some of the combinations many examples are available, for others only a few have so far been investigated. Some of the solutions are adequate, others are only makeshift.

Single Point—Simple Time Series—Single Climatic Element

The classical problem in this category is that of insurance against an adverse weather factor. The most common weather risks for which insurance protection is offered are hail, rain, and violent windstorms.

According to Roth [72], hail insurance on crops has been written in Europe for over a century and in the United States since 1880. Insurance rates must be based on the frequency and severity of hailstorms. Data on these storms during the growing season determine the risk. At present, risks are calculated for

counties or townships. Records are based on Weather Bureau observations and claims against the various insurance companies. There is a considerable variation in occurrence of damaging hailstorms from locality to locality. These variations are quite startling. Roth reports that the probability of hail damage varies so much that in extreme eastern Nebraska the insurance rate is 3 per cent of the amount insured, while in western Nebraska 18 per cent is charged.

Rain insurance is generally written for the protection of outdoor events against losses. These include carnivals, fairs, and sport contests. The insurance is paid when rainfall at the particular locality exceeds a minimum amount on the specified days or within the specified hours. This amount is often fixed at 0.1 in. It is easy to calculate the risk for such an occurrence from the extensive available rainfall statistics. At present, the insurance companies use weekly rainfall statistics. The frequency of occurrence of amounts above the minimum during the period of record is established and the risk is computed. Many decades of records are needed for this purpose.

Windstorms are similarly treated. In that case the damaging limit has to be established. Normally, wind speeds of less than 30 mph do not cause appreciable damage. However, it is not quite as easy to establish the frequency of windstorms above this or a similar limit as it is to establish the frequency with which rainfall exceeds a certain amount. At many stations winds are merely estimated according to the Beaufort scale. Not until recent years has there been a widespread network of stations equipped with recording anemometers. Even so, older instruments (such as the U. S. Weather Bureau triple register) are not very useful for determining peak gusts, because the instruments integrate the wind velocity over a period of time, but the peak gusts are the ones that cause the damage.

In this last case, the climatologist has to use his meteorological judgment to supplement the records. Where upper-air data are available, atmospheric-stability criteria can serve as useful guides to estimate gustiness. The frequency of incidence of hurricanes and tornadoes is also helpful. Yet the statistics are quite incomplete and we may find that one station which was affected by a severe storm has recorded winds in excess of gale force, while another close by has not, during the period of record, experienced such an unusual event. Adjustments of the records by analogies have then to be made in order to arrive at an appropriate risk factor.

In areas with sparse records this procedure may also have to be applied to other elements for calculation of insurance risks. Long-record stations will give the climatologist the probable shape of the frequency distribution. This shape, but not the absolute values, can then be used for the adjustment of shorter records at nearby stations.

In all these cases it is tacitly assumed that the past performance of climate is likely to continue essentially in the same fashion in the future. By and large this condition is fulfilled, but risks. once established, should be recalculated whenever a sufficient number of new observations become available.

Single Point—Simple Time Series—Multiple Climatic Element

Examples *par excellence* for this category of problems can be found in the field of design and specifications for protection of man against his environment. Clothing and housing fall into this group. When we build a house we want it to compensate for the exigencies of climate so that conditions indoors approach the zone of comfort as much of the time as possible. This aim governs all design factors. Nearly all climatic elements have to be considered for the purpose. In a preliminary fashion, analyses of the effects of various weather factors upon architecture and construction have been made [38, 77].

Table IV. Climatic Factors in Housing Design*

Climatic factors		Site, orientation, planting	Interior layout	Roof and walls	Openings	Foundation and basement	Mechanical equipment (for house)
Thermal heat	Temperature frequencies	X	X	X	X	X	X
	Frequency of hot and cold days	X					X
	Degree days		X				X
Radiation	Sunshine hours	X			X		X
	Clear and cloudy days		X	X		X	X
	Solar intensity	X	X	X	X		
	Solar height		X	X	X		
Winds	Wind direction	X	X	X	X		X
	Wind speed	X	X		X		X
	Strong winds			X			
Atmospheric moisture	Precipitation	X	X	X	X		X
	Snowfall	X		X	X		
	Excess precipitation	X		X	X	X	X
	Rainy days	X	X		X		
	Fogs				X		
	Thunderstorms						X
	Humidity		X	X	X	X	X

* Crosses (X) are entered for those elements which are of some importance for design of the particular feature.

Without going into the details of the reasoning involved, we can give a schematic picture of the relation of various weather factors to diverse phases of housing construction (Table IV).

The design of the individual features of the house will have to be geared to the climatic conditions. Houses are

built to withstand the elements for half-a-century or more. Hence the most probable values and the extremes of climatic elements must be known. As almost all climatic elements are involved, there will have to be compromises. For example, a roof design best for good drainage of rainfall and for avoiding excessive accumulations of snow may result in undesirably high absorption of solar radiation in summer. But at least the designer should have the basic information on climate available, so that he can plan for optimal results.

Even though the side-by-side data on all the climatic elements are valuable, they are by no means the final solution for problems of housing climatology. This is best illustrated by the specific case of heating requirements. Insulation and heating plants are commonly planned on the basis of *heating degree days*.[2] This factor is a derivative of temperature alone but is supposed to be an index for the heat loss of a house. Obviously, there is correlation between the temperature gradient "indoors-outdoors" and the heating required. Yet the degree-day index completely neglects the radiation and wind factors. The latter is quite important. Hence the design engineer would prefer to get from the climatologist a suitably combined factor of temperature and wind speed, reflecting the total cooling power of the atmosphere. Such joint relationships have been elaborated by the use of test houses [43]. However, this factor is neither directly measured nor generally used in design, because it is not usually known to both architects and meteorologists. This factor can be calculated approximately from simultaneous observations of temperature and wind speeds. But wind observations obtained on airport towers or other high structures are not directly applicable. They have to be reduced to the height of the building for which the design values are desired. For this purpose one can safely assume a logarithmic distribution of wind speed with height. Somewhat more refined aerodynamic procedures are also available in case very high accuracy is needed [69]. These calculations should not be applied to *mean* values of temperature and wind because in all practical cases the frequency distribution of cooling power is actually needed and, furthermore, at many stations there exist peculiar interrelations of temperature and wind speeds. At present these calculations are somewhat tedious, but they could easily be reduced to a minimum because the pertinent equations are readily adaptable to nomographic solutions.

We run into an analogous problem when evaluating climatic data for clothing design or issue. Several aspects of this last problem have recently been discussed in the literature [9, 18, 22, 60, 82]. The dry atmospheric cooling power in the shade as a function of wind speed and temperature has been measured as well as calculated. Among the measuring devices, the frigorigraph of Büttner and Pfleiderer has undergone rigid tests [48] and proved to be a rather satisfactory instrument for climatological purposes. Very few stations have this, or even more primitive types of equipment such as the katathermometer, and series of records are scarce. Therefore, for some time to come, climatic problems in clothing design have to be solved by means of the empirical formulas developed by Hill, Büttner, Plummer, Siple, and others. These are fairly satisfactory for the dry cooling power and they are reduced to nomographic form [60, p. 202]. These formulas, however, cover only part of the heat or cold stress problem. Radiation influence is neglected and evaporative cooling from sweating is not included. Formulas considering these additional terms are much less reliable for conversion of climatological data into values applicable to clothing problems. The experimental bodies used in establishing the relations give only a very poor first-order approximation of human physiological reactions. The real physiological process includes such complicated functions as the effect of body position on radiative heat exchange and the effect of rate of activity on sweating and hence evaporative cooling. These factors cannot be reduced readily to universally applicable quantitative terms that can be translated into the simple climatic elements which are observed at many stations. A further complication is that no *standard* human being exists. Individual reactions to environment cover a broad band. Studies on various types of individuals, registering their subjective feeling of comfort, will have to be made in order to provide the link for the proper interpretation of climatic data. A start in this direction is contained in the formulas proposed by Burton [16]. These are designed to give the effect of clothing on heat loss from the surface of the human body. It has been pointed out that these formulas apply only to equilibrium conditions. They make no allowance for convection inside the clothing, nor are the empirical constants of the equations well determined. From theoretical considerations Lee has arrived at some qualitative criteria for the design of tropical clothing. The complexity of clothing design can be grasped from a list of the climatic conditions and the clothing factors (Table V).

Table V. Climatic and Clothing Factors in Tropical Clothing Design

Climatic factors.....	Solar radiation, radiation from hot objects, hot dry air, hot humid air, tropical rain
Clothing factors.....	Color, texture, thickness, permeability for water vapor, weight, extent of body covering, openings, fit, water repellency, underwear

Quantitative solutions for five to ten variables, as listed in Table V, are not yet within reach. As in the case of housing design, microclimatic problems are encountered; for example, the reduction of shelter-measured temperatures to various body heights and the application of wind data from the usual meteorological exposures to "surface-near" data. The difficulty of ob-

2. Heating degree-days per day = 65F minus actually observed mean temperature for the day when below 65F.

Heating degree-days per month or year = cumulative values of degree-days per day.

taining satisfactory climatic information for the design of shoes from these conventional observations needs no further elaboration [10].

Single Point—Complex Time Relation—Single Climatic Element

In most existing works on problems of climatology the whole series of climatological observations has been used. All observations are accorded equal weight. The time factor enters only as length of the series or as a periodic variation, such as diurnal or annual changes. Many problems, however, include the duration of a climatic condition, or duration above or below limiting values, or intensities per unit time. These complex time relations present difficulties. Usually one difficulty is the lack of continuous records of the climatic element.

A very simple example in this group is the design of a heating system for the melting of snow falling on a pavement. For this purpose, the rates of fall of snow (or sleet) must be known. We must construct amount-frequency-duration curves for the particular locality (generally only for temperatures below the freezing point). Complications are encountered in establishing rates and durations of snowfalls in the absence of heated recording precipitation gages. A first approximation can be arrived at from cumulative 12-hr or 24-hr precipitation measurements, temperature data, and hourly or three-hourly reports of the synoptic network.

A rather simple example of the type of problem associated with this class is afforded by the evaluation *in situ* of the "frost hazard" to tender fruits and vegetables. For each fruit or plant (at a given stage of development) there exists a minimum temperature which constitutes an upper limit for damage through freezing. Freezing damage to the fruit, blossom, or plant will not occur at temperatures above this level. Nevertheless, experience and experiment have shown that freezing damage is a function of both the actual minimum temperature and the duration of temperatures below the critical limit for freezing. For example, a fifty per cent commercial damage to a particular crop can result from a minimum temperature of 28F with a duration of four hours, whereas no more than fifteen per cent damage might result from the situation in which the minimum temperature is 26F for a duration of thirty minutes, below 27F for forty-five minutes, and below 28F for one hour. For this reason, attempts to correlate frost damage and the frequencies of minimum temperatures have seldom been successful, although excellent results have been obtained when the temperature-duration data have been substituted [89].

In evaluating the probability of frost damage, *time* becomes a further complex to the extent that the dates of susceptibility of the plant or crop to freezing may vary widely from year to year. For example, in the case of deciduous fruit crops, the most susceptible period for freezing damage is during the blossom (or small green fruit) stage of development. However, the phenological data from a single orchard between earliest and latest dates of blossoming may show an extreme range of five weeks. Obviously, the probability of a complete or partial loss of crop through frost damage is much greater in an early-bloom year than when the blossoming period occurs so late in the season that there exists a small probability that freezing temperatures will occur. A somewhat similar problem exists in the evaluation of the frost hazard during autumn to the seed corn crop at points in the Middle West [79-81]. In the latter case the probability of frost damage to the crop is increased during a season of late seed maturity. In neither case will a simple frequency distribution of the occurrences of critical temperature durations according to calendar dates give a true measure of the probability of frost damage to the crop. In all such agricultural problems, the reference base is the "crop calendar," not the Gregorian calendar.

Another intensity-duration problem of a single element at a single locality concerns solar radiation as related to the design of window surfaces, blinds, and awnings. Their practical purpose may be either the use of solar energy for supplemental heating, or the elimination of excessive radiation. Theoretically, the problem is readily solved. Practical difficulties arise from the lack of radiation measurements, especially on exposures other than the horizontal surface [35, 40].

Another example which illustrates this category of problems deals with the exploitation of wind for power production. In contrast to other possible examples we are favored in this instance by the existence of an excellently documented case history [71]. A brief summary of the meteorological phases of the New England experiment for engineering exploitation of the wind is given below. In it we follow Putnam's conclusions closely.

There are many basic questions concerning wind behavior that have to be answered in the selection of a site for a wind power plant. In the specific case analyzed here, the choice of a suitable locality for a test site was an important problem. We can pass over the restricting operational requirements, such as accessibility and proximity to the center of maximum electric load, although in any practical case these limit the choice between all meteorologically favorable sites.

The basic ingredients for a preliminary climatic evaluation are:

1. Free-air wind speed at mountain-top height.
2. Effect of the geometry of a mountain on retardation or speed-up of wind flow over the summit.
3. Prevailing wind direction.
4. Influence of the turbulence structure of the wind on design.
5. Influence of the wind structure on estimates of output.
6. Influence of atmospheric density on estimates of output.
7. Influence of estimates of icing at higher elevation on design.

Factor 6 can be derived from theoretical considerations with sufficient accuracy; factor 7 is of importance

in some regions only. We will therefore omit further discussion of these factors and concentrate on the other five factors, all of which relate directly to the wind.

The energy of the wind varies with the third power of the speed. Hence, speed is the most important single consideration. The preliminary climatological investigation must proceed from the available anemometer data, the pilot-balloon information, the known relations of gradient winds to pressure distribution, and the general change of the wind vector with height. In most areas this will yield a first approximation on strength and constancy of winds at possible power sites. The lack of direct meteorological observations can often be supplemented by ecological studies in the area. Strong, steady winds will produce deformations in trees, which act as suitable integrators of the wind vector.

Factor 2 mentioned above will, after preliminary selection of possible sites, have to be supplied by wind-tunnel tests on models of the particular terrain. The local structure of the wind, such as gustiness, cannot be predicted with a great deal of confidence. It depends on local topography and roughness. A series of observations at various heights in the low layers above the ground will be necessary. Such special observations do not have to extend over a long period of time. Sets of measurements under a variety of conditions of stability can give enough information. The data thus collected are not only of value for the mechanical design of the wind turbine and its support, but are also essential for the choice of the best height of the windmill above the ground at the site.

The frequency distribution of hourly wind speeds must be known. The New England study has revealed several points that are universally useful. The speed-frequency curves are Pearson Type III curves with sharp modes, skewed toward lower speeds. Such a curve can be approximated quickly from a number of fixes. The following relations hold (at least for the New England area):

1. A small percentage of hours are calm.
2. The sides of the frequency curve are concave upward.
3. The skewness increases with the mean speed.
4. The most-frequent speed is lower than the mean speed, but increases with higher mean speeds.
5. The number of hours during which the wind blows at the most-frequent speed decreases as the mean speed increases.

Similar general relationships can undoubtedly be developed for other environments.

The mean velocity at a possible site can be established from a short record by reduction to a nearby long-record station, since the ratios of simultaneous values stay essentially constant. The study also established the requirements for the length of simultaneous records needed in order to establish this ratio of mean speeds within specified limits of error. For a 10 per cent error limitation, the lengths of record listed in Table VI were found.

Of further importance are the periodic variations (seasonal and diurnal changes) and the reliability of wind power (maximum positive and negative departures from the mean). These will require long records, but a nearby station (within about a fifty-mile radius and with not too different an exposure) can be used for this purpose. This means in essence that short-record sta-

Table VI. Length of Record Needed to Reduce Short-Record Station to Long-Record Station

Horizontal distance (mi)	Difference in elevation (ft)	Length of time (days)
38	+2000	90
12	+2000	30
10	+400	10
Same locality	+65	0.6

tions, operating for about three months, will yield the most essential basic data for a climatological analysis of this type of case. The low-layer vertical distribution at the auxiliary stations should also be established by anemometers at three or four levels between the surface and two hundred feet elevation. If possible, the three months of records should not be continuous but should be composed of several periods of two or three weeks each, spread over the seasons.

Single Point—Complex Time Relation—Multiple Climatic Element

Instead of establishing classes, the climatic influences can often be expressed by a formula. This scheme has been employed for determining deterioration of materials under weather influences. According to C. E. P. Brooks [14], the following elements enter into the problem:

1. Effects due to high temperatures alone.
   - *a.* Mechanical effects.
   - *b.* Speed-up of chemical reactions by heat.
   - *c.* Flowing due to decreased viscosity.
   - *d.* Increased activity of bacteria.
2. Effects due to high temperature in conjunction with moisture.
   - *a.* Corrosion of exposed surfaces.
   - *b.* Effect of diurnal cycle—*breathing* of partly sealed packages.
   - *c.* Organic changes—molds and rotting.
3. Effects due to impurities in the air.

The combined effects of temperature, humidity, atmospheric impurities, and wind can be stated in a general equation of the form

$$A = ax^t + b\frac{(H - K)}{100}(1.054)^t(1 + cI)(1 + 0.067W),$$

where A is the rate of deterioration of a fresh sample, t the temperature in degrees centigrade at the surface of deterioration, H the relative humidity at the temperature t, I the concentration of effective impurities, bacteria, fungi, or spores, and W the effective wind speed in mph. The quantities a, x, b, c, K are constants which have to be determined in each case. The complex

time relation enters into the rates of deterioration because of variable diurnal and seasonal durations of the effective factors, which often have upper and lower limits.

The analysis by Brooks covers only a limited group of cases. For example, deterioration of rubber, modifications of glass by certain ultraviolet wave lengths, and the weathering of paint by light and rainfall coupled with high temperatures have not been considered. In this field much work remains to be done before climatic data can be applied intelligently to the many practical problems of *weathering*.

An example of the lack of weather planning of this type is the recent case in which importers of a particular type of French silk print accused the European exporters of dumping an inferior product on the American market because the silks did not hold up as they had done in Paris. The silk prints, long-lived in a damp, cloudy atmosphere, literally went to pieces in a few weeks under the influence of the more frequent and intense direct sunlight found in New York, Chicago, and other American cities. The prior determination of the comparative sunshine (solar radiation) conditions would not have been a difficult task for the climatologist [46].

Multiple Point or Areal Problem—Simple Time Series —Single Climatic Element

The prime examples in this category are to be found among the large number of problems which are concerned with the marketing, on a state-wide or national scale, of the so-called "weather goods." Weather goods are referred to by those in the trade as goods whose use and sale are primarily determined by nonperiodic changes in such weather elements as temperature or precipitation. The national distributing system, as it serves the retailer, is more closely adjusted to the distribution of population than to the synoptic weather probabilities. As a result, it happens only too often that the bulk of the saleable goods are found where there are the most people but not where the actual demand is greatest.

There are of the order of one hundred cities in the United States with populations of 100,000 and over and these cities account for well over a third of the country's total population. Yet, how many climatologists (or distributors) have a knowledge of the number and location of cities that are likely to be affected *simultaneously* by the same abnormal weather conditions? A knowledge of the geographical extent and relative locations of areas which may be affected *simultaneously* by abnormally high temperatures is extremely important in planning the marketing and distribution facilities for such "hot-weather items" as lemons or soft drinks. If a study of synoptic weather data reveals that the odds are 5 to 1 that a summer "heat wave" which affects cities in the East with fifteen million people will also affect areas in the Middle West or South with a population of ten million people, then it is obviously poor planning to have 75 per cent of the product warehoused in New York and only 10 per cent in Chicago [28].

Another example chosen to illustrate this category of problem involves the evaluation of the possible increased demand for electric power that may result when heavily populated but dispersed urban areas, which are serviced largely by a common interlocking power facility, are affected simultaneously by abnormal and sudden decreases in the natural illumination. Such a condition can be caused by a sudden development of widespread thunderstorm activity, by a fortuitous grouping of individual thunderstorms or by the unusually rapid development of a heavy cloud or fog layer over the area. It is doubtful if such information can be obtained by inspection of data available from the usual station network. A synoptic-analytic approach appears to be the most promising method for the solution of such a problem at the moment, although there is the possibility that a statistical analysis of the data acquired by special thunderstorm-research projects might give quantitative information on the probable spacing of thunderstorms or on the areal extent of thunderstorm activity as related to the spacing (or areal extent) of the service areas. However, actual observational material may subsequently become available through radar storm-detection programs or, perhaps, through special thunderstorm-observing programs similar to those sponsored by the Amateur Weathermen of America.

Multiple Point or Areal Problem—Simple Time Series —Multiple Climatic Element

This group of problems is peculiarly the province of the synoptic-climatic technique of approach. An example is the planning for the aerial mapping of large geographic areas or boundaries. The technique of aerial photographic mapping requires that given strips of the earth's surface of the order of 100 or more miles in length be free of clouds at all levels below the flight altitude; that illumination conditions (as determined by solar altitude, haze, and upper-cloud conditions) be favorable; that there be no turbulence at flight altitudes; and that the surface to be photographed be free of snow cover and not otherwise obscured. The element of time is not a complex, because the photographic flight strips are run in a matter of minutes or hours. The possibility or probability of occurrence of the proper combination of meteorological circumstances over the area cannot be obtained from frequency data at a single station alone. However, the analysis of the several meteorological factors as they occur simultaneously over large areas will reveal whether or not the aerial photographic technique is feasible, determine the favorable period for mapping, and may also serve to indicate the most economical photographic flight plan (flight altitude, order, and orientation of photo strips).

Another application could perhaps be cited with equal validity under the types discussed in the next section. It is the use of the synoptic-climatic method for weather forecasting. This procedure is now frequently referred to as *objective forecasting*.

The use of climatological information for forecasting is really nothing new. The recognition of certain persistencies in series of observations led to some early forecasting rules. Other autocorrelations in time-series of meteorological observations at one point have been discussed repeatedly as aids to forecasting [36, 54].

The modern approach substitutes an areal net of observations for local time series. From series of synoptic charts, stratified or multiple correlations of meteorological parameters at distant stations are used as *predictors* for the locality for which a forecast is desired. Frequency distributions of antecedent and subsequent events are obtained and stochastical relations are worked out so that the probability of a given event (in dependence upon previous events) can be obtained. Machine tabulations have facilitated this approach. Subjective forecasting experience, which in many cases is nothing but a subconscious system of statistics, is replaced by an objective summarization of the past performance of weather. While this is a potent tool, it should not be overlooked that it is essentially a static extrapolation technique and disregards dynamic and energetic factors which can influence the weather in the interval for which the time-lag relations are worked out. The procedures have been adequately presented in several recent papers to which reference only is made here [91]. Much useful work in this direction remains to be done.

Multiple Point or Areal Problem—Complex Time Relation—Single Climatic Element

Problems of this class are most frequently encountered among the large number of questions concerned with the probable maximum runoff of flood waters as determined by the areal patterns of intensity-durations of rainfall. Such information is essential in the evaluation of possible peak floods or in the design of storage or flood-control dams, levees, culverts, storm sewers, and the like. The hydrometeorological literature abounds with examples of the employment of the technique which, in its best form, makes use of the detailed precipitation intensity-duration data accumulated from the records of a large number of precipitation gages (preferably recording gages) distributed over a given drainage area. Where such detailed data are not available, the use of theory is required to supplement the inadequacy of the observational record. The technique, itself, is so commonplace that there appears to be little requirement that it be further elaborated upon in this section. For further details the reader is referred to the extensive and readily available hydrologic literature on the subject [34, 76].

Another example of this class is afforded by the necessity for evaluating the frost hazard or assessing frost damage to crops as the problem applies to large agricultural areas of diverse topography. In an earlier section the evaluation of the frost hazard as it applies to a single point in an orchard or to a small agricultural plot was discussed. In the present case the identical temperature-duration data are required, but now we are concerned with the pattern of such conditions as they occur over a large area during a given time interval. For the rapid evaluation of the probable freezing damage from a single frost, the temperature-duration data from a large number of representative points within the agricultural area for the single frost night may be considered sufficient. For crop estimation purposes, however, the cumulative data for the several series of frost occurrences during the entire growing period are required.

The variations in minimum temperatures and durations that occur over a region are due largely to differences in (1) exposure to winds, (2) relative elevation (in the local sense), and (3) nature of the soil cover (whether wet or dry, cultivated or fallow). Of the three factors, only (2) remains constant throughout the growing season. The most exposed areas will present fewer frost nights and the duration of critical temperatures will be shorter than in more sheltered locations, due to the "mixing" effects of the frequent nocturnal winds. The temperature effect of the wind, however, will depend upon the degree of stability that has been attained in the lower layers of the atmosphere at the time the wind is initiated; the degree of exposure to winds will vary from night to night depending upon the direction of the pressure gradient during the frost period. It is doubtful if the pattern of damaging temperatures recorded for an area during a given frost period is ever other than *unique*.

For the purpose of rapidly evaluating the frost damage to the crop, the citrus industry in the Far West makes extensive use of temperature-duration data obtained from thermographs exposed in a large number (approximately 400) of representative citrus plantings. Through the mechanical or manual integration of the individual thermograms, the analysts can estimate the probable frost damage with surprising accuracy within a day or two after the frost occurrence. Similar estimates based on field sampling of the fruit would require weeks to accomplish.

Multiple Point or Areal Problem—Complex Time Relation—Multiple Climatic Element

As we expand the variables into all dimensions the applied climatological problems become very involved. At the same time well-analyzed examples become rare. The problem of agricultural land usage in relation to climate has probably had more attention than any other. In agriculture we are dealing with complicated questions that can be disentangled only by the cooperation of four sciences: soil science, plant biology, land management, and climatology.

But even with proper soil, proper tillage methods, proper fertilization, and proper seed material, the climatic conditions determine to a large extent what plants will or will not grow, mature, and ripen at a given locality. A most intricate system of time sequences of various climatic elements enters into the problem. We find that photosynthesis is dependent on radiation intensities in specific wave lengths. In some

cases the amount of sunshine received at special times in the growth cycle of the plant will determine the commercial value of the fruits (an example is the August-September influence of sunshine on the sugar content of grapes). The rapidity of development depends on cumulative temperatures above a limiting value, that is, a certain *degree-day* value. The occurrence during the growing period of a temperature below a critical value, such as the freezing point, can determine the partial or complete loss of a crop. Further, the proper water balance is of particular importance in plant development. There are times when too much water is as detrimental as is insufficient moisture at other times. The requirements within the life cycle of maximal and minimal amounts of water are very different from one plant species to the other. The water balance is not determined by a simple climatic element, such as rainfall, but is governed by precipitation, transpiration of the plant, and evaporation. This last element itself is a function of humidity, temperature, and wind speed. Thornthwaite has tried to establish some of these relations of *evapotranspiration* [83].

The literature abounds with studies of crop yields in relation to weather conditions [90*e*, pp. 293–476]. These relations can be used to establish facts about optimal growing conditions. In turn, these can serve to establish from long climatic records the risks to a given crop at various localities. A great deal depends here again on microclimatic conditions. This is particularly true of frost risks [47]. An area may, for example, be in general quite favorable for apple or peach orchards and yet some topographically poor exposures may be veritable *frost holes* and entirely unsuited for this type of cultivation.

Climatological principles have been applied to the introduction of plant seeds from one territory to another [66]. These studies of comparative climato-ecology are still in their infancy and amenable to much improvement. The difficulties involved have already been pointed out (see page 980) and allusion was made to the problem of climatic influences upon pests.

Experience has shown that a definitive *fire weather* exists, especially in forests. For effective fire-fighting procedures it is useful to estimate the *danger class* of each day and season. This will result in better dispositions for lookout posts, fire-fighting equipment, and personnel. The climatic information serves both for planning purposes and for the anticipation of fire-weather stations. These are later supplemented by special weather forecasts.

For many forest areas of the United States special *fire-danger meters* have been worked out. These are composite slide rules which have been described in the literature [41]. A case history for an urban area may, however, be of interest. It shows the technique of arriving at a fire-danger scale. This scale was derived for grass and rubbish fires in the city of Raleigh, N. C. It was assumed that the immediate causes for such fires, namely flying embers, carelessly dropped burning matches and cigarettes, etc., remain essentially constant so that weather conditions contribute most to the variability of incidence.

The climatic elements mainly entering into a classification of fire hazards are precipitation, relative humidity, and wind speed. Precipitation and relative humidity are important for the period preceding the fire, while the wind speed enters at the time of ignition. For establishing the scale, the conditioning and coexisting weather conditions were noted for 90 days on which three or more grass or rubbish fires were observed. It is not surprising that precipitation and humidity values immediately preceding the fires were more important than those further back in time. This required weighting factors for the sequences of these elements. These were simply obtained by multiplying the values on the *fire day* (three or more rubbish or grass fires) by 4, on the preceeding day by 3, on the second preceding day by 2, and by using the actually-observed values for the third day prior.

The statistics showed immediately that 70 per cent of the fire days had a relative-humidity factor of less than 40 per cent, while only 18 per cent of *all* days had a humidity factor of less than 40 per cent. Over 80 per cent of the fire days also had a weighted precipitation factor of 0.01 in. rainfall or less (of *all* days only 37 per cent have a precipitation factor of less than 0.01 in.). The simultaneous wind speed did not show quite as good a discrimination, although 88 per cent of the days with five or more fires had wind speeds of 13 mph or more. However, 75 per cent of *all* days have winds exceeding 12 mph at one hour or another.

These statistics resulted in the following scale:

	Weighted relative humidity %		Weighted precipitation *in.*		Wind speed *mph*
Good fire days	$\leqq 40$	and	$\leqq 0.01$	and	$\leqq 13$
Poor fire days	$\leqq 60$	and	$\leqq 0.05$		
	or $\leqq 70$				
	or		$\leqq 0.25$		

Days that did not fall into the classes *good* or *poor* were classified *indifferent*, that is, the weather would not contribute to the fire hazard nor would it definitely retard fires. The discrimination of the scale is shown for four test months (January–April = 120 days) for four years (total 480 days) (Table VII).

Table VII. Fire Occurrences and Fire Danger Class

Fire Class	Total days	Total fires	Fires per day
Good fire day	110	226	2.05
Indifferent fire day	224	193	0.86
Poor fire day	146	32	0.22

The table shows that the scale is quite satisfactory. The *good fire days* had over nine times as many fires as the *poor fire days*.

This group of problems is also peculiarly the province of the synoptic-climatic technique of approach. An example is the establishment of a time schedule for

flights optimally adapted to weather conditions. A flight operation will be the more economical the better the weather conditions are at the terminals and over the route. Poor terminal conditions, as expressed by weather minimums of combined ceilings and visibilities, prohibit or delay take-offs and landings. Hazardous conditions en route, such as severe icing and severe turbulence due to fronts and thunderstorms, may require detours or delays and hence are not conducive to the maintenance of a smooth flying schedule. A good flying day can be defined in terms of absence of such *limiting weather conditions* in sequence at the take-off point, en route, and at the landing terminal. This sequence of events cannot be established readily by spot observations. But it can be obtained on sequences of synoptic weather maps. Statistics of the annual variation of good and poor flying days, the change from year to year, at least in terms of largest and smallest number, will give a planner the weather factor for a decision on whether or not a specific air route can be operated profitably.

The identical technique was applied to logistic and other planning problems during World War II. Strategic bombardment operations were a case in question. The limiting weather conditions for take-off and landing in friendly territory, the route conditions, and the availability of targets for visual or blind bombing entered the picture. This application of synoptic-climatic techniques has been discussed in more detail by Jacobs [45, pp. 20–22].

This type of study for airline operations can be considerably refined by considering, for example, diurnal variations of limiting conditions for each hour at the terminals and en route. There is no reason to elaborate on this because every meteorologist knows that during certain times of the day in various seasons limiting conditions are particularly prevalent at certain airports. It is elementary to expect higher frequency of dense fogs in the hours before sunrise. Also the high summer frequency of fogs in areas like Newfoundland is notorious. The climatic analysis can place such knowledge on a systematic basis. The same applies to the seasonal variations of air trajectories. These can be obtained from the upper-level synoptic maps and used for laying out the most advantageous routes over great distances in accordance with the principles of *pressure pattern flying*.

Additional Unsolved Problems

The foregoing review of the state of the art of applied climatology had to point to a great variety of only partially solved problems. In fact, each new application will require its own specific solution. In many respects this part of the *Compendium of Meteorology* is a series of *confessions of ignorance*.

We have already pointed out in sufficient detail the need for new types of observations, many of which require new instrumental approaches. The requirement for statistical solutions better adapted to the specific problems has also been discussed. These are not the only deficiencies. The classical methods of summarization and graphical presentation need radical overhauling in many respects [55]. Some very remarkable advances of graphical and pictorial presentation of climatic data, growing out of wartime experiences, have been placed in the scientific record [45, pp. 44–51].

Yet we still lack adequate methods for multidimensional representation of data. We need more nomographic charts to simplify the presentation of complex relations and to allow for a quick solution of empirical equations.

We have indicated that, in contrast to classical climatology and climatography, applied climatology is a broad, two-way meeting ground between the synoptic meteorologist and the climatologist. A fruitful field for joint efforts lies in the methods of *air-mass* climatology. There has been insufficient space to discuss these in detail here. This approach, however, has shown increasing promise for problems of bioclimatology [24, 33, 56, 62]. It is likely that further work along this line can find useful applications.

Other advances in applied climatology are intimately tied to progress in *microclimatology*. The research needs in this field have recently been presented by Baum and Court [11]. Their very pertinent remarks can be summarized as follows: There is need for standardization of microclimatic procedures and equipment. More measurements are needed on the vertical distribution of the usual climatic elements in the boundary layer over various forms of terrain and vegetation. Coincidentally, measurements of the following factors should be made:[3]

1. Temperature gradient in the soil.
2. Soil conductivity as a variable quantity.
3. Insolation, direct and diffuse.
4. Atmospheric radiation.
5. Terrestrial radiation.
6. Eddy conductivity in the air.
7. Evaporation.

As basic problems Baum and Court propose studies and causal understanding of the heat balance at the surface of the earth, the existence and extent of laminar layers near the surface, and the interrelation of microclimatic factors and evaporation.

Lastly, it should again be stressed that applied climatology should be in each instance a cooperative venture. It deals with borderline fields. Climatology is on one side; on the other side are a great many other disciplines (medicine, ecology, agronomy, hydrology, pedology, architecture, strategy, business management, and various phases of engineering). The solutions to the many joint problems can best be found by teamwork.

Appendix

Table VIII lists areas of practical problems in applied climatology, by classes. Some of these areas have

3. Some of these have been mentioned in Table I, above, but are repeated to present Baum and Court's ideas completely.

been studied and reference is made to pertinent papers that have not already been quoted. The list is by no means complete, but it indicates what a broad and

TABLE VIII. LIST OF TOPICS IN APPLIED CLIMATOLOGY

Field of application	Class of problem	Type of analysis	Reference
Advertising and marketing	*C*	*a, b*	[28]
Aerial photography	*C*	*c*	[75]
Airfield construction	*A*	*b*	[45]
Airline operation	*C*	*f*	[63]
Air pollution control	*B, C*	*c*	
Architecture and housing construction	*A, B, C*	*d*	[58]
City planning	*B*	*b*	[51]
Clothing	*A*	*b*	[60]
Crop planning and protection	*C*	*g*	[89]
District heating	*C*	*b*	
Forest fires	*C*	*d*	
Gas and oil dispatching	*C*	*f*	
Health resorts	*D*	*f, g*	[61]
Heating and cooling plants	*A*	*b*	[59]
Highway construction	*B*	*f*	[57]
Hydroelectric power	*B*	*b*	
Insurance	*A*	*a*	[72]
Land utilization	*D*	*f, g*	[47]
Lubrication	*A, C*	*b*	
Military operations	*C*	*f*	[45]
Open air theater operation	*B, C*	*a*	
Power dispatching	*C*	*f*	[29]
Power lines	*A*	*f*	[20]
Shipping and transportation	*C*	*g*	[8]
Storage (food, oil, photographic material, etc.)	*A, B*	*b*	
Wind power	*A, B*	*c*	[71]

fruitful field lies before those climatologists who break away from the traditional archivistic or descriptive approach and who are willing to meet the challenge of new frontiers.

The following notation is used in Table VIII:

Class of Problem

A. Design and specifications.
B. Location and operation.
C. Planning of operation.
D. Relation of climate to biological problem.

Type of Analysis

Point	*Element*	*Time relation*
a. single	single	simple
b. single	multiple	simple
c. single	single	complex
d. single	multiple	complex
e. multiple	single	simple
f. multiple	multiple	simple
g. multiple	single	complex
h. multiple	multiple	complex

REFERENCES

1. ADOLPH, E. F., and MOLNAR, G. W., "Exchanges of Heat and Tolerances to Cold in Men Exposed to Outdoor Weather." *Amer. J. Physiol.*, 146: 507–537 (1946).
2. AIR WEATHER SERVICE, *Machine Methods of Weather Statistics*, 59 pp., Washington, D. C., 1948.
3. AMELUNG, W., *Klimatische Behandlung innerer Krankheiten*, Abhdlg. Geb. I, Bäder & Klimaheilkunde Nr. 4, 92 SS., Berlin, 1941.
4. AMERICAN SOCIETY OF HEATING AND VENTILATING ENGINEERS, "Physiological Principles" in *Heating Ventilating Air Conditioning Guide 1948*. Vol. 26. New York, Industrial Press, 1948. (Chap. 12, pp. 197–231)
5. ARMY AIR FORCES, HEADQUARTERS, *Study of Length of Record Needed to Obtain Satisfactory Climatic Summaries for Various Meteorological Elements*. Weather Information Branch, Rep. No. 588, Washington, D. C., Nov. 1943.
6. BARTELS, J., "Anschauliches über den statistischen Hintergrund der sogenannten Singularitäten im Jahresgang der Witternng." *Ann. Meteor.*, 1:106–127 (1948).
7. —— "Zufallszahlen für statistische Versuche." *Ann. Meteor.*, 1: 209–216 (1948).
8. BATES, C. C., and GLENN, A. H., "Meteorological Engineering Aids Highway Transportation." *Civ. Eng.*, 17: 606–609, 640 (1947).
9. BAUM, W. A., *The Climate of the Soldier, Pt. III, Vertical Temperature Distribution Surrounding the Soldier*. Environmental Protect. Sec. No. 124, 69 pp., Office of the QMG, Research and Development Branch, Jan. 1949.
10. —— "The Temperature at Your Feet." *Weatherwise*, 2:75–78 (1949).
11. —— and COURT, A., "Research Status and Needs in Microclimatology." *Trans. Amer. geophys. Un.*, 30:488–493 (1949).
12. BAUR, F., "Zur Frage der Echtheit der sogenannten Singularitäten im Jahresgang der Witterung." *Ann. Meteor.*, 1: 372–378 (1948).
13. BERG, H., *Einführung in die Bioklimatologie*. Bonn, H. Bouvier, 1947.
14. BROOKS, C. E. P., "Climate and the Deterioration of Materials." *Quart. J. R. meteor. Soc.*, 72: 87–97 (1946).
15. BROOKS, C. F., "Recommended Climatological Networks Based on the Representativeness of Climatic Stations for Different Elements." *Trans. Amer. geophys. Un.*, 28: 845–846 (1947).
16. BURTON, A. C., *An Analysis of the Physiological Effects of Clothing in Hot Environments*. Report to Aviat. Med. Council of Canada, C 2754 SPC 186, Nov. 24, 1944.
17. BÜTTNER, K., *Physikalische Bioklimatologie*. (*Probl. kosm. Phys.*, Bd. 18.) Leipzig, Akad. Verlagsges., 1938.
18. CAMPBELL, R. D., *Preparation of Clothing Almanacs*. Environmental Proteet. Section Report No. 144, 17 pp., Office of the QMG, Research and Development Branch, April 1, 1949.
19. CONRAD, V., *Methods in Climatology*. Cambridge, Mass., Harvard University Press, 1944.
20. COREY, C. P., "The Effects of Weather upon the Electric Power Systems." *Bull. Amer. meteor. Soc.*, 30:239–241 (1949).
21. COURT, A., "Separating Frequency Distributions into Two Normal Components." *Science*, 110:500–501 (1949).
22. —— "Wind Chill." *Bull. Amer. meteor. Soc.*, 29:487–493 (1948).
23. DILL, D. B., *Life, Heat and Altitude: Physiological Effects of Hot Climates and Great Heights*. Cambridge, Mass., Harvard University Press, 1939.
24. DINIES, E., "Luftkörper-Klimatologie." *Aus. d. Arch. dtsch. Seew.*, Bd. 50, Nr. 6, 21 SS. (1932).
25. DORNO, C., *Das Klima von Davos*. Braunschweig, F. Vieweg & Sohn, 1930.
26. —— *Das Klima von Agra (Tessin)*. Braunschweig, F. Vieweg & Sohn, 1934.
27. —— *Grundzüge des Klimas von Muattos-Muraigl*. Braunschweig, F. Vieweg & Sohn, 1927.
28. DRAKE, L., *Putting the Weather to Work for You*. American

Retail Federation, Washington, D. C., 25 pp., 1946.
29. Dryar, H. A., "Load Dispatching and Philadelphia Weather." *Bull. Amer. meteor. Soc.*, 30:159–167 (1949).
30. Duenst, J. U., *Sauerstoffschwankungen der Atemluft in ihrer formbildenden Wirkung bei Mensch und Tier.* Leipzig, Haupt, 1937.
31. Eckert, W. J., *Punched Card Methods in Scientific Computation.* New York, The Thomas J. Watson Astronomical Computing Bureau, Columbia University, 1940.
32. Fitzgerald, S. E., *A Study of the Copper Man, Phase I; Physical Characteristics, Thermometry, Air Clo. Evaluations.* Office of the QMG, Clim. Res. Lab., Lawrence, Mass., 1946.
33. Flach, E., "Entwurf einer Wetter- und Klimadarstellung für Heilstätten und Kurorte." *Strahlentherapie*, 40:672–681 (1931).
34. Fletcher, R. D., "A Relation Between Maximum Observed Point and Areal Rainfall Values." *Trans. Amer. geophys. Un.*, 31: 344–348 (1950).
35. Fritz, S., and MacDonald, T. H., "Average Solar Radiation in the United States." *Heat. & Ventilating*, 48:61–64 (1949).
36. George, J. J., "Fog: Its Causes and Forecasting with Special Reference to Eastern and Southern United States." *Bull. Amer. meteor. Soc.*, 21:135–148, 261–269, 285–291 (1940).
37. George, M. C., "Annotated Bibliography of Some Early Uses of Punched Cards in Meteorology and Climatology." *Bull. Amer. meteor. Soc.*, 26:76–85 (1945).
38. Gray, J. L., *Climatic Elements Related to Building Research.* Paper presented before R. Meteor. Soc. Can. Branch and ASHVE (Toronto Chapter), mimeogr., 11 pp., Nov. 5, 1945.
39. Gumbel, E. J., "Probability Tables for the Range." *Biometrika*, 36:142–148 (1949).
40. Hand, I. F., "Solar Energy for House Heating." *Heat. & Ventilating*, 44:80–94 (1947).
41. Hayes, G. L., "Forest Fire Danger" in *Yearbook of Agriculture 1949.* U. S. Dept. Agric., Washington, D. C. 1949. (See pp. 493–498)
42. Hill, L., Griffith, O. W., and Flack, M., "The Measurement of the Rate of Heat-Loss at Body Temperature by Convection, Radiation and Evaporation." *Phil. Trans. roy. Soc. London*, (B) 207:183–220 (1916).
43. Hottinger, M., *Klima und Gradtage in ihren Beziehungen zur Heiz- und Lüftungstechnik.* Berlin, J. Springer, 1938.
44. Huntington, E., *Civilization and Climate*, 3d ed. New Haven, Yale University Press, 1935.
45. Jacobs, W. C., "Wartime Developments in Applied Climatology." *Meteor. Monogr.*, Vol. 1, No. 1, 52 pp. (1947).
46. —— "Weather Analysis—Aid to Business Planning." *Dun's Rev.*, 55:15–18 (1947).
47. —— *Frost Survey of Imperial County, California.* 31 pp. Publ. Agric. Dept. Imperial Co., El Centro, Calif., Feb. 1941.
48. Kestermann, A., "Abkühlungsstudien mit besonderer Berücksichtigung des Frigorigraphen nach Büttner und Pfleiderer." *Biokl. Beibl.*, 7:1–16 (1940).
49. Klein, W. H., *Calculation of Solar Radiation Intensity and the Solar Heat Load on Man at the Earth's Surface and Aloft.* Army Air Forces, Air Tech. Serv. Command, Eng. Div. Memorandum Rep., Aero. Med. Sec. Ser. No. TSEAA-695-64, 77 pp., Feb. 20, 1946.
50. —— *Ibid.*, Addendum I, 5 pp., June 17, 1946.
51. Kratzer, A., *Das Stadtklima*, (*Die Wissenschaft*, Bd. 90.) Braunschweig, F. Vieweg & Sohn, 1937.
52. Lahmeyer, F., und Dorno, C., *Das Klima von Assuan.* Braunschweig, F. Vieweg & Sohn, 1927.
53. Landsberg, H., "On the Relation Between Mean Cloudiness and the Number of Clear and Cloudy Days in the United States." *Trans. Amer. geophys. Un.*, 25:456–457 (1944).
54. —— "On Climatological Aids to Local Weather Forecasting." *Bull. Amer. meteor. Soc.*, 22:103–105 (1941).
55. —— "Critique of Certain Climatological Procedures." *Bull. Amer. meteor. Soc.*, 28:187–191 (1947).
56. —— "Air Mass Climatology for Central Pennsylvania." *Beitr. Geophys.*, 51: 278–285 (1937).
57. —— "Include Climate in Highway Plans." *Better Roads*, 17: 25–28 (1947).
58. —— "Microclimatology." *Archit. Forum*, 86:114–119 (1947).
59. —— "Use of Climatological Data in Heating and Cooling Design." *Heat. Pip. Air Condit.*, 19:121–125 (1947).
60. Lee, D. H. K., and Lemons, H., "Clothing for Global Man." *Geogr. Rev.*, 39:181–213 (1949).
61. Linke, F., "Klimatische Anforderungen an einen Kurort." *Biokl. Beibl.*, 5:7–20 (1938).
62. —— und Dinies, E., "Ueber Luftkörperbestimmungen." *Z. angew. Meteor.*, 47: 1–5 (1930).
63. Manos, N. E., and Molo, W. L., "A Technique for a Synoptic-Climatological Study of an Air Route." *Bull. Amer. meteor. Soc.*, 29: 401–407 (1948).
64. Mills, C. A., *Medical Climatology: Climatic and Weather Influences in Health and Disease.* Springfield, Ill., C. C. Thomas, 1939.
65. Neumann, J. von, and Morgenstern, O., *Theory of Games and Economic Behavior*, 2d ed. Princeton, N. J., Princeton University Press, 1947.
66. Nuttonson, M. Y., "International Cooperation in Crop Improvement Through the Utilization of the Concept of Agroclimatic Analogues." *Interagra*, Vol. 1, Nos. 3–4, 8 pp. (1947).
67. Petersen, W. F., *The Patient and the Weather*, Vols. 1–4. Ann Arbor, Mich., Edwards Brothers, 1934–1938.
68. Pollak, L. W., "Charakteristiken der Luftdruckfrequenzkurven und verallgemeinerte Isobaren in Europa." *Prager geophysikalische Studien I: Cechoslovakische Statistik*, Bd. 44 (Reihe XII, Nr. 6) (1927).
69. Poppendiek, H. F., *Investigation of Velocity and Temperature Profiles in Air Layers Within and Above Trees and Brush.* Los Angeles, Calif., U. of Calif., Dept. of Engng. (ONR Rep., Contract N6-ONR-275, T. O. VI, NR-082-036) dittoed, 45 pp., 1949.
70. Price, A. G., *White Settlers in the Tropics.* New York, Amer. Geogr. Soc., Spec. Publ. No. 23, 311 pp., 1939.
71. Putnam, P. C., *Power from the Wind.* New York, Van Nostrand, 1948.
72. Roth, R. J., "Crop-Hail Insurance in the United States." *Bull. Amer. meteor. Soc.*, 30:56–58 (1949).
73. Rudder, E. de, *Grundriss einer Metcorobiologie des Menschen*, 2. Aufl. Berlin, 1938.
74. Ryzhkov, Cmdr. K., *Soviet Climatology for 30 Years.* U. S. Weather Bureau Seminar Talk, planogr., 16 pp., April 7, 1948 (Quotes work by the following U.S.S.R. authors: P. Tomashevich, B. P. Veinberg, O. A. Drosdov, A. A. Shepelevsky, U. N. Korotkevich, E. I. Abramova, G. L. Gaken, R. F. Sokhrina, and E. S. Rubinstein.)
75. Sette, F. J., "The Planning of Aerial Photographic Projects." *Trans. Amer. Soc. civ. Engrs.*, 112:595–611 (1947).
76. Shands, A. L., and Ammerman, D., "Maximum Recorded

U. S. Point Rainfall." *U. S. Wea. Bur., Tech. Pap.* No. 2, 36 pp., Washington, D. C. (1947).

77. Siple, P., "American Climates." *Bull. Amer. Inst. Archit.*, 3:16–36 (1949).
78. Thilenius, R., und Dorno, C., "Das Davoser Frigorimeter." *Meteor. Z.*, 42:57–60 (1925).
79. Thom, H. C. S., *Analysis of Charts on Corn Maturity and Frost Data.* Weather Div., Iowa Dept. of Agric., mimeogr., Sept. 1945.
80. —— *Evaluation of Corn Progress Based on Phenological Data.* Weather Div., Iowa Dept. of Agric., mimeogr., Sept. 1946.
81. —— "What's in the Weather." *Iowa Farm Sci.*, 2:3 (1948).
82. Thornthwaite, C. W., *The Climate of the Soldier, Pt. IV, Determining the Wind on the Soldier.* Environmental Protect. Sec. No. 124, 13 pp., Office of the QMG, Research and Development Branch, Feb. 1949.
83. —— "An Approach Toward a Rational Classification of Climate." *Geogr. Rev.*, 38:55–94 (1948).
84. U. S. Weather Bureau, *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere, Sea Level, 1899–1939.* Washington, 1943, 1944.
85. Weger, N., "Obstbaumblüte und Wetter." *Umschau*, Nr. 13 (1942).
86. Wellington, W. G., "Conditions Governing the Distribution of Insects in the Free Atmosphere, Pts. I–III." *Canad. Ent.*, 77:7–15, 21–28, 44–49 (1945).
87. Wexler, H., and Tepper, M., "Results of the Wartime Historical and Normal Map Program." *Bull. Amer. meteor. Soc.*, 28:175–178 (1947).
88. Wolfe, J. N., "The Use of Weather Bureau Data in Ecological Studies." *Ohio J. Sci.*, 45:1–12 (1945).
89. Young, F. D., "Frost and the Prevention of Frost Damage." *U. S. Dept. Agric., Farmer's Bull.* No. 1588, 62 pp., Washington, D. C. (1935).
90. *A Few Sources of Climatological Data.*
    a. Army Air Forces, Headquarters, *General Climatic Information Guides.* Weather Information Branch, Washington, D. C., 1943–1944.
    b. Clayton, H. H., "World Weather Records." *Smithson. misc. Coll.*, Vols. 79, 90 (1921–1930); 105 (1931–1940), Smithsonian Institution, Washington, D. C. (1927, 1934, and 1947).
    c. McDonald, W. F., *Average Precipitation in the United States for the Period 1906–1935 Inclusive.* U. S. Weather Bureau, Washington, D. C., 1944.
    d. Meteorological Office, London, "Monthly and Annual Summaries of Pressure, Temperature and Precipitation at Land Stations." *Réseau mond.*, 1910–1931 (issued in parts 1917–1938).
    e. U. S. Department of Agriculture, *Climate and Man: Yearbook of Agriculture.* Washington, D. C., 1941.
    f. U. S. Weather Bureau, *Airway Meteorological Atlas for the United States.* Washington, D. C., 1944.
    g. ——*Atlas of Climatic Charts of the Oceans.* Washington, D. C., 1938.
    h. —— *Climatic Summary of the United States* (by sections). U. S. Dept. of Agric., Washington, D. C., 1932–1936.
    i. —— *Climatological Data for the United States* (by sections). Washington, D. C. (appears in annual volumes).
    j. —— *Weather Guides for Long Range Planning.* War Advisory Council on Meteorology, Washington, D. C. 1943.
91. *Literature on Objective Forecasting Techniques.*
    a. Brier, G. W., "Predicting the Occurrence of Winter Time Precipitation for Washington, D. C." *U. S. Wea. Bur. Res. Rep.* (Dec. 1945).
    b. —— "A Study of Quantitative Precipitation Forecasting in the TVA Basin." *U. S. Wea. Bur. Res. Pap.* No. 26 (Nov. 1946).
    c. Counts, R. C., "An Objective Method of Forecasting Rain for Portland, Oregon." *Mon. Wea. Rev. Wash.*, 77:133–140 (1949).
    d. Dickey, W. W., "Estimating the Probability of a Large Fall in Temperature at Washington, D. C." *Mon. Wea. Rev. Wash.*, 77:67–78 (1949).
    e. Jorgensen, D. L., "An Objective Method of Forecasting Rain in Central California During the Raisin-Drying Season." *Mon. Wea. Rev. Wash.*, 77: 31–46 (1949).
    f. Penn, S., "An Objective Method for Forecasting Precipitation Amounts from Winter Coastal Storms for Boston." *Mon. Wea. Rev. Wash.*, 76:149–161 (1948).
    g. Miller, J. E., and Mantis, H. T., "An Objective Method of Forecasting Visibility." *Bull. Amer. meteor. Soc.*, 29:237–250 (1948).
    h. Vernon, E. M., "An Objective Method of Forecasting Precipitation Twenty-four to Forty-eight Hours in Advance at San Francisco, California." *Mon. Wea. Rev. Wash.*, 75:211–219 (1947).

MICROCLIMATOLOGY*

By RUDOLF GEIGER

University of Munich

INTRODUCTION

Origin and History of Microclimatological Research. Microclimatology is the science of climate on the smallest scale. It developed as a separate branch of our meteorological science in two entirely different ways. First, the progressive development of a network of meteorological observation stations attracted attention to local peculiarities of climate. Data gathered in mountainous regions, at stations on slopes, in valleys, and on mountain peaks revealed characteristic peculiarities. Soon differences were detected also in level regions, between stations in larger cities and those in the open country, near forests, or in agricultural territories. The attempt to distribute observation instruments expediently led, as early as the end of the nineteenth century, to the recognition that the type and condition of the soil, as well as the vegetation, had a considerable effect on the climatic data obtained. At the time, emphasis was placed on large-scale climate rather than on local peculiarities, and consequently measuring instruments were placed at a height where they would lie above the region of surface effects (about two meters above the ground). Only later did investigation of the climate near the ground become of interest. This interest was prompted by the fact that the thermodynamics and hydrodynamics of the earth's surface determine changes in the state of the atmosphere above.

The development of microclimatology would never have taken place so rapidly had it not been for the fact that new demands were placed on meteorologists by entirely different fields such as agriculture, forestry, and horticulture, as well as by industry, transportation, and the building industry.

As the population in each country grew and the degree of cultivation was extended, it became necessary to exploit the natural resources of the land to an ever-increasing extent. Simultaneously, the demands placed on climatological data rose. Information recorded in almanacs became quite insufficient for the requirements of practice. The situation is best illustrated by an example (Table I).

Table I refers to a German forest range selected at random, located in flat country (Eberswalde, near Berlin); it shows differences in the danger of late frost in the spring of 1939. The forester is concerned with the climate prevailing at the height of sprouting pine trees, 10 cm above ground. Table I provides the pertinent information, namely the frequency with which late frosts occurred as well as the date of last occurrence. Furthermore, the temperature minima of a damaging frost-night in May and of one in midsummer are given. It is seen that frost incidence and intensity are entirely different at the nearest climatological observation station only a few kilometers away. The measurements were confirmed by the existence of large areas of frozen vegetation.

TABLE I. DIFFERENCES IN THE INCIDENCE OF LATE FROST IN A GIVEN FOREST RANGE (*After Geiger and Fritzsche* [13])

Location	No. night frosts May 14–July 19 1939	Last frost	Temperature minimum May 29–30 1939 (°C)	Temperature minimum July 11–12 1939 (°C)
Meteorological station at Eberswalde, given for comparison.	0	April 10	+3.0 (−1.2)*	+11.2 (+7.5)*
Forest range outside Eberswalde at the same altitude, 10 cm above ground level.				
1. Farm area plowed two months previously to a depth of 24 cm.	7	June 12	−3.4	+5.9
2. Uncultivated ground overgrown with grass.	11	June 21	−6.5	+1.5
3. Abandoned farm area covered by weeds.	17	July 7	−9.0	−2.5

* Minimum at 5 cm.

Every forester, farmer, and horticulturist must select his plant varieties and his cultivating methods on the basis of the climate actually prevailing in small regions—we call this the *microclimate*. Location considerations in plant growth are always microclimatological in nature; whether a plant is located in an open field, or at the north side of a large boulder, or at the foot of a high oak tree is a decisive factor with respect to the conditions affecting its growth. The vineyard owner is interested in knowing the temperature conditions to which a ripening grape is subjected when it is located at a height above the ground which can be determined by the method of cultivation. Similarly, every engineer must know what climatic conditions his work will be subjected to, for instance the frequency with which his concrete road will be subjected to freezing and thawing in spring and fall. The scientist also requires such data for research purposes. The biology of bees can be understood only if one knows the microclimate of the beehive. The list of examples could be continued indefinitely. It is the problem of the microclimatologist to investigate these small-scale climates.

* Translated from the original German.

In some of these cases one deals with a single weather process, that is, a meteorological event. Thus there exists also a science of *micrometeorology*. Thornthwaite [31] ascribed the difference between these disciplines to the research methodology used, stating that

> Micrometeorology deals with the physics of a layer and aims at discovering physical principles, whereas microclimatology is concerned with the geographical distribution, both areally and vertically, of the various properties of the air, and seeks patterns of geographical distribution.

It has so far been customary (and will probably also prove expedient in the future) not to distinguish between two separate disciplines, but to assemble all small-scale investigations under the collective term *microclimatology*. For just as modern climatology with its method of frequency statistics accepts the synoptic weather processes in the domain of its studies, so microclimatology will also include small-scale single weather processes, that is, micrometeorological events.

THE PRESENT STATE OF RESEARCH

Microclimate as the Climate near the Ground. We shall first consider microclimate as the special climate prevailing in a layer of air about two meters in height adjacent to the surface of the ground. In this layer, friction between the air and the earth's surface plays a decisive role. In the lowest few decimeters, wind velocity is very low; mixing of the air due to turbulence is therefore also slight. However, the greatest portion of radiation from the sun and ground is absorbed at the surface of the ground. Heat radiation during the night also takes place at the surface. Precipitation is absorbed by the ground. The ground furnishes water vapor as well as carbon dioxide and other substances to the atmosphere. Exchange of heat and water thus takes place in the boundary layer between ground and atmosphere. The layer near the ground is the first intermediary between ground and atmosphere, and it is precisely here that the vertical transport of all properties is greatly impeded by wind resistance along the ground. It is therefore in this layer that we encounter the greatest differences in the smallest space.

Temperature Conditions. The slight air movement, the strong radiation reflected by the ground surface, and the large vertical temperature gradients make it difficult to obtain exact measurements of the temperature of the air layer near the ground. The problem of the measuring technique is not as yet satisfactorily solved.

The use of thermometers which are artificially ventilated and protected from radiation is possible only at heights above roughly one meter. The first exemplary experimental arrangement of this kind was applied to microclimatology by Flower [7] in England. Closer to the surface, ventilation destroys the natural temperature stratification that is to be measured. Moreover, no suitable protection from radiation by the sun, the sky, and the earth's surface has yet been found. In most cases, the unavoidable radiational error increases systematically with the approach of the thermometers to the ground surface. Thereby, excessively large gradients are simulated when the insolation is strong. In addition, the protective device also shades the ground and thus affects the measuring field.

At present, Albrecht's method of temperature measurement [1], involving a single platinum wire as an electrical resistance, is probably the best. It is well known that the radiational error of a wire exposed to sunshine decreases with decreasing diameter. If wires of only 0.015-mm diameter are used, the error becomes negligibly small even in calm air, a fact that now can also be proven theoretically. Such a wire is stretched horizontally at the desired height of measurement. Its length serves to average the temperature of many eddies at the selected height of measurement. This is an advantage considering the extraordinary temperature fluctuations near the ground. This arrangement has proved satisfactory even for temperature measurements in the Gobi Desert.

The most important characteristic of the air temperature near the ground is the unusually large vertical temperature gradient, especially by day. Table II refers to measurements by Mal, Desai, and Sircar [21]; it lists the frequency distribution of temperature gradients observed at the Karachi airport by means of ventilated electric-resistance thermometers; values are converted to degrees Fahrenheit per 100 m for the purpose of comparison. The measurements were not actually made in the layer near the ground. They serve to illustrate, however, how variation in the prevailing gradient increases as the ground is approached. According to measurements made by Best in England, by means of ventilated thermoelements, the gradient in the layer between 1.2-m and 0.3-m altitude at 12 noon, even in the middle of June, amounts to 139F per 100 m, and as much as 1230F per 100 m in a layer between 0.300 and 0.025 m, if conversion to a 100-m altitude is made.

Table II. Yearly Percentage Frequencies of Lapse Rates of Various Magnitudes in Karachi

Lapse rate (°F/100 m)	+57	+43	+31	+21	+13	+7	+3	+1
Layer 4–56 ft (%)	0.1	0.4	2.0	8.5	13.0	13.0	12.0	7.1
Layer 56–156 ft (%)	—	—	—	—	0.6	2.5	24.1	23.7

Lapse rate (°F/100 m)	−1	−3	−7	−13	−21	−31	−43	−57
Layer 4–56 ft (%)	6.6	5.3	7.6	6.7	6.2	6.4	3.6	1.1	0.3
Layer 56–156 ft (%)	16.7	7.9	7.3	7.2	6.3	2.7	0.7	—	—

These high gradients are accompanied by a number of other phenomena which are equally characteristic. Rapidly recording electric thermometers reveal extraordinarily great temperature fluctuations which never subside and are particularly strong during periods when the incident radiation is large. They are caused by the rising of hot air parcels in proximity to descending

cold air parcels. These fluctuations are visible in the form of scintillations which can be observed at noon above heated roads or railroad embankments. If, in making this observation, one moves his eye downward until a layer directly above the ground is observed, one will be surprised to note a considerable increase in scintillation. The high lapse rate, moreover, gives rise to reflections by air strata, also known as *road mirages* which also occur only in the microclimate near the ground.

From the great decrease in temperature with altitude at noon (incoming-radiation type) and the great increase during the night (outgoing-radiation type) it follows that daytime temperature fluctuations near the ground are considerable. If a meteorological station located 2 m above ground observes a daytime fluctuation of 10C, the value at 1 m increases to about 15C, at 10 cm to about 20C, and at the ground to a value of 30C or even 40C. This variation can occasionally be observed in the erosion of vertical rocks or buildings.

Effect of the Type of Ground. Since microclimate is controlled by the ground, it follows that the type and condition of the soil must have a considerable effect on the microclimate near the ground. Surface properties determine, first of all, how much radiation is absorbed or reflected. Dark areas (peat bogs) absorb more, light areas less; moist ground absorbs more than dry ground. Surface properties are usually different for the absorption of short-wave radiation from sun and sky during the day than they are for the emission of longer wave lengths during the night. The thermal conductivity of the surface controls the distribution of absorbed heat between the ground and the air. If the thermal conductivity is large, the ground absorbs much heat during the day and then conducts much heat to the radiating ground surface during the night. Temperature conditions in the air layer near the ground are consequently moderate. Ground of poor conductivity transmits much heat to the air during the day and removes heat from the air during the night. In such cases harmful air temperatures are attained at noon, and at night the danger of frost is correspondingly great. Thermal conductivity depends on the condition and composition of the soil, and on its structure. The latter is determined by the treatment which the soil has been given. A table of various soil constants (which can serve merely as a qualitative indication in view of the great variety of soil conditions) will be found in the textbook on microclimatology by Geiger [11].

The microclimatologist can understand the climate near the ground only if he is familiar with the climate prevailing in the ground itself. The farmer, however, by his practice of cultivating and fertilizing the soil, changes its condition and thus affects the climate near the ground—the climatic environment of the plants he cultivates. Even changes in the ground's surface and its properties often have a great effect, for instance covering the ground with evergreen twigs, straw, etc. (mulching) and covering peat bogs with sand. The greatest obstacle to the investigation of this ground climate is the difficulty of measuring in a simple manner the water content and the water balance of the ground. Nevertheless, at the present time, investigations that raise hope for progress in the near future are being conducted in various countries. If soil volumes of more than one cubic meter are sunk into the ground flush with the surface, and if the weight of this soil can be measured with sufficient accuracy, the water loss (evaporation) or the water gain (dew and frost formation) of the soil can be determined from the precipitation and the seepage. In 1930–1937, J. Bartels and J. Schubert [3, 10] obtained a good series of measurements of this kind in Eberswalde. However, the instrumental equipment is very expensive because of the extensive installation work and the balances that must have a high accuracy (about 0.1 kg) under great loads (about 1500 kg). For this reason, employment of this method at many locations can hardly be expected. At present, C. W. Thornthwaite is undertaking promising experiments at Seabrook, N. J., on the evapotranspiration of natural and cultivated soil surfaces. In India, Ramdas [25] found a method for measuring the water content of the soil without changing (digging) the soil itself; this method consists in electrically heating a soil thermometer (mercury) and measuring the subsequent rate of temperature change. As soon as progress has been made in these instrumental problems, it will be possible to study the influence of the type and state of the soil more adequately than has been the case thus far.

The air layers near water and snow are of special interest for the further systematic development of microclimatology. The uniformity of the water surface and of the freshly fallen snow cover facilitates the elimination of all advective influences. Because of its small heat conductivity, the snow cover also eliminates the influence of the soil. Investigations by Bruch [4] and Roll [26] of the temperature and wind field over water surfaces have greatly promoted knowledge of the air layer near a water surface. Moreover, these researches have furnished new information regarding the process of wave formation, the change of an air mass in moving from land to sea, and the usefulness of the temperature of the water surface reported by ships in the synoptic weather service.

Moisture Conditions. The measurement of the atmospheric humidity near the ground is nowadays attempted in three different ways: (1) by hair hygrometers in which the orientation of the hair and the construction of the instrument are adapted to the horizontal stratification of the humidity (Diem [6]), (2) by psychrometers with very fine thermoelements whose radiational error can be neglected (Franssila [9]), (3) by aspirating air from the desired height into a modern dew-point hygrometer (Thornthwaite [31, 33]). The first two methods are distinguished by their simplicity, the third method by its high accuracy.

Unfortunately, extensive series of measurements of the water-vapor distribution near the ground are as yet unavailable. As with temperature, this distribution is quite different from the conditions prevailing at a height of 2 m. The temperature at ground level may

be lower at times and higher at other times than the air above; deviations in humidity, however, are almost always in the same direction. Just as wind velocity is lower near the earth's surface, so the absolute and relative humidities are always higher. Only in the morning hours is the ground surface sufficiently colder than the air for water vapor, in the form of heavy dew, to condense there in a quantity sufficient to dry out temporarily the lowest meter of the atmosphere. (Often this applies only to a layer several centimeters deep.) A relative humidity at 10 cm which is 30 to 50 per cent higher than that at 2 cm is not rare. Thus the climate near the ground is also entirely different in this respect.

Rapid fluctuations of humidity with time are as characteristic of the microclimate near ground as are temperature fluctuations. Air parcels enriched in moisture content rise upward from the ground, dry parcels sink downward. Because of poor mixing they cause the indicators of special hygrometers to move back and forth in jumps. When the ground is covered with drought cracks, saturated air rises from the depths of these crevices and causes deep layers to dry out more quickly.

Moisture layers near the ground are sometimes visible in the form of flat fog banks in enclosed spaces or as the steam above highways which is so annoying to the motorist (for instance after a shower strikes a hot road surface). However, the formation and dissolution of fog above the ground, as well as the formation of dew, have not yet been investigated sufficiently. Exact investigation is rendered difficult by the discontinuous structure of the layer near the ground, varying conditions of exchange, and the effects of colloido-chemical and aero-electrical processes superimposed on the thermodynamic changes. Moreover, humidity changes in the ground air as well as condensation within the ground must also be taken into consideration. Investigations along these lines seem promising and may provide new information basic to the entire field of microclimatology.

Wind Conditions. Today we are well informed on the variations of the wind speed in the air layer near the ground. All measurements in the air layer near ground and water surfaces are, according to investigations by Thornthwaite and Halstead [32], represented with sufficient accuracy by the equation

$$u^p = (\log z - \log z_0)/\log a,$$

where u is the wind speed (m $\sec^{-1}$) at the height z (m); p is a parameter whose value depends on the magnitude of the vertical temperature gradient. If $p = 1$, the equation above is transformed into the logarithmic law that was formerly in general use. In this case, the observed values, when graphically shown in a coordinate system with u as abscissa and $\log z$ as ordinate, lie on a straight line whose slope is determined by the value of $\log a$ and whose point of intersection with the ordinate is determined by the magnitude of $\log z_0$.

However, in the same coordinate system, the measured values do not lie on a straight line but on a curve that is concave upward, if the temperature stratification is unstable. In reverse, during a stable nocturnal inversion a curve is found that is convex upward. The parameter p in the equation above takes care of this influence of the vertical temperature gradient. When a strong temperature decrease exists with increasing height, then $p > 1$; with great instability, p may reach the value of 2. For stable stratification $p < 1$. For the smallest amount of turbulent mixing of the air that may occur in practice, p has the value of approximately 0.5.

According to the equation given above, a value $u = 0$ is reached at a height z_0. The value of z_0 (roughness height) depends on the nature of the surface and on the existing meteorological conditions. Over water, snow, and level bare soil, a value of $z_0 = 0.02$ m can be used as an approximation.

Eddy Diffusion. Whereas the variation of wind with height above the ground is very well known, the determination of the austausch coefficient A represents one of the most difficult and immediate problems of microclimatology.

At the surface, a boundary layer of about 1-mm thickness is assumed, in which molecular processes (molecular heat conduction, and diffusion) are almost exclusively effective. Above this layer, eddy diffusion determines the vertical exchange of all properties in the air layer near the ground. Although the value of A is of greatest importance for the understanding of the total heat and water balance at the ground, only a few measurements of A in the air space near the ground are available.

The measurement of A is accomplished by the simultaneous observation of the short-period fluctuations of a meteorological element (temperature, momentum, water-vapor content) and its vertical gradient. This method assumes that, for example, with a very unstable temperature stratification a relatively warm air particle comes from below, a relatively cold one from above. Since the same mixing process exchanges all properties simultaneously, it might be expected that, at a given place and time, the same value of A would be found independently of the meteorological element that is employed for the determination of the magnitude of A. This universal character of the austausch forms the basis of many computations of the heat balance of our earth. However, more recent measurements show that quite different values of A are obtained when different elements are used. Thus, the austausch proceeds in a different manner for the different elements (unless the differences are attributed to the properties of the instrumental arrangement). None of the several explanations found in the literature is completely satisfactory.

Frankenberger [8] made simultaneous measurements of the microstructure of the horizontal and vertical wind speed by means of a very sensitive instrument and showed that there exist two entirely different types of austausch motion. As regards the magnitude of this motion, he proved not only that faster air particles come from above, slower ones from below (which represents vertical motion promoting aus-

tausch), but that in more than one-third of the cases faster air particles come from below, slower ones from above. The latter motions impede austausch and increase, rather than diminish, the existing gradient. Future measurements of A must, therefore, be made separately for the two types of vertical austausch motion. This makes the measuring technique even more difficult. Nevertheless, all heat and water vapor problems of the air layer near the ground will not be solved unless an exact and, from the practical point of view, not excessively difficult determination of the quantity A is made possible.

Contour Microclimates. Another group of microclimates is caused by topographic formations. The differences between sunny and shady locations or between valleys and mountain peaks which are known in large-scale climatology occur even in minute dimensions. An anthill of some 10-cm height, for instance, with its conical shape, possesses all the climate typical of slope locations, even though only within the air layer immediately adjacent to the surface: there is a warm, dry, south side (which the ants systematically use for breeding) and a shady, cool, moist, north side. Typical differences between east and west are also observed; in the case of uniformly incident radiation on cloudless days these differences are caused by the fact that the morning sun serves primarily to dry out the ground, while the afternoon sun mainly serves to heat it up.

Plowed furrows represent mountain chains on a small scale; the corresponding microclimate is determined by the orientation of these furrows. Kaempfert [17, 18] has investigated the radiation conditions for this case in detail. Variations are occasionally visible in the form of differences in the variety of weeds growing on the two sides of the furrow. Every large rock represents a climatic divide on a small scale. Ullrich and Mäde [34] even found that the air, which is stratified in curved layers during the night, reveals a temperature stratification, in millimeter dimensions, entirely analogous to the temperature distribution in flat valleys.

A continuous range of intermediate conditions exists between these small-scale phenomena and the large-scale climatic differences observed by the mountain, valley, and slope stations of meteorological observation networks. Frequently, however, the laws of microclimatology are different from those applicable to large-scale climatology. For instance, in the case of high mountains located in a region of predominant west winds the west slopes are rich in rain because here the water vapor in the rising air is forced to condense. In the case of low hills, on the other hand, such thermodynamic processes have no effect. On the contrary, in this latter case the east side is the more humid because the field of motion of the wind determines the distribution of precipitation. Precipitation is lacking on the windward side of obstacles. During snow this can sometimes be observed directly, but it should not be confused with the redistribution of already fallen snow by the wind. We do not know as yet just where the line between "large-scale" and "small-scale" is to be drawn in this sense.

The effect of exposure on the microclimate is also determined decisively by the large-scale climate. In tropical regions where the sun is almost vertical at noon there is little variation between differently oriented slopes. The variation increases with decreasing height of the noon sun, that is, with increasing latitude. On the other hand, it is only the direct radiation from the sun which produces differences due to the direction of exposure, while diffuse radiation from the sky affects all slope orientations almost equally. The portion of total radiation which is diffused increases with latitude; differences in exposure are consequently decreased again in polar climates. In the far north, where there is a deficiency of heat, plants flourish only on southern slopes. In subtropical steppes, on the other hand, where there is a deficiency of water, it is often found that microclimatological conditions are adequate for plant growth only on the north side of rocks or hills.

Mountain and valley winds are known to arise locally in mountainous regions. They frequently make the microclimate *dependent*. A microclimate is referred to as *independent* if it can be explained solely on the basis of local meteorological conditions; it is *dependent* if it is determined also by extraneous effects in surrounding regions. This distinction is essential to a rapid and complete understanding of the various microclimates. A field immediately adjacent to a concrete road, or a meadow next to a pond, has a different microclimate than it would have if situated in surroundings of the same type. In any territory it is the wind peculiar to the time of day which makes extraneous effects felt over wide regions. This is particularly true of night winds, which move cold air toward depressions in a slow but steady flow persisting during the entire night. Local pools of cold air result, an example of which has already been offered in Table I. They are decisively affected by the "source regions," that is, the space from which cold air can be supplied. Railroad embankments, hedges, and forests are sufficient to stop or deflect these shallow cold air masses; underpasses, excavations, or gulleys may cause them to drain off. This explains the fact that forests are located everywhere above the vineyards along the Rhine; these forests prevent cold air from flowing down into the vineyards during the frost period in spring.

A normal decrease of temperature with altitude and the accumulation of cold air in valleys are responsible for the fact that a "thermal belt" is found to extend along a line halfway up most slopes; it is here that plants sprout earliest in the spring, that damage due to late frost is slightest, and consequently that the plant and fruit varieties most sensitive to climate thrive. Table III lists measurements made at twenty-four observation stations on the east slope of the Arber (Bavaria) during spring nights in 1931 and 1932. Here the thermal belt is located at about 820 m above sea level during radiation nights (continental winds); during west-weather nights (maritime winds) it is situated several decameters lower. The last column shows the

frequency with which the thermal belt can be found at certain altitudes. The maximum (16 per cent) in the valley (660 m) includes the cases of stormy weather

TABLE III. THERMAL BELT ON THE EAST SLOPE OF THE ARBER, BAVARIA
(Summary of Measurements Made at 24 Observation Stations)

Altitude above sea level (m)	Temperature minimum		Frequency of inversion (%)
	36 spring nights with continental breeze (°C)	24 spring nights with maritime breeze (°C)	
860	8.2	6.5	10
820	9.0	7.0	45
780	8.2	7.4	17
740	6.8	7.0	6
700	5.3	6.5	6
660	4.1	6.1	16

when no thermal belts were formed and consequently, in view of the normal temperature gradient, the foot of the valley was warmest. Otherwise, however, the highest night temperatures were found regularly at altitudes of about 820 m. Such facts are of the greatest importance in farming and fruit growing.

Microclimates in the Vegetation Layer. An intimate mutual dependence exists between a plant and the microclimate within the stand of vegetation. As soon as a young sprout has pierced the ground and developed its first leaves, it is no longer passively subjected to climatic conditions, but helps to form them. The surface is uniformly roughened. The air layer near the ground becomes more quiet, plant organs absorb radiation from the sun, casting a proportionate amount of shadow on the ground. Incident heat radiation is distributed vertically, extremes of surface temperature are decreased. In order to live, the plant must evaporate water which it obtains, if necessary, from layers deep below the ground. Humidity is thereby increased and then maintained nearly constant in the layer of stagnant air near the ground.

Microclimate is affected to an even greater extent by an entire society of plants. It is well known that plant societies grow more vigorously as soon as single plants have become large enough to touch the leaves of their neighbors. This is primarily due to the effect of the change in microclimate which also exerts a favorable influence on the nature of the soil. Research has been initiated during the past few years to investigate the dependence of these conditions on the type of plant, the density of planting, and the cultivation of the soil. This will offer the farmer, gardener, and horticulturist numerous artificial methods by which the growth of plant societies and their yield can be improved.

In high, close-knit plant societies, most clearly in the case of forests, new conditions arise because of the fact that an independent microclimate develops in an enclosed air space, separated from the free atmosphere. The ground surface has become a secondary boundary. Transfer of heat and water vapor takes place almost exclusively at the surface of the plant society. The interior climate is characterized by the fact that all meteorological quantities change very slowly, air humidity is high, the air is extremely quiet, and temperature conditions are moderate. This microclimate can often be understood most readily if one conceives of the space occupied by the plant society as being part of the ground, and then works with a heat conductivity and air exchange which are larger than within the ground, but very small compared to the conditions in the free atmosphere.

Now every plant society produces microclimates at its periphery which are widely different in nature, depending on the direction of exposure. To a first approximation, these can be compared with the corresponding microclimates near vertical walls. However, the analogy is not entirely accurate, in view of the fact that air exchange takes place between the interior climate of the plant society and the peripheral climates. In forestry these peripheral climates have been exploited for about a century to favor young forest growth. Young forests are raised in the climatic protection of old forests. The effect of peripheral climate, however, is not always favorable. At a southern periphery it may become too dry and hot, whereas a northern edge may be too shady; at an eastern periphery plants may be deprived of dew too soon. Frequently the danger of late frost is increased because of the forced quiescence of air near forests, particularly where plants have been stimulated to early sprouting, as at southern peripheries. These microclimatological conditions, which are known and exploited in practice, have hardly been investigated scientifically so far. Work in this field seems very promising. In particular, questions concerning the extent to which experience gained in one forest range can be applied to another region cannot be decided without a scientific basis.

APPLICATIONS AND FUTURE PROBLEMS OF MICROCLIMATOLOGY

Agriculture and Microclimatology. No branch of economics is associated with microclimatology as intimately and in as many ways as agriculture and the related fields of horticulture, viticulture, and fruitgrowing. Preceding sections have shown that the location environment of all cultivated plants depends upon the microclimate and the way in which it is affected by topography, surroundings, soil treatment, and the type of plant. An agriculture designed for maximum yield must therefore be based on a knowledge of and an expedient control of microclimatological effects. Without microclimatology it is not possible to deal effectively with the three major problems which face farmers all over the world. These are (1) wind protection, (2) frost protection, and (3) artificial watering.

The problem of *wind protection* has been comprehensively treated in the international literature. However, the present state of our research is very unsatisfactory because many occasional observations, but few systematic experiments, have been made. Distinction must be made between the changes of the wind field, which are caused by the measures taken for wind protection, and the effect of these changes on the microclimate and the harvest yield.

The changes of the wind field depend on the height, width, substance, and arrangement of the shelter belts. Most recently, the best pertinent investigations were carried out by Naegeli [22–24] in Switzerland from 1943 to 1947. Thus far, the following results are certain: An obstacle of height h reduces the wind speed at 1 ½ m height above the ground (or plant surface) by at least 10 per cent; the distance to which this effect is felt amounts to about $5h$ to the windward, about $25h$ to the leeward. The effect is almost independent of the wind speed; also the protective effect decreases only slightly up to the height h. Within the sheltered area, the distribution of wind speed is always nonuniform. Directly behind the shelter belt, the reduction of wind speed is greatest and amounts to 60 to 80 per cent. Easily penetrable obstacles always produce a better effect than do impenetrable obstacles (walls, very dense pine hedges). The minimum width of the shelter belts is determined by the requirement that the wind-breaking effect must be fully realized. (Otherwise, the width of the belts is determined by silvicultural and botanical considerations.)

The effect of artificial wind protection on the harvest yield does not depend directly on the changed wind field but on the changed microclimate. The macroclimate, the local features, and the existing weather conditions determine the reaction of the microclimate to the changed wind field. However, the same change of the microclimate may have an entirely different influence on the harvest yield. For example, in a semiarid area (such as the Ukraine) the reduction of evaporation by wind protection will improve the water balance and increase the harvest yield. In a humid marshland near a stormy coast (such as parts of Holstein), on the other hand, the reduced evaporation can be detrimental to the harvest. For this reason, harvest statistics are hardly a suitable approach to the microclimatic problem of artificial wind protection. In order to study the microclimate in a wind shelter, not only wind measurements, but, especially, simultaneous measurements of the heat and water balance are needed (radiation, temperature, humidity, evaporation, dew formation, austausch coefficient). This could best be achieved by strictly comparable observations made over a period of several years under different conditions of wind protection in an experimental field whose construction and equipment is reserved for this purpose. With increasing population of the earth and increasing utilization of the soil, the problem of artificial wind protection will become more urgent in the future. Microclimatological research should, in anticipation, prepare the necessary material.

The problem of *frost protection* is a different matter. Today, we are well informed regarding the possibilities of artificial frost protection. Kessler and Kaempfert [20] have summarized in their work all the necessary fundamentals. Only the measurements of the radiation balance within and without the artificial smoke clouds should be extended. Otherwise, the problem of frost protection is only a question of economic feasibility. This depends entirely on the location, the climate, and the market for the fruits which are grown. In the individual case, the microclimatologist must, above all, furnish statistics on the frost probability. Such statistics are the most important basis for the computation of the economic feasibility and must take into consideration the microclimate of the object to be protected.

As with the problem of wind protection, *artificial irrigation* assumes increasing importance. Irrigation was originally employed to help avoid crop failures. Today, however, it generally serves to increase the crop yields by offering the most favorable growing conditions to the plants. In some instances, irrigation makes possible multiple crops in a single growing season, where otherwise the natural aridity would prevent this (for example, the secondary crop yield in western Europe).

In localities where water for irrigation is not available in unlimited quantities, the timing and the amount of irrigation must be adjusted to the weather and climatic conditions. The effect of irrigation by day is different from that by night, because of varying temperature and humidity conditions in soil and air, and because of the varying temperature differences between the irrigation water and the plants. Likewise, frequent small amounts of water have an effect different from infrequent large amounts. Moreover, the plant in its various stages of development has a different sensitivity to water deficiencies. Also it is not yet certain what portion of the irrigation water returns through the soil to the ground-water table so that it is really not lost to the water balance of the land. This consideration is of great importance in countries in which conflicting interests in water (for household, industrial, shipping, and hydroelectric use) have arisen. All these questions can be answered only by carefully designed experiments. Work on these problems already has been initiated, and it is to be hoped that the status of microclimatology will soon be rapidly advanced.

Of importance, moreover, to future work is a microclimatological mapping project. There are agricultural maps which serve to give a general picture of the quality and yield of the soil (also used for the just assessment of taxes) and the possibility of planting various products; similarly, there is an increasing desire to map microclimate in an analogous manner. In 1948 Weger [35] mapped a vineyard region near Geisenheim on the Rhine from this point of view for the first time. In the same year Schüepp [30] in Switzerland mapped the possibility of raising potatoes in the Davos valley by distinguishing eight zones with different degrees of microclimatological frost incidence. Schnelle [28] is currently mapping large valleys in the Odenwald (Germany) in order to determine microclimatological conditions affecting fruitgrowing. These maps are to serve as the basis for a systematic expansion of fruit production and are held in high esteem by men engaged in practical work. All of these first attempts at mapping are thus connected with particular economic tasks, and they are therefore concerned with only a selected few of the microclimatological factors considered. The success achieved justifies a continuation of these

investigations. They provide the most effective methods of combating climatic damage; as W. J. Humphreys once said with regard to fruitgrowing, "The best time to protect fruit from frost injury is before the orchard is set out; obviously by carefully considering what kind, or even variety, to grow and where to grow it."

Often the farmer makes use of the artificial microclimate inside a greenhouse. In general, greenhouses are still being constructed extensively according to traditional practical experience. Investigations by Mäde and Kreutz concerning greenhouse climates reveal the extent to which microclimate depends on the materials and the methods of constructing and erecting various types of greenhouses. Air humidification by spraying has opened new possibilities of climatic control. Great progress is yet to be expected from systematic investigation along these lines.

After a farmer has brought in his harvest, the durability and quality of his product are determined by the microclimate of barns, shocks in the open air, silos, and storehouses. In the American occupation zone of Germany farmers are currently being advised by radio, on the basis of microclimatic measurements in experimental shocks, when they should ventilate their shocks or cover them more thoroughly. This practice has contributed greatly to the preservation of agricultural products. The health of dairy cows, hogs, etc., and the work capacity of animals depend on the microclimate of stables. Here also, research is in its infancy and these factors are only beginning to be considered in the design of stables.

The additive effect of changes in microclimate can influence the fertility of an entire country. The elimination of innumerable small and minute windbreaks is known to have laid waste entire regions. The cultural planning of cities and industrial areas has caused disorder in the hydrology of many a region. Fortunately, however, opposite instances are not lacking. Thousands of small hedges have afforded wind protection which increased agricultural yield and even improved large-scale climate. An interesting new example is offered by work done in Yakut in Siberia (Keil [19]). In 1939, under the pressure of war, workers began to clear large areas of a cushion of moss which had formerly served as a heat insulator and had consumed much heat by evaporation. The exposed soil now absorbed heat from the sun. The frozen soil thawed to a greater depth from year to year; soon the first agricultural plants began to thrive. The virgin soil bore a good harvest. Today the region in question supports a large population, and this success was actually achieved by the application of microclimatological experience.

Phenology as an Auxiliary Science. Phenology is currently developing into an important auxiliary science of microclimatology (Schnelle [28]). Phenology of plants concerns itself with the chronology of plant growth and its dependence on weather and climate. Suitable plants are used to observe the visible stages in growth, such as the development of the first leaves in spring, the first blossoms, the ripening of fruit, and the discoloration of foliage. Numerous observers, in Germany sometimes as many as ten thousand farmers, gardeners, etc., make systematic entries in this "phenological network."

The plant is thus used as an instrument for meteorological observations. This procedure has the advantage that, with a judicious choice of plants, a large number of widely scattered observation points are obtained. In addition each plant reacts to the sum of all meteorological factors at its location near the ground (heat, rain, radiation, wind), thus representing a microclimatological instrument. The disadvantage of the method is the fact that every plant, as a living organism, is an individual. Not only are different plant varieties affected differently by given weather conditions, but a different behavior is shown by every specimen of a single species. This disadvantage can be compensated by the choice of particularly suitable varieties and by the observation of a large number of single plants. Wild plants are more suitable than cultivated plants because the development of the latter may still be affected by cultivating methods such as choice of the time at which seeds are planted, treatment of the soil, or fertilization.

As early as fifty years ago Ihne [15] prepared maps showing the arrival of spring in a region on the basis of such systematic phenological observations. These maps represent a survey of orographic microclimates on a correspondingly large scale. In judging the microclimate of a region it is therefore advisable to observe the development of the plants very closely. Such observations supplement a small number of accurate measurements made with the psychrometer and anemometer and aid in interpolating microclimatic data between points at which climatic conditions have been accurately determined by ordinary observing stations.

Forestry and the Microclimate. A threefold connection exists between microclimatology and forestry. Young forest growth develops either on large cultivated areas, in which case it is subject to local microclimatic conditions as are agricultural plants, or else it grows by natural seeding or is planted by a forester at the edge or in the midst of the mother forest. The section on "Microclimates in the Vegetation Layer" already described the effect of all silvicultural measures on the microclimate of locations. Microclimatology is thus an auxiliary science of forestry.

Secondly, the effect of weather damage in forests depends to a great extent on local conditions. Wind damage always begins at the weakest point. The practice of eliminating such danger points by silvicultural measures requires a knowledge of the meteorological wind conditions. Topographic and vegetative obstacles must be considered in this connection. Snow damage and frost damage are also greater for certain tree formations than for others. The effect of a drought depends upon the method of planting, the mixture of tree varieties, and the treatment of the soil. Even forest fires require certain conditions of the ground and the undergrowth which can be affected by cultivating techniques. The macroclimate and the location of the forest in question determine which kind of weather damage

is most to be feared. The practical forester must consult with the microclimatologist in an effort to prevent such weather damage or at least to reduce it to a minimum.

Thirdly, the microclimatologist aids in combating insect damage in forests. Insects breed partly in the ground, partly in forest litter, and partly in the trunks or crowns of trees. The rate of reproduction of insect pests is determined only by the meteorological conditions prevailing at these locations. In this connection we might quote the English biologist Buxton [5]:

> ...the meteorologist has studied the temperature in a ventilated white screen, [but] the biologist wants to know what differences exist between the forest and the shore, or the rat's hole and the bird's nest: in the white screen he finds no living thing, unless it be an earwig.

Today these microclimates are being investigated to an ever-increasing extent. In connection with pest control by airplane dusting, the distribution of the insecticide and its effectiveness in the infected growth are determined not only by the weather conditions prevailing above the forest, but also by the temperature stratification, air humidity, and field of air motion within the individual forest.

Microclimatology in City Planning and Industry. Every newly built house produces a quantity of opposing special climates. Radiation from the sun which formerly impinged on a horizontal surface is now absorbed by vertical walls and a sloping roof. A radiant, hot, and usually also dry microclimate is produced at the south wall; frequently varieties of fruit thrive in this climate which could normally grow only in a macroclimate many hundreds of miles nearer the equator. The north wall of the house, on the other hand, suffers from a rough, moist microclimate, deficient in radiant heat, of the type usually found in regions nearer the poles. In temperate latitudes of the Northern Hemisphere, the south wall receives a larger total of incident radiation on a clear day in January than it does in July, and consequently the wall climate possesses characteristic traits not found elsewhere. Within the house every room from the basement to the attic has a different climate depending on the exposure of windows, the height of the ceiling, and the size of the room; these climates have been investigated in great detail for bioclimatological purposes. In designing houses and apartments, clinics and schools, factories and warehouses, it is possible to allow for and to influence the microclimate on the basis of previous experience by the choice of location, orientation, and distribution of rooms. Frequently even artificial climatization of rooms (operating and treatment rooms in hospitals, warehouses for perishable goods, shipholds) is possible and necessary.

The over-all effect of buildings which are grouped into settlements, towns, and finally modern cities influences the microclimate to an ever-increasing extent. The climate of large cities is distinguished primarily by a concentration of waste gases from industrial and domestic furnaces as well as of dust; in winter and on quiet days the first predominates, in summer and in windy weather the latter predominates. Parks are most effective in filtering out impurities. The first turbid layer is found above roads, near the ground; a second is found at the altitude of smoking chimneys. Seen from a distance, a large city appears to be surrounded by a hemisphere of haze.

These impurities in the air attenuate incident radiation from the sun and radiation emitted from the ground during the night. A city is therefore warmer at night than the surrounding countryside; this is disagreeably noticeable, particularly on hot summer nights. The decreased amount of solar radiation incident during the day is compensated by intensive absorption by walls and roads and by the absence of heat loss due to evaporation. Consequently the climate of a city is also hotter than its surroundings during the daytime, more so if a city contains few parks. In the absence of a gradient wind this becomes noticeable on clear summer mornings because of the flow of air into the city from all sides (as shown by chimney smoke). Trees in a city consequently begin to sprout earlier and terminate their vegetation period sooner than those in the country.

Impurities present in city air favor the formation of fog and precipitation. London fogs are world-famed. Wherever a sufficiently long series of investigations is on record, an increase in the frequency and density of fog is seen to accompany the growth of a city. Only recently has the increased pollution of city atmospheres been halted by the increased application of electric power and a more efficient use of coal by industry and heating furnaces. Precipitation in the form of drizzle is favored in cities and industrial areas. Occasionally precipitation of larger droplets may be triggered by large cities.

Elsewhere, also, the engineer is constantly creating new microclimates; for example, the pile of tailings from a mine represents an artificial mountain surrounded by a variety of new slope-climates. A railroad embankment or a railroad cut not only creates two new embankment climates, but it frequently also affects the surroundings over a wider range, for instance, by controlling the flow of cold air. A concrete road possesses a microclimate of its own, and the road may also affect the local field of wind flow (*e.g.*, storm damage in forests along roads) as well as the runoff and evaporation characteristics of the soil. It is maintained by some that the destruction of flourishing cultures which occurs again and again in world history may be ascribed primarily to a disruption of the dynamics of nature by the hand of man.

THE METHODOLOGY OF MICROCLIMATOLOGICAL RESEARCH

There are many problems of microclimatological research which still await solution, and the practical applications of this research are almost unlimited.

Techniques learned from general climatology are applicable only to a limited extent. There are two reasons for this. In the first place, it is impossible to measure

the great variety of microclimates by building an ever more dense network of observing stations. Moreover, since the essential characteristics of microclimates are repeated everywhere, it suffices to study them in special experimental areas. This was done, for instance, by Schmidt [27] in Austria with a special climatic network near Lunz, which achieved world-wide fame. Similar networks were previously and subsequently constructed in Germany. The knowledge of microclimatic types obtained in such experimental fields may be applied to unknown regions located in the same macroclimate.

Furthermore, the necessity for working in small spaces requires the creation of a special field of instrumentation to provide instruments which will not affect natural conditions and which are capable of making accurate measurements even near the ground, with inadequate natural ventilation, and often in the presence of intense radiation (both direct and reflected). We might agree with Thornthwaite [31] that "instrumentation remains the basic problem of the investigation." In microclimatology as well as in other sciences, the development of new measuring instruments has stimulated research to advance by leaps and bounds. Thus, in addition to the experimental fields mentioned above, in which geographic considerations are the most essential, microclimatology also requires experimental areas well equipped with instruments, perhaps attached to observatories or universities. An excellent example of such a setup is the laboratory of climatology of the Johns Hopkins University in New Jersey, under Thornthwaite.

Finally, we might point out that it is of the greatest importance to the development of microclimatology that future research be conducted in the largest possible variety of the macroclimates of the earth. Our knowledge has already been extended greatly by the numerous measurements made in India by Ramdas, Ramanathan, and others, as well as by the observations which Haude [2, 14] made during the last Sven Hedin expedition in the Gobi Desert; scattered measurements made in tropical, arid, and arctic regions have raised many new questions. Here lies a great future for basic research in microclimatology.

REFERENCES

1. Albrecht, F., "Thermometer zur Messung der wahren Lufttemperatur." *Meteor. Z.*, 44: 420–424 (1927).
2. —— *Ergebnisse von Dr. Haudes Beobachtungen.* Rep. Scient. Exped. to NW Prov. China under Leadership Dr. Sven Hedin, Vol. IX, Met. 2, Stockholm, 1941.
3. Bartels, J., "Verdunstung, Bodenfeuchtigkeit und Sickerwasser." *Z. Forst- u. Jagdw.*, 65: 204–219 (1933).
4. Bruch, H., "Die vertikale Verteilung von Windgeschwindigkeit und Temperatur in den untersten Metern über der Wasseroberfläche." *Veröff. Inst. Meeresk. Univ. Berl.*, Heft 38 (A) (1940).
5. Buxton, P. A., *Insects of Samoa*, Part IX, Fasc. 1. British Museum, London, 1930. (See p. 10)
6. Diem, M., "Messungen der Feuchtigkeit in der bodennächsten Schicht." (Preliminary Rep., *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 12, S. 266 (1950).)
7. Flower, W. D., "An Investigation into the Variation of the Lapse Rate of Temperature in the Atmosphere near the Ground at Ismailia, Egypt." *Geophys. Mem.*, Vol. 8, No. 71, pp. 5–87 (1937).
8. Frankenberger, E., "Über den Austauschmechanismus der Bodenschicht und die Abhängigkeit des vertikalen Massenaustausches vom Temperaturgefälle nach Untersuchungen an den 70 m hohen Funkmasten in Quickborn/Holstein." *Ann. Meteor.*, Beiheft, SS. 3–23 (1948).
9. Franssila, M., "Mikroklimatische Untersuchung des Wärmehaushaltes." *Mitt. meteor. Zent-Anst. Helsingf.*, No. 20 (1936).
10. Friedrich, W., "Messung der Verdunstung vom Erdboden." *Dtsch. Forsch.*, 21: 40–61 (1934).
11. Geiger, R., *The Climate near the Ground.* Cambridge, Mass., Harvard University Press, 1950.
12. —— *Das Klima der bodennahen Luftschicht*, 3. dtsch. Aufl. Braunschweig, F. Vieweg & Sohn, 1950.
13. —— und Fritzsche, G., "Spätfrost und Vollumbruch." *Forstarchiv.*, 16: 141–156 (1940).
14. Haude, W., *Ergebnisse der allgemeinen meteorologischen Beobachtungen und der Drachenaufstiege an den beiden Standlagern bei Ikengueng und am Edsen-gol 1931/32.* Rep. Scient. Exped. to NW Prov. China under Leadership Dr. Sven Hedin, Vol. IX, Met. 1, Stockholm, 1940.
15. Ihne, E., "Phänologische Karte des Fruehlingseinzuges in Mitteleuropa." *Petermanns Mitt.*, 51: 97–108 (1905).
16. Johnson, N. K., "A Study of the Vertical Gradient of Temperature in the Atmosphere near the Ground." *Geophys. Mem.*, Vol. 5, No. 46, 32 pp. (1929).
17. Kaempfert, W., "Einfluss der Pflanzenrichtung, -weite und -höhe auf die Besonnungszeit und -dauer." *Biokl. Beibl.*, 10: 148–153 (1943).
18. —— "Ein Phasendiagramm der Besonnung." *Wetter und Klima* (in press).
19. Keil, K., "Frostbekämpfung im hohen Norden." *Meteor. Rdsch.*, 1: 40–41 (1947).
20. Kessler, O. W., und Kaempfert, W., "Die Frostschadenverhütung." *Wiss. Abh. D. R. Reich. Wetterd.*, Bd. 6, Nr. 2, 243 SS. (1940).
21. Mal, S., Desai, B. N., and Sircar, S. P., "An Investigation into the Variation of the Lapse Rate of Temperature in the Atmosphere near the Ground at Drigh Road, Karachi." *Mem. Indian meteor. Dept.*, Vol. 29, Part 1 (1942).
22. Naegeli, W., "Über die Bedeutung von Windschutzstreifen zum Schutze landwirtschaftlicher Kulturen." *Schweiz. Z. Forstw.*, 11: 265–280 (1941).
23. —— "Untersuchungen über die Windverhältnisse im Bereich von Windschutzstreifen." *Mitt. schweiz. Anst. forstl. Versuchsw.*, 23: 221–276 (1943).
24. —— "Weitere Untersuchungen über die Windverhältnisse im Bereich von Windschutzstreifen." *Mitt. schweiz. Anst. forstl. Versuchsw.*, 24: 657–737 (1946).
25. Ramdas, L. A., (See Monin, A. U., "A New Simple Method of Estimating the Moisture of the Soil *in situ*." *Indian J. agric. Sci.*, Vol. 17, Pt. 2 (1947).)
26. Roll, H.-U., "Über die vertikale Temperaturverteilung in der wassernahen Luftschicht." *Ann. Meteor.*, 1: 353–360 (1948).
27. Schmidt, W., "Observations on Local Climatology in Austrian Mountains." *Quart. J. R. meteor. Soc.*, 60: 345–352 (1934).
28. Schnelle, F., "Kleinklimatische Geländeaufnahme am Beispiel der Frostschäden im Obstbau." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 12, SS. 99-104 (1950).
29. ——"Studien zur Phänologie Mitteleuropas." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 2 (1948).

30. Schüepp, W., "Frostverteilung und Kartoffelanbau in den Alpen auf Grund von Untersuchungen in der Landschaft Davos." *Schweiz. landw. Mh.*, SS. 57–59 (1948).
31. Thornthwaite, C. W., *Micrometeorology of the Surface Layer of the Atmosphere*. Johns Hopkins Univ. Lab. of Climat., Interim Reps. Nos. 1–10, Seabrook, N. J. 1946–50.
32. —— and Halstead, M. H., "Note on the Variation of Wind with Height in the Layer near the Ground." *Trans. Amer. geophys. Un.*, 23: 249–255 (1942).
33. Thornthwaite, C. W., and Holzman, B., "The Determination of Evaporation from Land and Water Surfaces." *Mon. Wea. Rev. Wash.*, 67: 4–11 (1939).
34. Ullrich, H., und Mäde, A., "Studien über die Ursache der Frostresistenz, I—Untersuchungen des Temperaturaustausches an Rizinusblättern durch Messung der Oberflächentemperatur." *Planta*, 28: 344–351 (1938).
35. Weger, N., "Die vorläufigen Ergebnisse der bei Geisenheim begonnenen kleinklimatischen Geländeaufnahme." *Meteor. Rdsch.*, 1: 422–423 (1948).

GEOLOGICAL AND HISTORICAL ASPECTS OF CLIMATIC CHANGE

By C. E. P. BROOKS[1]

Ferring, Sussex, England

THE FACTS AS KNOWN AT PRESENT

The Mild Climates of the Greater Part of Geological Time. The oldest known rocks have been dated by the uranium-lead ratio as having been formed about 1600 million years ago, but the evidences of climate given by these very early deposits are scanty, and it is not until nearly the beginning of the Cambrian period, about 500 million years ago, that a picture of world climate begins to emerge. (The succession of geological periods is shown at the bottom of Fig. 4.) All we can say of the pre-Cambrian period is that at intervals of a few hundred million years glaciers or ice sheets covered various parts of the world; of the intervening periods we can say nothing. From the beginning of the Cambrian onwards our knowledge of the general level and zonal distribution of temperature becomes increasingly detailed, and it is quite clear that climate has alternated between mild and glacial, but that mildness has prevailed for nearly nine-tenths of the time. It seems appropriate, therefore, to begin this review of geological climates with a study of the warm periods. Two epochs may be selected as typical, the Jurassic and the Eocene.

The Jurassic was the age of corals which extended into fairly high latitudes, indicating that in 50°–60°N the temperature of the water in shallow seas was about 60F, or nearly 10F higher than the highest present-day ocean temperatures in those latitudes. Nearer the poles, however, corals were dwarfed or absent, and the general assemblage of animals differed from that in the tropics. This shows that climatic zones existed at that time, though they were less marked than at present. We know less about the climate of the land areas, but numerous salt beds in lake deposits point to a rather scanty rainfall which most probably fell in heavy showers of short duration instead of in prolonged cyclonic storms. There must have been sufficient vegetation to support the dinosaurs and other great land reptiles, but this was probably limited mainly to the river valleys. There is no sign of ice anywhere in the world; even the polar regions were too mild for ice sheets to form, and the absence of mountain ranges prevented the formation of glaciers.

In the Middle and Upper Eocene the climate was generally similar to that of the Jurassic, except that the rainfall of temperate regions was heavier. Land vegetation was abundant as far north as northern Greenland, and Chaney [8] has given us a clear picture of the climatic zones. The most northerly flora known was near Cape Murchison in Grinnell Land (72°N) and included horsetail, yew, pine, spruce, poplar, birch, hazel, and grass—a cool-temperate flora but one which was sufficiently remarkable when we think of the barrenness of the region at present. Everywhere the northern boundary of the temperate flora appears to have been 15° to 20° nearer the pole than at present. At the same time the subtropical flora also advanced northward by about 10° of latitude, extending well into the United States and in places almost reaching the Arctic Circle. The subtropical as well as the polar regions appear to have been warmer than they are today, but the difference decreased from high to low latitudes. The subtropical flora differed completely from that of the Arctic, indicating well-marked climatic zones.

The rainfall of temperate regions in the Eocene probably exceeded the present rainfall. In Europe it occurred mainly in winter and early spring, giving place to a long, hot, and dry summer—the Mediterranean type of climate. The vegetation often shows damage by hail, even quite early in spring, pointing to instability showers. In western United States, however, where the relief was more pronounced, rainfall was more evenly distributed through the year and may have averaged seventy inches.

The picture we form of the climate of middle latitudes during the warm periods is one of mild winters and hot summers, and the absence of cyclonic depressions and gales. The "polar front" as we know it either did not exist or lay far to the north. The subtropical anticyclones lay farther north than now. The arctic basin had prevailing west or southwest winds and mild humid weather with probably a good deal of rain, favouring a rich vegetation. Warm ocean currents extended into high latitudes, and zonal differences of temperature were reduced to a minimum.

Ice Ages. The great "Ice Ages" stand out in sharp contrast to the long mild periods. The climate of large areas of the earth was very severe, but it was also very changeable from one millennium to another. The latest and best-known, the Quaternary Ice Age, began roughly a million years ago, and in this (geologically speaking) short period there have been four major advances and retreats of the ice, as well as a number of minor oscillations. It is convenient to distinguish between the *Ice Age* as a whole, and the *Glacial* and *Interglacial Periods* into which it is divided. The details of the latter of these periods are described by R. F. Flint in this volume,[2] but as they are important for the study of the causes of glaciation, they are briefly recapitulated below. In North America, northern Europe, and probably also parts of northern Asia, great sheets of inland ice, thousands of feet thick, spread from high ground over the plains and even crossed the shallow epicontinental seas;

1. Retired from the Meteorological Office, London.

2. Consult "Climatic Implications of Glacier Research" by R. F. Flint, pp. 1019–1023.

ice from Scandinavia, for example, spread over eastern Britain and ice from Scotland occupied northern Ireland. In North America the southern margin of the ice at its greatest extension was everywhere south of the Canadian border and in the Mississippi Valley it reached almost to 37°N. In Europe the Scandinavian ice extended as far south as 50°N, and large piedmont glaciers spread out from the Alps. Mountain glaciers also reached far below their present limits in the Himalayas and the mountain ranges of the U.S.S.R., in the Rockies and coastal ranges of North America, the Andes, and New Zealand, and even on the equator in East Africa. Glaciers formed on many mountains now ice-free, including those in New Guinea and southeastern Australia. The latest estimates of the maximum area occupied by land ice during glacial periods are about 13 million square miles, including 5 million in the Antarctic, 4½ million in North America, 1¼ million in Europe and at least as much in Asia, and over 800,000 square miles in Greenland. The remainder is made up of a number of relatively small areas. The present ice-covered area is about 6 million square miles, almost entirely in the Antarctic and Greenland. In addition there was a great extension of floating ice in the oceans, especially the North Atlantic and probably also in the Antarctic. Altogether, nearly one-tenth of the earth's surface must have been ice-covered.

It is not likely that the ice everywhere reached its maximum extension and thickness simultaneously, since there was undoubtedly some migration of the centres of maximum accumulation. Flint [11] estimates that if the maxima were everywhere contemporaneous, sea level would have fallen by about 390 ft below the present level. This is a maximum figure, and is higher than previous estimates; it would have been considerably reduced by isostatic adjustment due to the weight of the ice. Shore deposits, coral reefs, etc., point to a lowering of sea level by more than 260 ft. The two figures are sufficiently close to provide good evidence that all the major ice sheets were in existence at the same time.

During glaciation, the snow line over the whole world was lowered by an average of about 2500 ft, more in the snowier and less in the drier regions. If this were due entirely to a fall of temperature, it would indicate a world temperature more than 7F below the present level. Probably the lowering of the snow line was due in part to heavier snowfall, but we can safely put the mean temperature at the peak of the Quaternary glaciation as at least 5F below the present.

It is difficult to estimate the mean global temperature at the peak of a warm period. In the Arctic it was enormously higher than it is now, and was probably of the order of 50F, but in lower latitudes the rise was much smaller. A rough estimate of the increase for the world as a whole would be 10F, which is not excessive considering that we are still in an Ice Age, though in an interglacial period. Thus we arrive at a difference of at least 15F as a measure of the contrast between a mild period and an Ice Age; less near the equator and much more near the poles.

Table I. Glacial Sequence, Quaternary Ice Age

	Alps [27] (thousands of years ago)	North German Plain	North America	Dating by Zeuner [36] (thousands of years ago)		
Glacial	Wurm (40-18)	Weichsel Warthe	Wisconsin { Mankato, Cary, Tazewell, Iowan }	Late glaciation	III II I	25 72 115
Interglacial	Riss-Wurm		Sangamon			
Glacial	Riss (130-100)	Saale	Illinoian	Penultimate glaciation	II I	187 230
Interglacial	Mindel-Riss		Yarmouth			
Glacial	Mindel (430-370)	Elster	Kansan	Antepenultimate glaciation	II I	435 476
Interglacial	Gunz-Mindel		Aftonian			
Glacial	Gunz (520-490)	?	Nebraskan	Early glaciation	II I	550 590

The Quaternary Ice Age was not a single uninterrupted episode, but consisted of a number of advances and retreats of the glaciers and ice sheets. Penck and Brückner [27], in their classic researches into the glaciation of the Alps, recognised four main advances separated by recessions to a climate not more unfavourable than the present. A similar arrangement has been found in North America, but in northern Europe the earliest glaciation has not been recognised, probably because its remains were swept away by subsequent more extensive ice sheets. The sequence is shown in Table I. In some areas the Mindel and in others the Riss glaciation was the most extensive; the Wurm was nearly everywhere much smaller. The extent of the Gunz is not well known. The Wurm glaciation included three or four main readvances and the Riss at least

two. There seems little doubt that the main glaciations were contemporaneous on both sides of the North Atlantic. In most other parts of the world only two or three glaciations have hitherto been recognised, but even on the equator in East Africa three separate extensions of the mountain glaciers can plausibly be equated to the Mindel, Riss, and Wurm glaciations. In Kashmir four distinct glaciations have been recognised; the fourth includes four glacial substages. Here the second interglacial period was the longest and the third the shortest.

Penck and Brückner [27] made the first approach to a time scale for the Quaternary glaciation. Their methods were crude but effective, and their results have never been seriously challenged on geological grounds. Their dates, in thousands of years before the present, are shown in the left-hand column of Table I. It may be remarked that de Geer[3] estimated that the Wurm ice sheet left the coast of Germany about 18,000 B.C. The durations of the various glacial periods naturally varied from place to place, being longest near the centres of the ice sheets and shortest near the peripheries. Thus in Iowa and Ohio the latest glaciation ended roughly 25,000 years ago, whereas in Sweden the duration of postglacial time is taken as only 8500 years. Recent estimates quoted by Flint [11] suggest that the durations of the interglacial periods may have been somewhat longer than the figures given by Penck and Brückner.

In the last few years striking evidence of the succession of glacial and interglacial periods has been recovered from the floor of the Atlantic Ocean by the Swedish Deep-sea Expedition under the leadership of Professor Hans Pettersson, using a special "corer" capable of extracting cores up to 50 ft long from great depths. These cores are estimated to include the deposits formed during a period of at least a million years. An account of the results of this investigation is given by Ovey [26]. Glacial periods are represented by beds rich in erratic debris, showing that icebergs drifted at least as far south as 30°N. Interglacial periods are represented by foraminiferal oozes between the beds of debris. The species of foraminifera indicate the temperature of the surface water very clearly; in the tropics the alternation of glacial and interglacial periods is shown by the alternation of layers with cool and warm species.

A core obtained in the Caribbean Sea shows a succession of four "glacial" periods which have been provisionally correlated with the Nebraskan, Kansan, Illinoian, and Wisconsin. The Nebraskan had three substages, the Kansan and Illinoian two each, and the Wisconsin four or five, but the most interesting point is that the second interglacial (Yarmouth stage) was about twice as long as the first (Aftonian) and three times as long as the third (Sangamon).

Pluvial Periods. Outside the borders of the glaciated regions, the rainfall during glacial periods was in most places appreciably greater than at present. In the middle latitudes of the Northern Hemisphere this was no doubt due to the fact that ice sheets deflected the storm tracks southwards, so that the Mediterranean regions, for example, had a rainfall two or three times that at present, distributed fairly evenly through the year. The Mediterranean storms continued into southwest Asia; v. Ficker [10] calculated that at the maximum glaciation the rainfall in northwest Pamir was four or five times greater than that at the present time. The change was most notable on the northern margins of the subtropical deserts: wandering storms penetrated into the Sahara and this area, now desert, was one of the main centres of population. Similar conditions prevailed south of the main ice masses of North America, the lakes of the Great Basin spreading to form large inland seas, the best known of which are Lakes Lahontan and Bonneville [32].

The most remarkable development of the pluvial periods took place in East Africa, near the equator, where a succession of great lakes grew from the union of a number of existing small lakes, left their deposits, and disappeared. The pluvial periods in nonglaciated regions coincided with the advances of the mountain glaciers, and almost certainly represent the glacial periods of Europe and North America. A very early lake left deposits, now fragmentary, termed Kafuan; Wayland [33] thinks that this lake had two maxima separated by a period of earth movements, and he provisionally equates the lake to the Gunz and Mindel glaciations. This was followed, after a long dry interval, by the "Great Pluvial" in which the very large Lake Kamasia was formed; Nilsson [25] equates this pluvial to the Riss glaciation. Lake Kamasia then dried up completely, and at the same time the mountain glaciers disappeared. Then followed a period of renewed lake formation, the Gamblian pluvial, less intense than the Kamasian, since the lakes did not overflow their separate basins. Nilsson distinguishes four successive Gamblian lake systems, probably corresponding to the three maxima of the Wurm glaciation and a late-glacial readvance. Following the close of the Quaternary Ice Age there have been a number of minor fluctuations of lake levels, which appear to correspond with minor advances and retreats of the mountain glaciers, and these fluctuations present a succession very similar to the postglacial history of northern Europe, though the exact correlation is not yet defined.

Closed lake basins are very sensitive to variations of rainfall. From botanical evidence Moreau [23] estimates the average rainfall in the last of the Gamblian (Wurm) stages as 44–50 inches, while during the postglacial dry period, when the lakes dried up completely, it cannot have been as low as 27 inches. The present rainfall in this region averages 37½ inches. On this scale even the Great Pluvial period may have had a rainfall less than twice the present average. Nevertheless these changes indicate climatic disturbances of large magnitude, which are important for the theory of climatic oscillations.

Earlier Ice Ages. Great ice ages have recurred at intervals of roughly 250 million years—the Late Pal-

3. For a general summary of de Geer's numerous papers see *Geogr. Ann. Stockh.*, 16: 1–52 (1934).

aeozoic (Permo-Carboniferous or Permian) about 250 million years ago; the Upper Proterozoic and earliest Cambrian about 500 million; and two or three in the early Proterozoic probably from 700 to 1000 million years ago. The Proterozoic Ice Ages are now represented only by scattered deposits, mostly in temperate regions. Early Cambrian or uppermost pre-Cambrian glacial deposits are found in small areas in the Lake Superior region and possibly in Utah, in the Yantze valley of China, near Simla (India), extensively in South Africa where they extend to within 29° of the equator, and in eastern and southern Australia. This glaciation was apparently similar to but more severe than that of the Quaternary. These early glaciations are chiefly of interest because they effectively dispose of the theories of a cooling earth.

The Late Palaeozoic Ice Age, on the other hand, reached its greatest development in low latitudes on both sides of the equator, but mainly in the Southern Hemisphere. A great continent extending from South America across Africa and India to Australia carried true ice sheets which reached the sea in many places. There were large mountain glaciers in the eastern United States, but at the same time there was a rich valley vegetation in Europe, Asia, and North America resulting in the formation of thick beds of coal, even as far north as Spitsbergen. This peculiar distribution of climatic zones has been a great puzzle to geologists and climatologists, and was one of the chief bases of the theory of continental drift. A plausible explanation can be given in terms of the peculiar geography of the period, but this is deferred until the geographic factors of climate have been considered (see p. 1014).

Between the major ice ages there is occasional evidence of local glaciation, but not all of it is accepted by geologists. The most important of these are (1) late Silurian and early Devonian in Alaska, eastern Canada, South Africa and, possibly southeast of Kashmir, and (2) late Cretaceous and early Eocene in the Cordilleras of North America and in the Antarctic. These periods of glaciation occur about midway between the major ice ages.

Postglacial Climatic Changes. The changes of climate since the last maximum of the Wurm or Wisconsin glaciation are important for the theory of climatic change, because during that period the changes in land and sea distribution and in the elements of the earth's orbit were small. Postglacial climatic changes in northern Europe are summarized in Table II and are based mainly on the study by Movius [24]. The dating is based on the work of de Geer on the glacial "varves" or annual layers of fluvio-glacial sediment.

The retreat of the glaciers took place from the periphery inwards, mostly by melting and ablation. They did not decay at the centre, leaving large masses of "dead ice" at the margins to melt gradually, but continued active until they had nearly vanished. Their retreat was interrupted by three halts or slight readvances, which differed from the main substages of the Wurm only in being superposed on a steady withdrawal. In the Alps, Britain, and Ireland there were similar stages which took the form of definite readvances or even new formation of mountain glaciers. In North America there were similar halts in the recession, but these have not yet been definitely correlated with the European sequence.

Table II. Postglacial Succession in Europe

Date (B.C.)	Climatic stage	Climate	Vegetation
18,000–14,800	Pomeranian end moraine	Arctic	Tundra
10,000–8300	Allerød Oscillation	Sub-Arctic–temperate	Pine, sedge, peat
8300–7800	Fenno-Scandian end moraine	Arctic	*Dryas*
7800–6800	Pre-boreal	Dry, cool	Pine, hazel
6800–5600	Boreal Ragunda stadium 6800–6500	Dry, cold winters, warm summers	Alder, oak, elm
5600–2500	Atlantic "Climatic Optimum"	Warm, humid	Peat, oak, alder, lime, elm
2500–500	Sub-boreal	Drier, becoming cooler, variable	Oak, giving place to pine
500–0	Sub-Atlantic	Cool, wet	Peat, beech

The Scandinavian geologists consider that the Ice Age ended about 6500 B.C., when the last remnants of the ice sheet split into two parts, but by this date the climate of most of Europe had become temperate. Near the periphery of the glaciated areas the Ice Age ended much earlier (about 20,000 B.C.), while Greenland and the Antarctic are still in the Ice Age, so that the choice of the date 6500 B.C. is arbitrary.

The climate of the early postglacial period in Europe was continental, with hot summers and cold winters. In the sixth millennium B.C. there was a change to a warm humid climate, with a mean temperature up to 5F higher than the present mean and a heavy rainfall which caused a considerable growth of peat. This is known as the *Climatic Optimum*. In Scandinavia it was accentuated by subsidence of the land, which permitted a greater influx of warm Atlantic water into the enlarged Baltic known as the *Litorina Sea*, but the Climatic Optimum was so widespread—probably worldwide—that this cannot have been the only cause. Judging by the flora of Spitsbergen, the Arctic Ocean was free of ice.

The sub-boreal climate was peculiar. On the whole there was a gradual decrease of temperature and rainfall from the Climatic Optimum, but this was interrupted by long droughts in which the surface of the peat dried up, followed by returns to more rainy conditions. This alternation occurred several times, the main dry periods falling about 2200–1900, 1200–1000, and 700–500 B.C. The latter, which is termed the "Grenzhorizont," was the best-developed and caused

a widespread interruption in the growth of peat in Europe. It has been described as a "dry heat wave" lasting for perhaps 200 years. Lakes decreased in area and in a few places trees grew on their floors below the level of the outlet. From four such lakes in Ireland, Germany, and Austria it is estimated that the annual rainfall was only about half the present amount. The drought was not sufficiently intense to interrupt the steady development of forests, but it caused extensive migrations of peoples from drier to wetter sites. About 1300 B.C. there was a period of heavy rainfall and floods which destroyed some of the Alpine lake villages.

About 500 B.C. there was a great and rapid change to a colder and wetter climate. Over large areas the forests were killed by a rapid growth of peat. The levels of the Alpine lakes rose suddenly, flooding many of the lake dwellings, and most of the mountain area became uninhabitable, the few settlements being limited to the warmest and driest valleys. Traffic across the Alpine passes, which had continued steadily since 1800 B.C., came to an end. As many of the peat bogs formed during this period are now drying up, it seems that the rainfall of this sub-Atlantic period must have been greater than at present. This change of climate was by far the greatest and most abrupt since the end of the Ice Age, and its effect on the civilisation of Europe was catastrophic. It did not last long, however, for by the beginning of the Christian era conditions did not differ much from the present.

Our knowledge of postglacial conditions in other parts of the world is less detailed, but the climatic changes appear to have run closely parallel. In Asia and northeast Africa the evidence consists mainly of the migration of whole peoples, and of the occupation and abandonment of marginal sites, where under average conditions there is only just enough water to support life. In North Africa, West and Central Asia, and China there was a wet period between 5000 and 4000 B.C. In China the cultivation of rice extended about five degrees north of its present boundary, pointing to a higher temperature. This warm wet period was followed by a period of gradually increasing drought which culminated about 2200 B.C. Conditions remained dry until 1900 B.C., when there was a "complaint" that the marshes of the Nile delta had dried up and the river could be crossed on foot, presumably during the low-water stage. Then followed a period of fluctuating rainfall, with another drought which the Chinese records place between 842 and 771 B.C. Migration and war almost ceased after 500 B.C., when caravans crossed deserts which are now almost impassable. A passage in Herodotus suggests that the level of the Caspian Sea stood much higher than now, and there is evidence of abundant water supply in Egypt and in the Saharan oases. On the whole, however, the case for a great increase of rainfall about 500 B.C. is not so strong in Asia and North Africa as it is in Europe.

There is abundant evidence of postglacial climatic change in North America, but until the tree rings take up the story about 1000 B.C., the dating is uncertain. In eastern North America the rate of recession of the ice sheets at the close of the glaciation ran parallel with the recession rate in northwest Europe, and the peat bogs show similar alternations of wet and dry periods; there seems no reason to doubt that the changes were approximately synchronous. Hansen [12] gives the following succession:

Eastern North America	*Northwest Europe*
Ice retreat (Hudsonian)	Late-glacial
Spruce, fir (cool, moist)	Pre-boreal
Pine (warmer but still cool)	Boreal
Oak and hemlock (warm, moist)	Atlantic
Oak and hickory (warm, dry)	Sub-boreal
Oak, chestnut, spruce (cooler, moister)	Sub-Atlantic

The history of the lakes of the Great Basin, described by Jones [32] and van Winkle [35], points to a long dry period which, according to the salt content of the lakes, ended some time between 2000 B.C. and A.D. 0. This period was brought to an end by a rainfall greater than that of the present and was followed by a gradual decrease.

The evidence of the tree rings is unfortunately doubtful, as comparatively few trees date from before 500 B.C. The curves drawn by Huntington and Antevs [32] both indicate a wet period which had definitely begun by 660 B.C. and reached an absolute maximum between 480 and 250 B.C. All these facts are in sufficiently good agreement with the dating of the sub-Atlantic in Europe at 500 B.C.

The fluctuations of the lakes in equatorial Africa appear to run parallel with the alternations of wet and dry periods in Europe, though they cannot be dated precisely. On the other hand, in the dry subtropical regions south of the main storm tracks the evidence for climatic change is much less definite. In the temperate parts of the Southern Hemisphere the postglacial variations of climate appear to have been generally similar to those of the Northern Hemisphere, but they have not yet been dated. In southern New Zealand the Climatic Optimum is represented by lowland rain forest, indicating increased rainfall and higher temperature, but in the southern Argentine in the rain shadow of the Andes the increased warmth was accompanied by greater aridity.

Climatic Changes during the Christian Era. After the beginning of the Christian era we have an increasingly detailed and accurately dated knowledge of climatic changes. The evidence includes fluctuations of lakes and rivers, growth of peat bogs, succession of floras, rate of growth of trees as shown by annual rings, advances and retreats of glaciers, locations of settlements and migrations of people (for which climatic reasons may be assigned with some show of probability), literary records and old weather journals, and finally instrumental records. The results of this great mass of detail, summarised in Table III, show that climatic oscillations are not entirely local events but tend to fit together into a world pattern. Asia and western North America tend to vary together; Europe partly varies with them but has changes of its own in addition. North Africa, represented principally by the Nile valley, varies partly with and partly against Asia. An opposition between the temperate and subtropical regions is shown more clearly in Yucatan, as was first pointed

out by Huntington [15]. At present that country is covered by dense forests and the hot moist climate is enervating and inimical to a high culture. Buried in the forests are the ruins of magnificent Mayan cities, and it is inconceivable that these could have been built under present conditions. The first period of culture lasted from about 400 B.C. to A.D. 300. From A.D. 300 to 450 the forest encroached from the south, and from 450 to 900 it extended over the whole country; culture declined to a low level. There was a climatic recrudescence from 900 to 1100, deterioration from 1100 to 1300, and a slight improvement from 1300 to 1450, after which conditions were continuously unfavourable. The dry periods in Yucatan coincide almost exactly with the rainy periods farther north.

Table III. Climatic Variations during the Christian Era

A.D.	Europe	Asia	Western North America	Africa
0	As present	Slightly rainier than now	As present	Good Nile floods
100	Somewhat drier	Rainy		Drier
200	Rainy			
300		Dry	Dry	Rainier
400		Less dry Caspian −15 ft	Dry	Rainy
500	Drier	Dry		Rainy
600	Rather dry		Slightly rainier	
		Rainfall increasing		Rainy
700	Dry, warm	Rainy	Drier	Drier
800			Dry period ended	Dry
	Rainier	Rainy in China		
900		Caspian + 29 ft	Rainier	Rainier
	Drier		Slightly drier	
1000	Colder	Dry in China	Very rainy	Drier
1100	Heavy rain	Dry Caspian −14 ft	Dry	Very dry
1200	Rainy Very stormy	Dry Rainfall increasing	Dry	Rainy
1300	Glacial advance, drier	Rainy, Caspian, etc. high	Rainy	Rainy
1400	Glacial min.	Dry in China	Dry	Rainy
1500	Oceanic Continental	Rainy, Caspian +16ft		Rainfall maximum
1600	Rapid advance of glaciers	Rainy Caspian +15 ft	Rainier	Drier
1700	Dry in west Glacial max.	Near present Caspian rather high		
1800	Cold, rainier		Rainy	
1900	Rapid retreat of glaciers	Caspian falling	Drier	

The dry periods in Yucatan coincide almost exactly with the rainy periods farther north.

Special mention must be made of the advance of the glaciers in the Alps, Scandinavia, and Iceland, which began near the middle of the 16th century and was so striking that it has come to be known as the "Little Ice Age." It may be divided into six periods:

1. First advance, A.D. 1550–1650 (first maximum). About 1605, glaciers overran settlements which had been occupied since early days.
2. Recession, 1650–1680, followed by fluctuations about a generally small extent until 1715.
3. Rapid advance to a maximum about 1750–1760, which in many districts marks the greatest extension of glaciers since the Ice Age.
4. Retreat until about 1790, but with fluctuations.
5. Advance to a third maximum about 1850, with a minor maximum about 1815.
6. General retreat, interrupted about 1890, then becoming increasingly rapid. In Norway and Iceland the glaciers have not yet retreated to the positions they occupied in the 14th century.

Generally speaking, the glacial advances were associated with cold winters and cool springs, and the retreats with strong westerly winds and a mild maritime climate. The variations do not appear to be related to variations of total precipitation, but most probably indicate variations in the length of the period of ablation.

Similar but less accurately dated fluctuations have occurred in glaciers in other parts of the world. In Spitsbergen they reached a maximum about the middle of the 19th century and have since retreated or become stagnant. The marginal glaciers of northeast Greenland have been receding since the late 18th or early 19th century. Alaskan glaciers for the most part reached maxima in the 18th or 19th century and have since receded, but some are still advancing, probably because their sources are at considerable heights.

The broad pattern of climatic change since the end of the Ice Age is consistent with the hypothesis of an alternate weakening and strengthening of the planetary atmospheric circulation, associated with a poleward and equatorward shift of the wind zones. At times of minimum circulation the circumpolar vortex is contracted and anticyclones are frequent in middle latitudes. Winds are variable, rainfall is small, and climate continental, with cold winters and hot summers. This was the general situation during the long dry period from about A.D. 400 to 1000. When the circulation is stronger, westerly winds predominate, rainfall is

heavier, and climate more oceanic. The shifts of the wind zones may have been associated to some extent with fluctuations in the ice-covering of the arctic seas. References in the works of classical geographers suggest that near the beginning of the Christian era Iceland was icebound. During the period of the Norse colonisation of Greenland there was little ice, but glacial conditions returned about 1200 and remained predominant throughout the "Little Ice Age"; they are now improving. The parallelism is not exact, however, and in the past 2000 years it is more likely that the variations of atmospheric circulation have governed the variations of arctic ice than that the reverse is the case.

THEORIES OF CLIMATIC CHANGE DEPENDENT ON EXTERNAL INFLUENCES

Variations of Solar Radiation. The most obvious reason for climatic changes is that the radiation which the earth receives from the sun varies with time. It has frequently been suggested that ice ages were due to decreased solar radiation, either because the sun emitted less radiation, or because some obstacle such as a cloud of cosmic matter was interposed between the sun and the earth. Conversely, the warm periods were attributed to increased solar radiation. In 1929, however, Simpson [30] argued that glaciation should result from an *increase* of solar radiation.

Simpson's theory is, briefly, that glaciation depends not so much on temperature as on an excess of snowfall over melting. The total precipitation over the earth as a whole must equal the total evaporation, which in turn depends very largely on the solar radiation. In high latitudes and at high elevations most of the precipitation falls as snow. Now consider the effect of two cycles of solar radiation (Fig. 1). Starting with a minimum of radiation most of the precipitation falls as snow, but the total amount is small. As the radiation increases, the proportion of snow to rain decreases, but at first this is more than counterbalanced by the increase of total precipitation, while summer melting is still unimportant. At this stage there is an accumulation of snow, resulting in glaciation. Eventually, however, the rise of temperature due to increased radiation reaches a stage at which the actual amount of snowfall begins to decrease and is exceeded by the summer melting. The glaciers and ice sheets break up and a mild rainy interglacial sets in. As the radiation passes its peak and decreases, the process is repeated in the reverse direction, bringing another glaciation, but eventually the decrease of precipitation starves the glaciers and ice sheets, causing a second interglacial, but this time with a cold dry climate, which persists until another increase of solar radiation begins a new cycle.

Simpson considered that the stage of glaciation is rather near the peaks of the solar cycle, so that the warm moist interglacials were short and the cold dry interglacial was long, giving the well-known Alpine succession (p. 1005). In low latitudes each solar cycle is represented by a single cycle of rainfall, giving two pluvial periods covering respectively Gunz plus Mindel and Riss plus Wurm.

The facts do not entirely support this theory. The Mindel-Riss interglacial was not cold; at its height Europe and North America were warmer than at present. In East Africa the Riss and Wurm were not combined and Wayland [33] thinks that the first pluvial also was double. A third objection, that there is no physical reason for the existence of a cycle of solar radiation of the order of 500,000 years, has been partly met by Hoyle and Lyttleton [13], who showed that the passage of the sun through a cloud of interstellar matter would cause an increase of solar radiation.

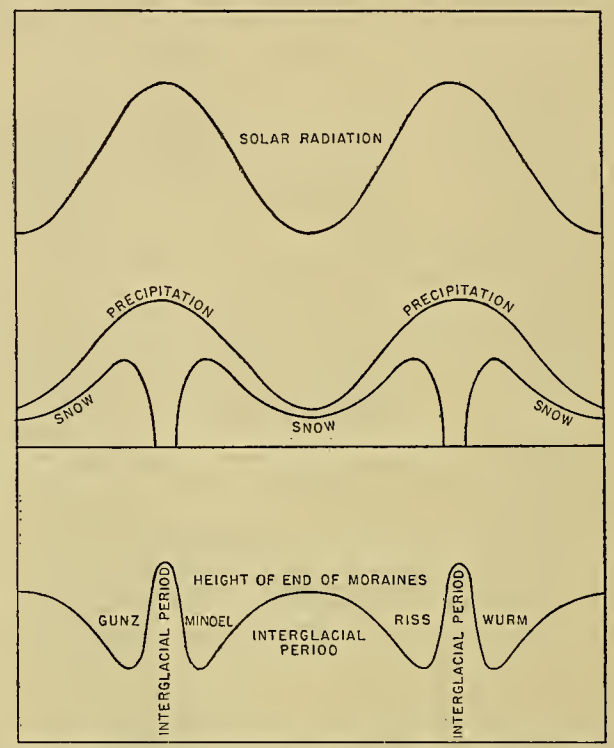


FIG. 1.—Effect of two cycles of solar radiation on glaciation. (*Reproduced by courtesy of Sir George Simpson and the Manchester Literary and Philosophical Society.*)

Willett [34] has pointed out that some of the objections to Simpson's theory are removed if we suppose that the Quaternary represents four solar maxima instead of two, the radiation never reaching the level necessary for a warm interglacial, and also that the cycle of radiation takes place mainly in the ultraviolet.

Another theory of solar control was put forward by Huntington and Visher [15]. There is some evidence that when sunspots are numerous and large, the storm tracks are displaced towards the equator and the climate of the earth as a whole is stormier, rainier, and cooler than at times of sunspot minimum. Since observations began, sunspots have varied, not only in the well-known 11-year cycle, but also over much longer periods, reaching a high maximum about 1372 and being almost absent from 1676 to 1725. Huntington and Visher suggested that there have been sunspot cycles of geological length, associated with changes in the distances of the nearest fixed stars, and that these have caused the geological changes of climate. This

theory raises many difficulties, and it breaks down completely over the Late Palaeozoic glaciation. It is true that the variations in the rainfall of the north temperate zone since A.D. 1000 have run fairly parallel with the variations of sunspots, but the geological variations of climate were on a scale many times greater and would require enormous and prolonged sunspot outbursts which seem quite improbable.

Changes of solar radiation could account for some, but not all, of the postglacial changes of climate. It is difficult to find any other explanation for the warmth and heavy rainfall of the Climatic Optimum, but the rainfall maximum about 500 B.C. was accompanied by a *fall* of temperature. The alternating halts and retreats of the late glacial period could have been due to solar changes, but other explanations have been put forward. Lewis [21] considered that a glacial period could be initiated by a comparatively small increase of precipitation, which might be due to the temporary diversion of an ocean current, and that once the ice sheets reached a size sufficient for the development of a glacial anticyclone, they grew by a "runaway" process until the increasing loss of solar energy by reflection from the ice and snow lowered world temperature, and consequently the precipitation level, below the subsistence level of the ice sheets. In this connection the recent recession of the glaciers of the "Little Ice Age" is especially interesting. By comparing the decrease in the volume of ice since 1850 with the rise of sea level, Ahlmann [1] found that the recession had been confined to the marginal glaciers, the central ice sheets of Greenland and the Antarctic being almost untouched. There has been no appreciable decrease of snowfall, so that the recession must be due to increased ablation. In the marginal areas ablation is due mainly to heat conduction from warm air, that is, to a stronger atmospheric circulation. On the main surfaces of the ice sheets ablation is due to solar radiation, and Ahlmann infers that the warming of the Arctic is associated with an increased atmospheric circulation without any appreciable change of solar radiation.

Changes in the Elements of the Earth's Orbit. With constant solar radiation, the heat reaching the outer limit of the earth's atmosphere in a year remains practically constant in any latitude, but the seasonal distribution changes. There are three variables:

1. The obliquity of the ecliptic, or the angle which the plane through the equator makes with the plane of the earth's orbit round the sun. The greater the obliquity, the greater the contrast between the heat received in summer and winter. Milankovitch [22] calculated the variation as $2\frac{1}{2}°$ in a period of 40,400 years; the last maximum was about 8000 B.C.

2. The eccentricity of the earth's orbit, which varies from 0.0 to about 0.07 in a period of 100,000 years. The hemisphere with winter in aphelion has a short hot summer and a long cold winter, that with winter in perihelion has a short mild winter and a long cool summer. At present the Northern Hemisphere is in perihelion in winter, but the excess of land quite outweighs the solar effect.

3. The precession of the equinoxes, by which the season in which perihelion falls advances through the year in a period of 21,000 years. About 8500 B.C. the Northern Hemisphere had winter in aphelion and a more extreme solar climate than now.

Many attempts have been made to account for the succession of glaciations by these astronomical changes. The early theories, such as that of Croll, require alternate glaciation in the two hemispheres; they were unsound meteorologically and the astronomical dating did not agree with the geological time scale. These defects in the theory have been overcome to a large extent, and the latest exposition, by Zeuner [36], is *qualitatively* in good agreement with the facts. Zeuner shows that the variation of the present snow line with latitude follows closely the variation of radiation in the summer half-year. A rise of winter temperature increases the snowfall and the corresponding decrease of summer temperature enables the snow cover to persist through the year. Hence small obliquity with high eccentricity and summer in aphelion lead to glaciation. Zeuner's dating of the various glacial stages is shown on the right of Table I, and is seen to agree reasonably well with the estimates by Penck and Brückner.

Undoubtedly these changes in the seasonal distribution of radiation must have some effect on climate, but they were almost certainly *quantitatively* insufficient to account for the enormous range between glaciation and deglaciation, even when secondary effects are exploited to the full. The variations of radiation are small and complex near the equator, and cannot possibly account *directly* for the alternation of pluvial and interpluvial periods. Moreover, as Zeuner recognises, such factors cannot account for the Ice Age as a whole. They have presumably been continuously in operation throughout geological time, but traces of them are rare. According to Bradley [3] the Eocene of Colorado, Utah, and Wyoming, covering several million years, shows a periodicity of 21,000 years. The Cretaceous in the United States also shows a cycle estimated as of this length. It is possible that the coal seams in the Upper Carboniferous (Pennsylvanian) represent periods of great eccentricity with northern winter in perihelion and a small obliquity of the ecliptic, giving little annual range of temperature; this would account for the absence of annual growth rings. The intercalated sandstones would represent the opposite condition of summer in perihelion.

The continental climate of the pre-boreal, about 8000 to 7000 B.C., fits in with the large obliquity and winter in aphelion, but could equally well have been due to the remains of the ice sheets and, in Europe, to the elevation of Scandinavia which converted the Baltic into the fresh-water *Ancylus* Lake. Astronomical changes cannot possibly account for the Climatic Optimum, the subsequent deterioration in the sub-Atlantic, and the "Little Ice Age" of the 17th to 19th centuries A.D. The trend of modern thought is against the astronomical theory.

Tidal Variations. We must mention here a suggestion by Pettersson [28] that changes of climate during

the past few thousand years were due to variations in the tides, especially in the submarine tides at the boundary of the Atlantic and Arctic Oceans, where a thin layer of cold arctic water overlies the warmer, more saline, Atlantic water. According to Pettersson, this "tide-generating" force reached maxima about 3500 B.C., 1900 B.C., 250 B.C., and A.D. 1433. At tidal maxima the arctic icecap is more readily broken up than at tidal minima, and more ice drifts out into the Atlantic. This floating ice, by lowering the surface temperature of the oceans and increasing the local contrasts, shifts the storm tracks southwards and causes an increase in the number and intensity of depressions, consequently tidal maxima should also be maxima of storminess and rainfall. There were in fact rainfall maxima about 2000 B.C. and 500 B.C. and a period of great storminess in the 12th to 14th centuries, but the agreement breaks down before 3000 B.C., during the Climatic Optimum. I think Pettersson's "tide-generating force" may have contributed to the climatic variations since 3000 B.C., but to be effective it requires a nice adjustment of ice conditions in the Arctic such as can have occurred very rarely and, geologically speaking, only for short periods.

THEORIES OF CLIMATIC CHANGE DUE TO TERRESTRIAL CAUSES

The Hypothesis of Continental Drift. Most theories of climatic change rest on the implicit assumption that geological deposits were laid down in the latitudes and longitudes in which they are now found. That assumption was seriously challenged by Wegener [19], who put forward the theory that in geological time the continents have drifted over the earth, moving relatively to each other and also, very widely, relative to the poles. He side-stepped the problem of climatic change completely; to him a decrease of temperature in any district simply meant that that district was moving into higher latitudes. Glacial deposits, wherever they are now, were all formed in high latitudes, the coal measures mark the Carboniferous equator, and so on.

For some time this theory attracted wide support, but difficulties, both geological and meteorological, have multiplied. The statement that the great glacial deposits of late Palaeozoic age in South America, Africa, India, and Australia were formed in a single primeval continent (Pangaea) through which the South Pole followed a wandering course, breaks down under detailed examination. The distribution of temperate floras of early Tertiary age in zones surrounding the present North Pole amounts to a proof that at those times the pole occupied its present position and not a point in the North Pacific as depicted by Wegener. His correlation of Quaternary glaciations is regarded by geologists as impossible. Finally, Wegener himself realised that continental drift cannot explain the succession of glacial and interglacial stages. The theory has no single definite fact to support it, for even the supposed westward drift of Greenland has not been proved; as a theory of geological climates it is now almost obsolete.

The Geographic Control of Climate. The zonal distribution of climates at present is by no means perfect; owing to warm and cold ocean currents and to the positions of the continents, any isotherm may cross many degrees of latitude at sea level. The isotherm of 32F in January, for example, ranges from 35°N in eastern China to 70°N north of Norway. There is also a strong vertical zoning of climate, temperature decreasing upwards at about 3F per 1000 ft. From this it appears that, other things being equal, a geological period with large high continents should be cold and one with wide oceans and low continents broken up into islands should be warm. It will be shown that this geographic factor is quantitatively sufficient.

The Geographic Cycle. The geological record shows that the world has passed through a number of cycles of elevation and erosion. At the close of a long period of quiet, folding of the earth's crust begins. The continents rise and extend to the full limit of the continental shelves, and the ocean floor sinks. Erosion forms great thicknesses of sedimentary deposits which cause further subsidence and folding into great mountain chains, accompanied by volcanic activity. Glaciers form on the mountains and in favourable conditions grow into ice sheets. This locking-up of water reduces the level of the sea still further. Eventually the period of disturbance ends, the mountain ranges reach their greatest elevation and are rapidly worn down by erosion, accentuated in many cases by the glaciers. The general level of the land falls and that of the oceans rises, and the continental borders are again flooded. This brings in a long period of low relief and small land masses which continues until another period of disturbance begins.

Geographic cycles are not all of the same intensity. The major cycles, culminating in widespread glaciation, appear to run their course in about 250 million years, but these tend to be broken by minor cycles which lead to nothing more than local glaciation. The succession of events is shown diagrammatically in Fig. 2.

An important point is that in the Quaternary, and probably in the earlier periods of disturbance also, glaciation was not coincident with mountain building, but lagged some five to ten million years behind it. Various reasons have been assigned for this lag:

1. At the end of a long period of warm climate the oceans are warm, and a long time is required to cool them. As soon as the polar seas froze, or extensive glaciers reached the sea, cooling of the great mass of the oceans would be fairly rapid, but until some ice existed, they would cool very slowly.

2. A smooth dome is not a favourable basis on which glaciers can grow into ice sheets. The mountains must first be eroded into peaks and valleys, with further isostatic elevation.

3. Mountain building alone is not enough for glaciation, which must wait until some other factor, such as a decrease of solar radiation, acts as a trigger.

4. Recently Wagner [31] suggested that mountain building is accompanied by a great release of earth heat, and this raises the temperature of the ground

sufficiently to prevent glaciers from forming; Wagner mentions a figure of 10F. Jeffreys [16], however, states that any appreciable rise of surface temperature from such a cause is not possible.

Mountain Building and Climate. It is now generally agreed that the presence of mountains is a *necessary* condition for glaciation, the mountains forming a gathering ground and nucleus. It remains to be considered whether mountain building alone can cause world-wide cooling which is quantitatively *sufficient* to explain glaciation. Four factors are involved:

1. The decrease of mean temperature with height.
2. Reflection of solar radiation from the surface of clouds.
3. Cooling power of surfaces of ice and snow.
4. Loss of heat owing to increased evaporation.

The average height of the land surface at present is about 2500 ft above sea level. At the maximum glaciation the average elevation may be taken as 3500 ft.

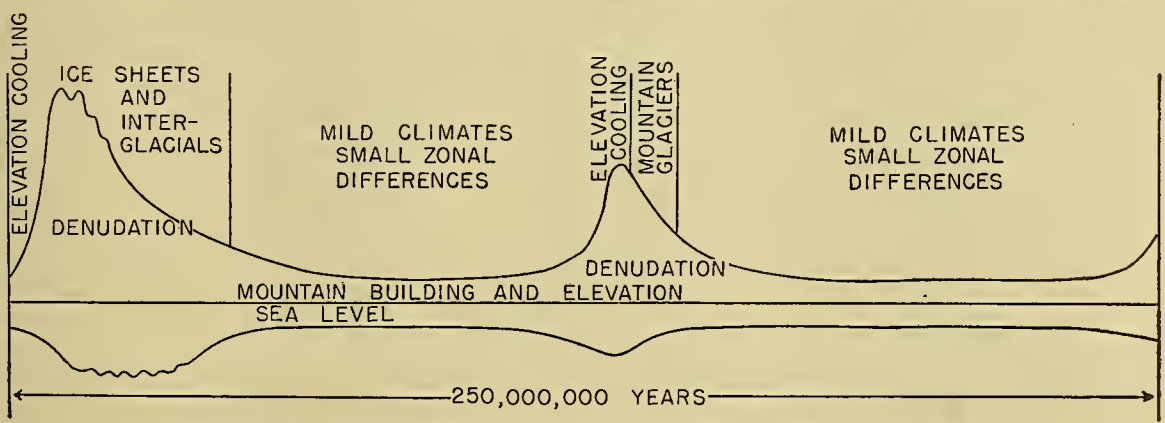


FIG. 2.—Geographic cycle of mountain building, sea level, and climate.

1. Owing to the cooling of ascending air by expansion the decrease of temperature with height is about 3F per 1000 ft. When the air descends again, it is warmed by compression, so that this effect is limited to the actual mountains. An elevation of 1000 ft is equivalent to an *average* cooling of 3F over the land areas, or 0.8F over the world as a whole.

2. Except in winter in high latitudes, clouds lower the mean temperature by reflecting from 60 to 75 per cent of the solar radiation back to space. On the average, an increase of one-tenth in the mean cloudiness lowers the mean temperature by about 6F. The level of condensation varies according to the humidity of the air, but probably averages around 5000 ft over land areas. A mountain range increases the cloud amount to windward and over the crest and decreases it to leeward, but to a less amount. As a rough estimate we may assume an average increase of the cloud amount by three-tenths of the sky. At present 12½ per cent of the land is above 5000 ft, and an increase in the average elevation by 1000 ft is estimated to increase this area to 21 per cent. At the time of maximum elevation, therefore, increased cloudiness would have lowered the mean temperature over the land areas by about 1.5F, or 0.4F over the world as a whole.

3. In calculating the cooling effect of increased areas of snow and ice, we must ignore for the moment the great Quaternary ice sheets, which were a consequence and not a cause of the initial cooling. Hence we must limit the calculation to the area above the snow line and the surface of mountain glaciers. At present this is about 3 per cent of the land surface, more than half of which is between latitude 70° and the poles. When the continents were at their greatest height, this figure was probably rather over 6 per cent. Brooks [5] has calculated that if a land surface in high latitudes, formerly bare, has one per cent of its area covered by ice or snow, the mean annual temperature would be lowered by 0.3F. Hence the increased snow cover would have resulted in an average cooling of about 1F over the continents. Since the Quaternary elevation was greatest in high latitudes, the figure was most probably greater than this, perhaps double. Something must be added for the formation of sea ice in polar waters, and the average figure for the world as a whole may be taken as 1–2F.

4. In the nonglaciated parts of the world the total precipitation, and therefore the total evaporation, during the great ice ages were considerably greater than they are now. The increase was due partly to the greater elevation of the land, and partly to stronger winds. It is difficult to form an estimate of the increase of rainfall, but considering all the evidence, a figure of 20 per cent is probably a minimum. If we put the net fall of surface temperature (after allowing for increased back radiation from the atmosphere) as 6F at present, we find that increased evaporation at glacial maxima would have lowered the temperature by another 1.2F.

Summing up, we find that the lowering of temperature at the maximum elevation in the Quaternary can be estimated as follows:

Decrease of temperature with height	0.8F
Increased cloud amount	0.4
Increased area of snow and ice	1.5
Increased evaporation	1.2

The total is about 4F, averaged over the whole world.

At the end of a long quiescent period the average

elevation may not have exceeded 500 ft. At such times the mean temperature of the world would be expected to be about 9F higher than now, merely because of the lower mean height. It will be seen later that the changes of land and sea distribution associated with mountain building and erosion bring other factors into operation which magnify the effect of elevation.

The Role of Polar Icecaps in Accentuating Climatic Changes. The geographic cycle affects the distribution of heat over the globe by means of ocean currents. Before discussing this factor we must consider the effect of floating ice. If the surface of the ocean were above the freezing point of sea water, there would be no ice. But if a general fall of temperature brought a small area near the North Pole below freezing point, a nucleus of ice would be formed, and this would act as an additional source of cooling, reducing the temperature still further. Brooks [6] showed that once the nascent ice sheet exceeded a diameter of about 250 miles, this additional cooling would exceed the rise of temperature due to decreasing latitude, and the floating icecap would grow until it nearly filled the arctic basin. The freezing point of sea water is 28F. Three independent calculations have shown [4] that if the arctic ice could all be cleared away, the mean winter temperature of the sea surface near the pole would be not far from 24F. This is the "nonglacial" temperature. A permanent rise of the nonglacial temperature by 5F would eventually sweep away the floating icecap and convert the Arctic Ocean into an open sea, with only some thin and scattered floating ice forming in winter and melting in summer.

The present low winter temperature of the Arctic, estimated as −40F near the pole in January, is due almost entirely to the existence of the icecap itself. A small access of heat only sufficient in itself to raise the temperature by 5F would result in an actual rise of nearly 70F, that is, in a complete change of climate which would permit cool temperate vegetation to extend to high latitudes.

There would also be a great change in the atmospheric circulation. The area of relatively high pressure in high latitudes is a surface effect, due to the weight of cold air in the lowest few thousand feet; at a height of 10,000 ft pressure falls continuously from low latitudes to the neighbourhood of the pole. The polar front is the boundary between the surface easterly winds associated with this area of high pressure and the westerly winds of middle latitudes. It is a reasonable inference that the removal of the surface cooling due to the floating ice would remove this layer of cold air, so that westerly or southwesterly winds would extend to much higher latitudes than at present. This in turn would accentuate the warming of the Arctic. Depressions would still occur, but in the absence of marked differences of temperature they would probably be weak; the necessary balance between easterly and westerly winds would be maintained chiefly by an extension of the subtropical anticyclones into higher latitudes.

The differences between completely ice-free and completely glaciated polar regions add up to the difference between the warm periods constituting most of geological time, which we may term "nonglacial," and the "glacial" periods such as the present, which culminated in ice ages. From this discussion it follows that the basic difference between a nonglacial and a glacial period is comparatively small—of the order of 10F in the temperature of high latitudes. Such a change could be brought about by a general rise of temperature over the whole earth, but it could also result from a greater transfer of heat from low to high latitudes by ocean currents.

Ocean Currents. At a time of maximum mountain building, the continents are most extensive and irregular in shape. The free circulation of the oceans between low and high latitudes is restricted and relatively little heat is carried to polar regions. At the end of quiescent periods the continents are small and ocean currents have free access to polar regions along several broad channels. The present situation in the Northern Hemisphere is rather unfavourable in this respect, access to the Arctic being possible only by the gap between Greenland and Europe, and part of this gap is occupied by the cold East Greenland Current, which owes its existence to arctic ice. There is no doubt that the Gulf Stream raises the temperature of the Arctic Ocean by several degrees, but this is not sufficient to keep the Arctic free of ice. In warm periods such as the Jurassic or early Tertiary, the Bering Strait was more open and there was a third channel, the Volga Sea, from the Indian Ocean across western Siberia. Brooks [4], on the basis of calculations by Kerner [18], estimated that the additional supply of heat was sufficient to raise the "nonglacial" temperature by 10–13F, more than enough to keep the Arctic free of ice. That, given the geography of these periods, such currents would exist has been shown experimentally by Lasareff [20] by means of models.

The Late Palaeozoic Ice Age. The greatest problem of geological climates is presented by the Late Palaeozoic—Upper Carboniferous (Pennsylvanian) and Lower Permian—in which an apparently highly favourable climate in the Northern Hemisphere, permitting the rich vegetation of the coal measures, coincided with or only preceded by a short time very extensive ice sheets in low latitudes of both hemispheres. The continental drift theory gives a plausible explanation of this distribution but as we have seen, it suffers from other objections. There remains the possibility that it was due to a peculiar distribution of sea, land, and mountains.

In the Late Palaeozoic a large continent, termed *Gondwanaland*, extended across South America, Africa, India, and Australia. As a result of a period of mountain building, this continent was most extensive and lofty in the Pennsylvanian and Lower Permian. To the north (Fig. 3) the Tethys Sea, a precursor of the Mediterranean, extended east-southeast to open into the equatorial Pacific, and was connected by the Volga Sea and the Atlantic with the Arctic. To the south the Southern Ocean had no direct connection with the

equatorial oceans. The result was a permanently large temperature difference between the seas north and south of Gondwanaland, giving rise to a permanent wind from south to north across the continent. The latter consisted of a lofty plateau or series of high mountain ranges extending well above the snow line. For most of its track the temperature of the air must have been below its dew point, causing an almost permanent cloud cover. In such circumstances glaciers could descend the mountain slopes and coalesce to form continental ice sheets, protected by the cloud cover from the tropical sun. On the other hand, the warm seas to the north encouraged the growth of rich vegetation. It is probable that the coal forests preceded the maximum glaciation, because when the ice sheets reached the sea the melting ice lowered the temperature of the ocean waters by almost 10F and spread the local cooling over the whole world. This fits in with the impoverishment of the fauna and flora and the period of drought which set in during the Permian.

The warmest conditions apparently occurred in Europe and western Asia, where the influence of the Tethys Sea was greatest, and here there is no good evidence of even small glaciers. In North America, where there was some local cooling due to a return circulation from north to south, there were large mountain glaciers, represented especially by the famous Squantum Tillite, but as the Arctic Ocean was apparently ice-free, the climate of the lowlands was favourable enough for coal forests.

The Late Palaeozoic Ice Age, like the Quaternary, was not continuous, but was broken up into glacial and interglacial periods. This suggests that either geographic conditions were not continuously favourable for glaciation or more probably that some other factor was superposed. This point is discussed on p. 1016.

Humphreys' Theory of Volcanic Dust. In 1913 Humphreys [14] showed that the presence of volcanic dust in the atmosphere lowered the mean temperature of the surface. The dust particles are large enough to scatter solar radiation, returning part of it to space unaltered, but too small to have any appreciable effect on long-wave terrestrial radiation. He calculated that during the period of instrumental observations explosive volcanic eruptions have lowered the mean temperature of the earth by about 1F. At present the amount of vulcanicity is small compared with some of the geological periods, and it is possible that volcanic dust has contributed towards the decrease of temperature in glacial periods. On the other hand, the complete absence of volcanoes could not alone have raised the temperature by more than 1F—a negligible amount. Vulcanicity can never have been more than a contributory cause; an attempt to estimate its effect quantitatively is described on p. 1016.

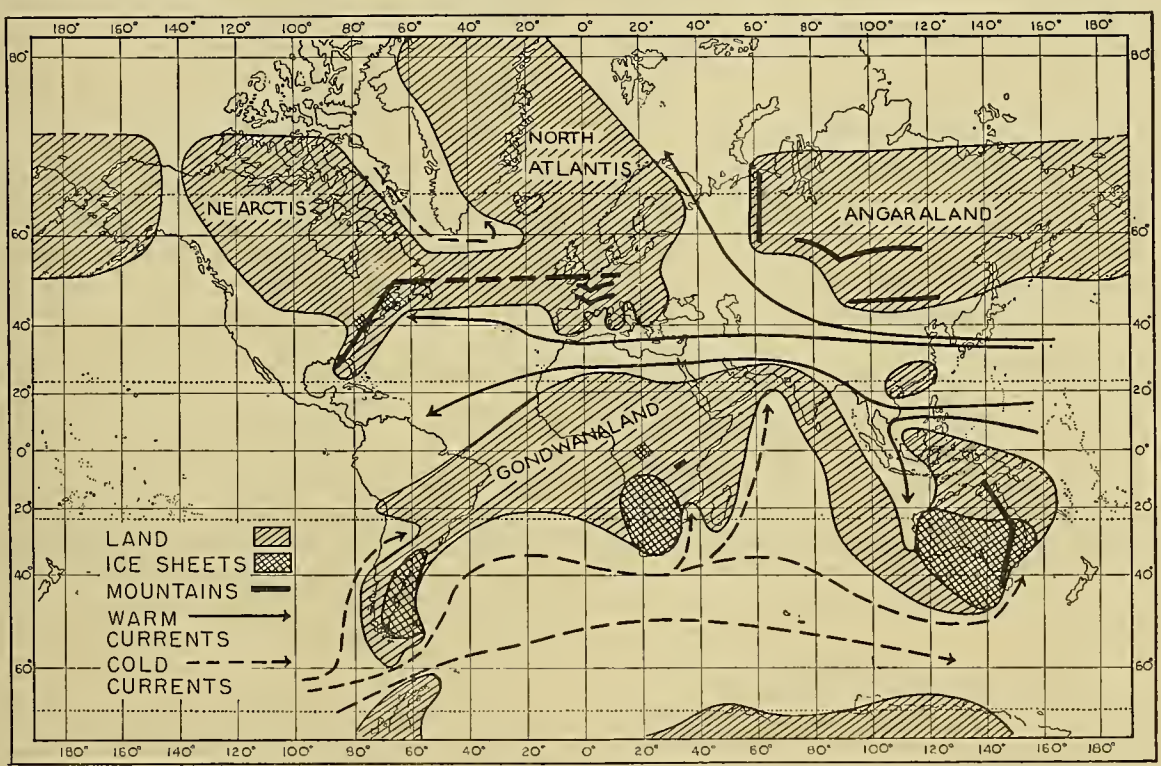


Fig. 3.—Geography, ocean currents, and ice sheets of the Late Palaeozoic.

Variations of Carbon Dioxide. Carbon dioxide absorbs long-wave radiation and so helps to maintain the temperature of the earth's surface above that at which it would otherwise be in equilibrium with solar radiation. The amount of CO_2 in the atmosphere must have varied greatly during geological time, being de-

pleted by the formation of limestones (carbonates) and coal measures, and replenished by volcanic action. Ordinarily the variation was slow, because a great reserve of CO_2 is dissolved in the oceans. Arrhenius and Chamberlin saw in this a cause of climatic changes, but the theory was never widely accepted and was abandoned when it was found that all the long-wave radiation absorbed by CO_2 is also absorbed by water vapour.

In the past hundred years the burning of coal has increased the amount of CO_2 by a measurable amount (from 0.028 to 0.030 per cent), and Callendar [7] sees in this an explanation of the recent rise of world temperature. But during the past 7000 years there have been greater fluctuations of temperature without the intervention of man, and there seems no reason to regard the recent rise as more than a coincidence. This theory is not considered further.

The Topographic Theory of Climatic Change. The theory that climatic changes are due to a combination of terrestrial factors—elevation, continentality, ocean currents, and vulcanicity—may be termed the "topographic theory." Palaeogeographers have given us a fairly complete history of the topographic changes since the Cambrian; palaeontologists and palaeobotanists have reconstructed the variations of temperature. From these data Brooks [4] made a statistical comparison of the topographic and climatic variations. The data used were:

1. Estimates of the mean temperature of middle and high latitudes (from 40° to 90°N) in each of thirty geological periods from Upper Proterozoic to Recent. For this purpose a curve given by Dacqué [9] for the zonal differentiation of climate was converted to mean temperatures using the present mean temperature of 33F and the assumed means of 53F in the Middle Jurassic and 28F in the Pleistocene. The assumption was made that the temperature of equatorial regions has varied little.

2. Estimates of the mean height of the continents based on a curve of mountain-building activity given by Dacqué, assuming a mean height of 3500 ft in the early Quaternary, 2500 ft at present, and 500 ft in the warm periods.

3. Estimates of "continentality," based on the areas of the continents in different latitudes from 80° to 40°N.

4. Estimates of the volume of the warm currents reaching the Arctic, as a percentage of that given by the most favourable conditions. Both these were based on the geographic reconstructions by Arldt [2].

5. Estimates of the vulcanicity on a scale of 0–10, from the amount of volcanic material in the different formations.

The mean temperature was compared with the four geographic variables by correlation, and the effect of one "unit" of each factor evaluated, separately for the Palaeozoic and for the Mesozoic and Tertiary together. The effect of one "unit" of each factor was also found from present-day conditions by independent, more or less theoretical calculations. The results are given in Table IV.

In spite of the roughness of the data and the approximations which had to be made in calculating the "theoretical" values, the three sets of figures agree well, except for Palaeozoic vulcanicity which was very difficult to estimate. The agreement strongly supports the theory that changes of orography and land distribution are a nearly complete explanation of the long-period changes of climate.

Table IV. The Temperature Effect of Terrestrial Factors in Climatic Change

Factor	"Unit"	Effect (°F) of one "unit"		
		Palaeozoic	Mesozoic	"Theoretical"
Height of continents	100 ft	−0.38	−0.47	−0.4
Continentality	%	−0.31	−0.32	−0.35
Ocean currents	%	+0.28	+0.28	+0.3
Vulcanicity	½ present	−1.68	−0.42	−0.5

Adopting the figures for Mesozoic and Tertiary as a basis, a partial regression equation was formed for temperature on the various geographic factors, and the "theoretical" temperature of each geological period was calculated. The result, compared with the "estimated" temperature from Dacqué's curve, is shown in Fig. 4. Here again the agreement is reasonably good, except for the earlier Palaeozoic, but there are some discrepancies. The three great glacial periods of the Upper Proterozoic, Late Palaeozoic, and Quaternary stand out clearly, but the "estimated" curve lags behind the "calculated" by five to ten million years, as if the earth took a long time both to cool and to warm up again in mountain-building periods. There is a steep drop in the "calculated" curve for the Lower Jurassic (Lias) which is barely represented in the "estimated" curve, either because the distribution of mountain ranges in relation to moisture-bearing winds was unfavourable for glaciation or because the cooling was not sufficient to freeze the polar seas. Finally, during the Palaeozoic the "calculated" curve is mostly above the "estimated," probably because the reconstructions showed too little land in high latitudes. Our knowledge of Palaeozoic geography is still rather fragmentary.

This statistical method of treatment automatically includes the secondary effects of freezing and thawing of the polar seas and changes in the atmospheric circulation, which would be difficult to handle quantitatively in any other way.

Shorter Climatic Oscillations—the "Solar-Topographic Hypothesis." The curves of temperature in Fig. 4 are generalised and do not show the fluctuations of relatively short period. For example, crowded into the narrow dip of the Quaternary were at least four oscillations with a range of more than 5F, and similar fluctuations undoubtedly occurred in the other Ice Ages. Even in the warm periods there must have been short-period oscillations, though the range was probably much less. The geographic factors (except vulcanicity) change slowly, and cannot account for these short-period changes.

We have seen that there are two causes which could have produced these shorter oscillations. Changes in

the elements of the earth's orbit (p. 1011) are known to have occurred, but are probably quantitatively insufficient. Variations of solar radiation could have produced most of the observed effects, and it is not inherently improbable that the sun is a variable star. The present trend of thought is therefore towards what Flint [11] terms the "solar-topographic hypothesis." A fluctuation of solar radiation by 10–20 per cent on either side of the mean value combined with the geographic cycle to produce the variations of geological climate. This hypothesis seems to offer the best basis for future research, but it is still doubtful whether the solar or topographic partner predominates. One view is that topography governs the major swings of climate but the solar factor determines the incidence of glaciation; the other view is that solar variations decide the main climatic variation but that mountains are essential for glaciation.

1006). The radioactive technique devised by Urry gives a prospect of dating the climatic changes revealed by these cores. Further, Ovey [26] foresees the possibility of tracing the sources of the marine glacial deposits and so tracking the paths of the icebergs, which will give us the ocean currents and winds of the glacial periods.

At present, especially as regards the pre-Tertiary periods, geologists are too apt to characterise the whole of a geological period, lasting millions of years, as having a certain unchanging type of climate. This is improbable, and careful detailed studies of the fauna, flora, and lithology will result in a picture of the minor fluctuations superposed on the broad swings shown in Fig. 4. If a rough time-scale (not necessarily absolute dating) can be added, this will go far towards solving the problem of causes.

The Factors of Climate. Parallel with the study of

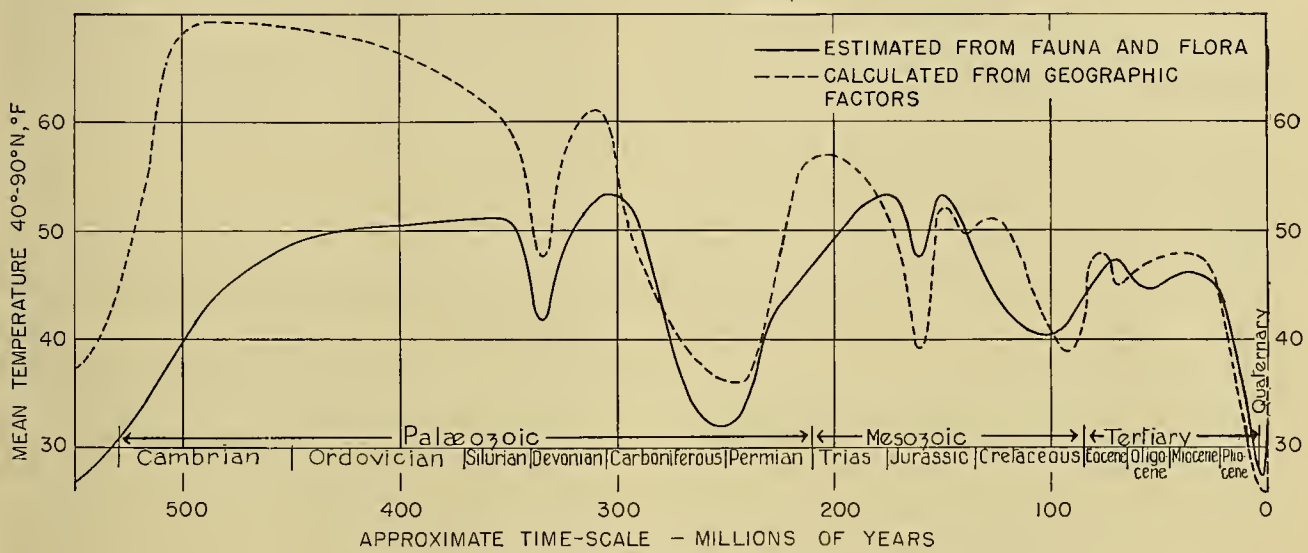


Fig. 4.—Variations of temperature in geological time.

The evidence from variations of rainfall is not clear. Increased solar radiation should increase total precipitation, and so far as we can determine, precipitation in the early Tertiary was greater than at present, On the other hand the long warm periods of the Mesozoic were characterised by desert deposits, pointing to a rainfall smaller than the present. This point needs further investigation.

THE FUTURE STUDY OF CLIMATIC CHANGES

The Palaeoclimatic Sequence. The greatest need at present is for a closer quantitative evaluation of climatic variations in geological time. We need better estimates of the distribution of mean annual temperature and annual range, and the amount and seasonal distribution of rainfall. These data, plotted in conjunction with the orography, may in time lead to an understanding of the wind systems. In particular, we need to know more about conditions over the oceans. This gap is being filled for the Quaternary and later Tertiary by the cores raised from the ocean floor (p. the palaeoclimatic sequence must go a more exhaustive quantitative study of the effect of the various solar and topographic factors on climate, and particularly on temperature and on the circulation of the earth's atmosphere. From a study of existing conditions we should be able to construct a more exact series of equations of climate which can be applied to any other topography. The temperature relations are already known to some extent, and the urgent problem now is to study the effect of topography on the atmospheric circulation.

When these parallel studies have reached a sufficiently advanced stage, we may be able to realise the dream of Kerner [18] that, from a large number of comparisons of quantitative estimates of local temperature from geological data with calculations of the same values from topography (both local and general), we can reconstruct the variations of extraterrestrial factors and especially of solar radiation. A few such comparisons already exist, made mostly by Kerner himself. An interesting case is the middle Eocene of southeast England.

Very detailed studies of the flora were made by Mrs. E. M. Reid and Miss Chandler [29], from which they inferred that the mean temperature was almost certainly not below 70F. On the other hand, from a careful study of the topography I estimated the probable upper limit of the mean temperature as 65F. In all other respects the climatic reconstruction agreed closely with the inferences from the vegetation, and it seems likely that the discrepancy of at least 5F was due to increased solar radiation.

It is only by the multiplication of such studies, both in space and time, that the fundamental problem of climatic variations can eventually be solved. The problem is not entirely academic, for signs are not wanting that we are even now in a period of climatic change which may have vital, but so far unpredictable, consequences for civilisation.

REFERENCES

A 201-item annotated bibliography has recently been prepared by the author. See C. E. P. Brooks, "Selective Annotated Bibliography on Climatic Changes." *Meteor. Abstr. & Bibliogr.*, 1: 446–475 (1951).

1. Ahlmann, H. W:son, "Glaciological Research on the North Atlantic Coasts." *R. geogr. Soc., Res. Ser.*, No. 1 (1948).
2. Arldt, T., *Handbuch der Palaeogeographie*, 2 Bde. Leipzig, Gebr. Borntraeger, 1919.
3. Bradley, W. H., "The Varves and Climate of the Green River Epoch." *U. S. Geol. Surv. Prof. Pap.* 158 E, pp. 87–110 (1929).
4. Brooks, C. E. P., *Climate Through the Ages*, 2nd (revised) ed. London, Benn, 1949.
5. —— "Continentality and Temperature—Second Paper: The Effect of Latitude on the Influence of Continentality on Temperature." *Quart. J. R. meteor. Soc.*, 44: 253–269 (1918).
6. —— "The Problem of Warm Polar Climates." *Quart. J. R. meteor. Soc.*, 51: 83–91 (1925).
7. Callendar, G. S., "The Composition of the Atmosphere Through the Ages." *Meteor. Mag.*, 74: 33–39 (1939).
8. Chaney, R. W., "Tertiary Forests and Continental History." *Bull. geol. Soc. Amer.*, 51: 469–488 (1940).
9. Dacqué, E., *Grundlagen und Methoden der Palaeogeographie*. Jena, G. Fischer, 1915. (See pp. 432, 449)
10. Ficker, H. v., "Die eiszeitliche Vergletscherung der nordwestlichen Pamirgebiete." *S. B. preuss. Akad. Wiss., Mat.-Nat. Kl.*, SS. 61–86 (1933).
11. Flint, R. F., *Glacial Geology and the Pleistocene Epoch*. New York, Wiley; London, Chapman, 1947.
12. Hansen, H. P., "Post-glacial Forest Succession, Climate and Chronology in the Pacific North-West." *Trans. Amer. phil. Soc.*, 37: 1–130 (1947).
13. Hoyle, F., and Lyttleton, R. A., "The Effect of Interstellar Matter on Climatic Variation." *Proc. Camb. phil. Soc.*, 35: 405–415 (1939).
14. Humphreys, W. J., "Volcanic Dust and Other Factors in the Production of Climatic Changes, and Their Possible Relation to Ice Ages." *Bull. Mt. Weather Obs.*, 6: 1–34 (1913).
15. Huntington, E., and Visher, S. S., *Climatic Changes, Their Nature and Cause*. New Haven, Yale University Press, 1922.
16. Jeffreys, H., *The Earth, Its Origin, History and Physical Constitution*, 2nd ed. Cambridge, University Press, 1929. (See p. 144)
17. Kerner-Marilaun, F., "Das akryogene Seeklima und seine Bedeutung für die geologischen Probleme der Arktis." *S. B. Akad. Wiss. Wien.*, Abt. IIa, 131: 153–185 (1922).
18. —— *Paläoklimatologie*. Berlin, Gebr. Borntraeger, 1930. (See pp. 455–464)
19. Köppen, W., und Wegener, A., *Die Klimate der geologischen Vorzeit*. Berlin, Gebr. Borntraeger, 1924.
20. Lasareff, P., "Sur une méthode permettant de démontrer la dépendance des courants océaniques des vents alizés et sur le rôle des courants océaniques dans le changement du climat aux époques géologiques." *Beitr. Geophys.*, 21: 215–233 (1929).
21. Lewis, G. N., "Thermodynamics of an Ice Age: the Cause and Sequence of Glaciation." *Science*, 104: 43–47 (1946).
22. Milankovitch, M., "Mathematische Klimalehre und astronomische Theorie der Klimaschwankungen," in *Handbuch der Klimatologie*, W. Köppen und R. Geiger, Bd. 1 Teil A, SS. 1–76. Berlin, Gebr. Borntraeger, 1930.
23. Moreau, R. E., "Pleistocene Climatic Changes and the Distribution of Life in East Africa." *J. Ecol.*, 21: 415–435 (1933).
24. Movius, H. L., Jr., *The Irish Stone Age, Its Chronology, Development, and Relationships*. Cambridge, University Press, 1942. (See Fig. 8, p. 60, and Table IV, p. 63)
25. Nilsson, E., "Quaternary Glaciations and Pluvial Lakes in British East Africa." *Geogr. Ann., Stockh.*, 13: 249–349 (1931).
26. Ovey, C. D., "Note on the Evidence for Climatic Changes from Sub-oceanic Cores." *Weather*, 4: 228–231 (1949).
27. Penck, A., und Brückner, E., *Die Alpen im Eiszeitalter*, 3 Bde. Leipzig, Tauchnitz, 1909. (See pp. 1168–9)
28. Pettersson, O., "Climatic Variations in Historic and Prehistoric Time. "*Svenska hydrogr.-biol. Komm. Skr.*, No. 5 (1914).
29. Reid, E. M., and Chandler, M. E. J., *The London Clay Flora*. London, British Museum, 1933. (See pp. 63–82)
30. Simpson, G. C., "Past Climates." *Mem. Manchr. lit. phil. Soc.*, 74: 1–34 (1929).
31. Wagner, A., *Klimaänderungen und Klimaschwankungen*. (*Die Wissensch.*, Bd. 92) Braunschweig, F. Vieweg & Sohn, 1940.
32. Washington, Carnegie Institution, Publication No. 352. *Quaternary Climates. Papers by J. C. Jones, E. Antevs, and E. Huntington*, 212 pp. Washington, D.C., 1925.
33. Wayland, E. J., "Rifts, Rivers, Rains and Early Man in Uganda." *J. R. anthrop. Inst.*, 64: 333–352 (1934).
34. Willett, H. C., "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6: 34–50 (1949).
35. Winkle, W. van, "Quality of the Surface Waters of Oregon." U. S. Geol. Surv. (Dept. of Interior), *Water Sup. Pap.* 363 (1914).
36. Zeuner, F. E., *The Pleistocene Period; Its Climate, Chronology and Faunal Succession*. London, Ray Society, 1945.

CLIMATIC IMPLICATIONS OF GLACIER RESEARCH[1]

By RICHARD FOSTER FLINT

Yale University

Glaciers as Climatic Indicators

Nature and Causes of Changes in Glaciers. Recognition of the relationship between changes in glaciers and changes in climate is older than the glacial theory as set forth by Agassiz. In 1821 Ignance Venetz [22], a Swiss civil engineer, announced an apparent correlation of fluctuations in Alpine glaciers with fluctuations of the regional snowline and of temperature. In land registers and other public records he found proof that these glaciers had been less extensive formerly than at the time of his investigation. Using glacier changes as a basis of inferences as to climatic changes, Venetz concluded that the average temperature fluctuates in an irregular manner, and that the abandoned end moraines of Alpine glaciers are very ancient and record earlier changes in climate. He even stated the opinion that the latest refrigeration had reached its climax, a statement that has proved to be remarkably true in the light of events since 1821.

Some years later Agassiz [1, pp. 237–239] placed the evidence of glacier behavior against that of thermometry which, it had been asserted in 1834, indicated no change in the average temperature at the earth's surface during historic time. In fact, Agassiz [1, p. 305] believed that the evidence of widespread former glaciation implied temperature changes of world-wide extent.

Thus glaciers were acknowledged as climatic indicators as soon as they were first studied scientifically. Refined observations, however, did not begin until long afterward. In the later part of the nineteenth century contemporary changes in glaciers so placed as to be capable of systematic observation were compared with changes in temperature and snowfall, both annually and over periods of years. Only recently, however, has an apparent correlation been established between glacier behavior and climatic changes[2] affecting wide regions.

Observed fluctuations of glaciers consist of changes in areal extent and thickness. Annual observations, begun in 1894, have been made on the positions of the termini of valley glaciers, notably in the Alps, western United States, western Canada, and Alaska (*e.g.*, Field [7]). Some of the data obtained are detailed, and will increase in significance as the record lengthens. More recently changes in thickness—more difficult to determine, but far greater in terms of volume—have been systematically recorded.

Since 1894 the assembling and recording of observations has been in the hands of capable committees. From 1894 to 1916 reports [19] were published by the International Committee on Glaciers of the International Geological Congress. From 1927 to date, reports have been published by the Commission Glaciologique of the Union Internationale de Géodesie et Géophysique, with the American Geophysical Union's Committee on Glaciers cooperating since 1931.

The value of the records obtained, as far as climatic implications are concerned, does not lie in fluctuations from year to year, which reflect only short-term changes in precipitation and temperature, as well as the avalanching of snow and other local factors. It lies rather in the cumulative changes, both during periods of several decades and during much greater spans of time. During the last few decades such cumulative changes have been so great that they have been described as "catastrophic."

It has become clear that change of temperature affects a glacier in at least two ways. First and most obviously, it determines the amount of ablation that occurs. Also, however, it determines the proportion of the precipitation that occurs in solid form, thus directly affecting the nourishment of the glacier. Hence it is apparent that a rise in temperature operates to diminish a glacier both through increased ablation and through decreased nourishment.

It has been observed, further, that decreased thickness is ordinarily accompanied by shrinkage in area, and vice versa. However, this is not true universally. Examples of valley glaciers that lengthened while becoming thinner were first noted very early [22, p. 15] and recently the Antarctic Ice Sheet, the world's largest existing glacier, has been thought [6, p. 392] to be increasing in extent while thinning. Accordingly the most refined observations take account of both thickness and area.

It is known further that although nearly all the glaciers within a region may be shrinking, one or two may be expanding at the same time, and vice versa. The exceptions are commonly explained as delayed responses to the latest climatic change, influenced by variations in local factors such as the volumes, slopes, and altitudes of the glaciers and in the relief of the surfaces on which they lie. An example is the Taku Glacier in coastal Alaska. Despite a general shrinkage in that region during recent years, this glacier has been expanding.

The most important qualitative-quantitative field study of glacier regimens ever undertaken is the work of Ahlmann [2] sustained throughout many years

1. Constructive reading of the manuscript by H. E. Landsberg and W. O. Field, Jr., is acknowledged.

2. The Climatological Commission of the International Meteorological Organization, meeting in Warsaw in 1935, distinguished between climatic *fluctuations* (differences between two 30-year means) and climatic *variations* (differences of a larger order, not specifically defined by the Commission). As far as practicable, this usage is followed in the present paper.

around the coasts bordering the North Atlantic Ocean. This work established that ablation of a glacier is controlled by two factors: direct radiation, which predominates in dry continental climates; and a combination of convection and condensation, which predominates in moist maritime climates. It established further that temperature is the dominant factor in the glacier regimen; glaciers are more responsive to changes in temperature than to changes in precipitation.

Correlation with Climatic Changes Determined Independently. Not until recently was the probability of a contemporary widespread temperature increase during the last 100 years established by actual compilation of weather records [10, 11], with the change affecting both polar hemispheres. Knowledge of this temperature increase was refined and enlarged through the subsequent appearance of several more compilations and interpretations of weather records (summarized and cited by Ahlmann [4, pp. 169–181]).

The modern temperature increase has been concomitant with a general shrinkage in the volumes of glaciers in both hemispheres [3, pp. 67–68; 15, p. 231]. It has been concomitant also with a reduction in the extent and thickness of floating ice in the Arctic Sea [4, p. 187] and with an apparent rise of sea level [9, 13], best explained as the result of an increment of water derived from the melting of glacier ice.

There appears, then, to be an obvious correlation among these observations: (1) secular increase of mean temperature, (2) shrinkage of glaciers, (3) shrinkage of sea ice, and (4) apparent rise of sea level generally within the span of the last 100 years.

This correlation seems to establish the relation between the fluctuations of climate, although data both more extensive and more precise would be desirable.

Ever since the pioneer announcement by Venetz, research workers have assumed that pronounced former variations in glaciers reflected changes in climate, and on this basis have inferred major climatic variations from the abundant geologic evidence of variations in glaciers. When the recent apparent correspondence between glaciers and climate has become more firmly established, the resulting quantitative data should prove useful in interpreting the climatic variations of earlier times.

Changes in Glaciers within Historic Time

As is true of most geologic phenomena, both the abundance and the quality of the evidence with which we have to work diminish conspicuously as we go backward in time. Therefore, in a review of the climatic evidence furnished by glaciers it is best to begin with the present and work back into the past.

Changes within the Last 100 Years. Abundant evidence supports the statement, made above, that glaciers have been generally shrinking during the last century. The evidence is derived from observations made in western North America (including Alaska), Greenland, Iceland, Spitsbergen, Scandinavia, the Alps, East Africa, South America, New Zealand, and the Antarctic Continent [3, 4, 15, 20]. Correlated with this are a contemporaneous shrinkage of lakes and a group of ecologic and oceanographic changes (summarized in [4]).

As Matthes [15] pointed out, the fact that glacier shrinkage has been occurring simultaneously in both polar hemispheres bears on existing hypotheses of the cause of climatic changes. This fact disfavors those "astronomic hypotheses" which demand, at least to some degree, temporal offsets of climatic effects between the polar hemispheres.

Changes since the Beginning of the Christian Era. Research in the historic period before the beginning of systematic glacier observations has consisted of extending the data on glacier changes backward in time through the study of official records made for other purposes. This was attempted first, and with considerable success, by Venetz [22]. Another example is a study of the archives of the town of Chamonix in the French Alps. In this study Rabot [18] succeeded in extending the record back to the year 1850 and demonstrated further that the present-day shrinkage commenced, in that district, about the middle of the nineteenth century, following a period of some 250 years during which the glaciers were relatively expanded. A more comprehensive reconstruction, covering a much wider area although it extended back only to the year 1700, was attempted by Brückner [5].

Similar research into the Iceland record by Thorarinsson [21] confirmed and extended these conclusions. Thorarinsson showed that glaciers on that island were far less extensive from the tenth century to the thirteenth than during the period since the beginning of the fourteenth century, and established changes of lesser degree within the latter period. This reconstruction is supported by various historical data not directly connected with glaciers (*e.g.*, [12, 23]).

Most of the first millenium of the Christian era is believed to have been relatively warm, but the evidence is derived from the fluctuations of lakes, from tree rings, and from other nonglacial data. Apparently nothing is on record as to the condition of glaciers during this time. A record may be present in the stratigraphy of glacial deposits, but if so, it does not seem to have been recognized.

Variations between the Wisconsin Maximum and the Beginning of the Christian Era

Still farther back in time the glacial evidence consists, not of observations on the glaciers themselves, but of stratigraphic and morphologic features. Many of these features are moraines built along the glacier margins and abandoned during shrinkage; from these, former glacier dimensions—extent or thickness or both—can be approximated. Research on this ancient period lies in the field of glacial geology rather than in that of glaciology.

The Wisconsin maximum is believed to represent the latest of a series of four major glacial expansions on a world-wide scale. It is generally assumed to date from about 60,000 years ago. The rough measurements available indicate that at that time glacier ice, mostly in the form of great ice sheets, covered about 27 per cent

of the world's present land area, as compared with a coverage of slightly more than 10 per cent by the glaciers of today [8, p. 451].

In North America, for example, the Wisconsin maximum is evidenced by the outer limit of a sheet of relatively fresh young drift stretching across the continent from the Atlantic near New York to the Pacific near Seattle. This drift is composite, consisting of thin, extensive layers of glacial deposits alternating in some regions with layers of loess and with beds of peat and other deposits of organic origin. Over wide sectors the individual layers of drift thicken to form series of subparallel, ridgelike moraines.

Each drift layer is interpreted as the product of a considerable glacial expansion, while the moraines are believed to have been built by lesser pulses of increase in the ice. As both drift layers and moraines generally occur in off-lapping relationship, like the clapboards on a frame house, the implied history since the great thrust at the Wisconsin maximum is one of gradual shrinkage interrupted by renewed but ever-lessening re-expansions.

The climatic implication of this sequence of events is, by analogy, a general increase of temperature punctuated by temporary, diminishing reversals of this trend. The implied climatic history is confirmed at various points by the ecologic relations of fossil organisms. No periodicity is apparent in the sequence.

The areal extent of the ice at various times is clearly recorded by the successive moraines and layers of drift. Owing to the low relief of most of the glaciated country, former ice thicknesses are rarely indicated directly. However, evidence of progressive thinning, particularly in the later part of the time span represented, exists in this fact: Inward from the periphery of the Wisconsin drift, moraines become fewer, smaller, and less continuous, and other features that indicate thin, slowly flowing or even stagnant ice become more abundant.

A generally comparable sequence exists in northern Europe; however, definite time correlations between events in Europe and those in North America have not yet been made.

The so-called *Climatic Optimum*, a warm period believed to have reached its climax some thousands of years before the beginning of the Christian era, is now widely recognized through the interpretation of organic evidence and through the history of fluctuations of lakes and sea level. The glacial record thus far includes very few indications of this warm time, doubtless because glaciers then were reduced in area. Thus the critical evidence is apparently covered up by present-day glaciers which, despite their recent shrinkage, are still large when compared with their condition during the Climatic Optimum. Fossil wood incorporated in young moraines at two localities in western United States is believed to record forests that flourished during the Optimum and were later overwhelmed by glacier ice. Matthes [14, p. 520] believed that most of the glaciers in western United States ceased to exist during the Optimum and have since been reconstituted in what he called the "Little Ice Age."

Major Variations within the Pleistocene Epoch

Glacial geology affords evidence of major climatic changes as far back as the beginning of the Pleistocene epoch, a million or more years ago. Because time and events have destroyed much of the record, it is far from complete, but the main outlines stand out with some distinctness [8].

In North America, and independently in Europe, four major glacial ages, of which the latest is the Wisconsin, are inferable. Geologists regard each of these ages as having been of the order of 100,000 years in duration, although it must be conceded that this figure is hardly more than a controlled guess.

As far as their positions are now known, the outer limits of the four major drift sheets are subparallel. This fact implies that over wide regions, at least, the major relief features of the lands, the mountains and lowlands, remained much the same throughout this whole time, for if there had been any widespread change the patterns of the successive drifts would have differed more than they seem to do. Mountain uplifts known to have occurred during the Pleistocene epoch, such as coastal mountains in California and the Himalayas in southern Asia, would have exerted only a limited influence on the world pattern of glaciers.

The area covered by ice when at its maximum extent was of the order of 32 per cent of the land area of the world, compared with about 10 per cent covered today.

That the climate was characterized by reduced temperatures and perhaps also greater precipitation in districts not covered by glaciers is established by two groups of phenomena. The first consists of large lakes and vigorous stream activity in western North America, central Asia, northern and eastern Africa, central Australia, and other regions now comparatively dry. The second consists of abundant evidence of former vigorous freezing and thawing of soil in districts where such activity is feeble or nonexistent today.

Additional glacial evidence of major climatic variation lies, in mountain districts, in the discrepancy between the altitudes of cirques occupied by valley glaciers today and the much lower altitudes of abandoned cirques. In some districts this discrepancy amounts to as much as 1200 meters. As there is an evident relationship between the general altitude of cirque floors in any district and the position of the regional snowline, the discrepancy between the two sets of cirques is a rough measure of the depression of the regional snowline during the glacial ages, as compared with its present position.

The extent of glaciers during the times between the glacial ages is virtually unknown. That climates in the temperate zones were mild by comparison with the glacial climates in those zones is indicated by the character of soils developed on glacial drift sheets and buried by the next succeeding drift sheets. The great thicknesses and deep alteration of such soils constitute the basis for a general belief that the interglacial ages were much longer than the glacial ages, that they may have lasted 200,000 to 300,000 years each. Direct evidence of the character of interglacial climates comes from

fossil plants and animals that occur between the layers of drift. In certain cases it has been possible to show that comparative ecologic assemblages are now living in climates milder than those prevailing at the fossil localities. From such relations it is inferred that interglacial climates, in some regions at least, were milder than those of today.

Accurate quantitative data on the fluctuation of mean annual temperature between glacial and interglacial times do not exist. Using altitudes of cirques, Penck [17] calculated the differences between glacial-age temperature and present temperature as being of the order of 7C to 8C on the Atlantic coast of Europe. From scanty organic evidence a difference of 2C between an interglacial maximum and the present time has been calculated. On this basis we have a total range of the order of 9C or 10C.

Owing to the presence of variables, however, these figures are more suggestive than accurate. Both apply to glaciated regions, and it may well be, as Willett [23, p. 43] suggested, that the differences would diminish in lower latitudes.

It seems probable that minor climatic variations, similar to those which have taken place during recent millenia, occurred also during earlier times, as irregularities superposed on the major glacial and interglacial fluctuations. For the most part, however, the geologic record is too meager to enable us to state definitely that such is the case.

Nonperiodic Character of Climatic Changes

Neither the facts of glaciology, evidencing recent climatic changes, nor the facts of glacial geology, evidencing more ancient ones, afford a basis for inferring a periodic recurrence of any particular climatic condition. Brückner [5] assembled a mass of data covering a modern 200-year period, from which he derived the concept of a climatic cycle with a period of about 35 years. Although widely quoted, this concept was later refuted by Willett [23, p. 36] and today has dubious standing. Paschinger [16] assembled data on regional snowlines, from which he concluded that climatic cycles occur in opposite phase in the north and south polar hemispheres. This conclusion was refuted by Matthes [15, p. 232].

Present Research Needs

The usefulness of glaciers as climatic indicators has been established. In view of this fact it is apparent that research on changes in glaciers should be pursued both more intensively and more extensively than hitherto. The research needed falls into two distinct fields, glaciology and glacial geology.

In the glaciologic field there is required a more nearly complete record of contemporary fluctuations of glaciers, so that after several decades curves prepared from annual observations will become available for comparison. The glaciers used for such observations should be carefully selected so as to include both polar hemispheres, both high and low latitudes, both maritime and continental glaciers, and both valley glaciers and ice sheets. If possible, one of the observed glaciers should be situated very close to the equator—in East Africa, the Andes, or New Guinea. Such a system of observations would require the existence of a vigorous international organization with centralized maintenance of records.

In the geologic field the greatest need is for a better knowledge of the areas of glaciers at various times during the last 10,000 years. In general, less is known about these comparatively recent variations than about variations of more ancient date. Specifically, what is required is the mapping and study of the abandoned younger end moraines of valley glaciers in selected mountain districts, and moraines and other deposits of the former ice sheet in northeastern North America, a region still little known.

Fulfillment of these two general needs would place our knowledge of climatic changes, as determined from past and present glaciers, on a firm basis.

Summary

In summary, an examination of the bearing of glacier research on the reconstruction of former climates leads to these conclusions:

1. A close relationship between changes in glaciers and changes in climate, particularly temperature, seems established.

2. Observations on glaciers themselves afford some climatic data for the last 1000 years.

3. Geologic features related to former glaciers afford much cruder climatic data for the last 1,000,000 years or more.

4. The climatic changes inferred from (2) and (3) are supported also by a variety of nonglacial evidence, much of it organic.

5. The climatic variations included four major cold ages, with extensive glaciers, three major warm ages, with greatly reduced glaciers, and lesser variations within the last few tens of thousands of years. The latest fluctuation consists of a general rise of temperature within the last 100 years.

6. As far as the existing evidence goes, the changes in climate seem to have been similar in character (though not of course in amount) throughout the world. There is no evidence that the changes recur in a periodic manner.

7. Rough calculation has suggested that in districts near the large ice sheets the amplitude of mean temperature fluctuation between cold ages and warm ages may have been of the order of 10C.

8. Present research needs include (*a*) a greatly expanded, coordinated system of observations on contemporary glacier fluctuations, and (*b*) a study of very late Wisconsin end moraines in order to throw more light on climatic changes during the last 10,000 years.

REFERENCES

1. Agassiz, L., *Études sur les glaciers*. Neuchâtel, Jent et Gassmann, 1840.

2. Ahlmann, H. W., "Researches on Snow and Ice, 1918–40." *Geogr. J.*, 107: 11–28 (1946).

3. —— "Glaciological Research on the North Atlantic Coasts." *R. geogr. Soc. Rc:. Ser.*, No. 1, 83 pp. (1948).
4. —— "The Present Climatic Fluctuation." *Geogr. J.*, 112: 165–195 (1948).
5. Brückner, E., "Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen der Diluvialzeit." *Geogr. Abh.*, Bd. 4, Heft 2, 325 SS. (1890).
6. Demorest, M., "Ice Sheets." *Bull. geol. Soc. Amer.*, 54: 363–400 (1943).
7. Field, W. O., Jr., "Glacier Recession in Muir Inlet, Glacier Bay, Alaska." *Geogr. Rev.*, 37: 369–399 (1947).
8. Flint, R. F., *Glacial Geology and the Pleistocene Epoch.* New York, Wiley, 1947.
9. Gutenberg, B., "Changes in Sea Level, Postglacial Uplift, and Mobility of the Earth's Interior." *Bull. geol. Soc. Amer.* 52: 721–772 (1941).
10. Kincer, J. B., "Is Our Climate Changing? A Study of Long-time Temperature Trends." *Mon. Wea. Rev. Wash.*, 61: 251–259 (1933).
11. —— "Our Changing Climate." *Trans. Amer. geophys. Un.*, 27: 342–347 (1946).
12. Manley, G., "Some Recent Contributions to the Study of Climatic Change." *Quart. J. R. meteor. Soc.*, 70: 197–219 (1944).
13. Marmer, H. A., "Sea Level Changes Along the Coasts of the United States in Recent Years." *Trans. Amer. geophys. Un.*, 30: 201–204 (1949).
14. Matthes, F. E., "Report of Committee on Glaciers, April, 1939." *Trans. Amer. gcophys. Un.*, 20: 518–523 (1939).
15. —— "Report of Committee on Glaciers, 1945." *Trans. Amer. geophys. Un.*, 27: 219–233 (1946).
16. Paschinger, V., "Die Schneegrenze in verschiedenen Klimaten." *Petermanns Mitt., Ergänzungsh.* 173 (1912).
17. Penck, A., "Das Klima der Eiszeit." *Int. Quartär-Konferenz*, 3d, Vienna, *Verh.*, 1: 1–14 (1936).
18. Rabot, C., "Récents travaux glaciaires dans les Alpes françaises." *La Géographie*, 30: 257–268 (1915).
19. Reid, H. F., "The Variations of Glaciers." *J. Geol.*, Vols. 3–24 (1895–1916).
20. Thorarinsson, S., "Present Glacier Shrinkage, and Eustatic Changes of Sea-Level." *Geogr. Ann., Stockh.*, 22: 131–159 (1940).
21. —— "Vatnajökull. Scientific Results of the Swedish-Icelandic Investigations 1936–37–38. Chap. 11, Oscillations of the Iceland Glaciers in the Last 250 Years." *Geogr. Ann., Stockh.*, 25: 1–54 (1943).
22. Venetz, I., "Mémoire sur les variations de la température dans les Alpes de la Suisse. Rédigé en 1821." Zürich: *Allg. schweiz. Ges. Naturw., Denkschr.*, Bd. 1, Heft 2, SS. 1–38 (1833).
23. Willett, H. C., "Long-Period Fluctuations of the General Circulation of the Atmosphere." *J. Meteor.*, 6: 34–50 (1949).

TREE-RING INDICES OF RAINFALL, TEMPERATURE, AND RIVER FLOW

By EDMUND SCHULMAN

University of Arizona

INTRODUCTION

History. Dendrochronology—the construction, based on tree-ring widths, of significant indices showing changes in climate from year to year and the use of these indices in dating and forecasting—had early beginnings. That the succession of wide and narrow growth rings of trees might provide a chronology of the wet and dry seasons of the past was noted at least as early as the fifteenth century by Leonardo da Vinci. As in other fields, research expanded tremendously during the 1800's, when botanists and foresters, particularly in central Europe and in Scandinavia, examined numerous details of growth-climate relationships [2]. About 1880, the Dutch astronomer, Kapteyn [19], succeeded in deriving a ring chronology several centuries long, which gave a first approximation to the annual rainfall of west Germany, but no clear index, which he sought, of solar activity. His concepts of tree selection and analysis, foreshadowing much later work, remained generally unknown, however, for some thirty-five years.

Dendrochronologic work was begun in the southwestern United States in 1904 by Douglass [5], who formulated the basic principles and who established the superiority of drought areas and coniferous species as sources of ring chronologies of rainfall. Since that time this research has been continuously developed by him and his co-workers at the University of Arizona. Although the main objectives of this research were climatic, there have been some fruitful offshoots, such as the setting up of a precise archaeological time-scale to A.D. 11 for the central part of the Pueblo area, which has been accomplished by the backward extension of the ring record in living trees, based on the ring dating of ancient timbers [6].

Dendrochronologic series have thus far been extensively developed in only three regions. The results of work in the semiarid western United States, as they bear on the climatology and hydrology of the dry lands, will form the bulk of the following survey. In arctic areas, Scandinavia has produced in recent years a number of long and significant chronologies of temperature changes, primarily for the June–July interval; specially long arctic chronologies, composites of records in living trees and ancient timbers after the fashion of those in the Southwest, have been derived for northern Alaska. Scant justice can be done here to a third region of active chronology work, the noncritical areas of the eastern and central United States; several thorough investigations have shown that chronologies of limited climatological significance can be obtained on some sites in this region, though many localities apparently cannot provide dendroclimatic indices.

Principles. It is now known that a great range of variability exists in all aspects of ring growth. In order to obtain climatic indices of maximum simplicity, fidelity, and length—the specific objective of dendrochronologic work in the Southwest—three basic principles have been developed: selection, crossdating, and sensitivity.

Proper field *selection* [25] of trees for sampling is of first importance, for many species, sites, and areas are found to yield ring chronologies of no apparent significance as climatic indices. By successive refinement of selection criteria, trees may be found in which the influence of one dominant factor, such as rainfall in the dry lands, is increasingly represented in the ring-width fluctuations.

The principle of *crossdating* [9] states that the approach to parallelism in ring chronology among the individual trees of a homogeneous set is a direct measure of the influence of some general factor or set of factors on the variations in ring-width from year to year. Such parallelism provides the scientific control for the discovery and evaluation of hidden irregularities in any ring series. This rather simple concept has proved very far-reaching in its thorough application to drought chronologies.

The approach to parallelism between highly sensitive ring series in a small area of the semiarid Southwest, where this phenomenon has been found to be near a maximum, is illustrated in Fig. 1. Although only a half-dozen or so specially long-lived Douglas firs have formed the basis of the record in the outer five centuries of the Mesa Verde and Tsegi indices, the similarity of these two series in light of their 115-mile separation indicates that the random term in individual tree growth has been largely cancelled in this interval; some earlier portions of both series are less well based. The Flagstaff index, more distant from Mesa Verde and based mainly on a different species (ponderosa pine), shows occasional substantial differences from the preceding two chronologies, yet is on the whole in good agreement. Fifty-year correlation coefficients for the twenty intervals from A.D. 900 to A.D. 1899 average +0.71 for Mesa Verde versus Tsegi (115 mi W by S) and +0.46 for Mesa Verde versus Flagstaff (220 mi SW). The high significance of the Mesa Verde index as a rainfall history is illustrated in a later section.

In ring series which crossdate with one another the average proportionate change in width from year to year is a quantitative index of *sensitivity*. Highly sensitive series have been found to show a close relation to rainfall in the dry lands of the Southwest; at the same time, even in this region, trees on sites which permit significant water carry-over from year to year show low sensitivity

Methods [8]. Some special techniques of dendrochronology which have been found of value but which can only be mentioned here, are (1) use of the Swedish increment borer, which permits wide and rapid core sampling, (2) mounting and surfacing of cores in such a fashion as to permit ease and rapidity of cross-comparison of gross features as well as good definition of cell structure under the microscope, (3) elimination of individual age trends (varying mean growth rates) in order to average the ring records in uneven-aged trees of a set and thus remove a part of the ever-present random term in the growth of each tree and make the data for different time intervals comparable, (4) use of extensive areal averages to remove still more of the random term in the mean growth at each site, and (5) use of skeleton plots as a reconnaissance tool in archaeological dating.

Errors in the Unit. Possible errors in dendrochronologic series are of two types: (1) identification of growth irregularities, such as locally absent or false rings, and (2) interpretation of the ring fluctuations in terms of climatic variation.

In particularly severe years much of the cambial area may be entirely inactive in terms of new cell growth. Locally absent rings may occur with a frequency of as much as 5 per cent or more in the most sensitive and slow-growing ring series in semiarid regions [27]; they are of rare occurrence in more humid regions [17], at upper timberline, and in the Arctic. Such rings are identifiable by comparison of approximately parallel patterns in climatically related trees of a site or area.

False annual rings are comparatively rare in conifers in high latitudes and near upper timberline, but are very common in the latitudes of southern Arizona. Nontraumatic false rings are found in general to occur as a function of species, climate, environment, and age. In such dendrochronologic species as Douglas fir and ponderosa pine they may be uniquely identified by crossdating as well as (in almost all instances) by absolute criteria [25]. Some species, such as Arizona cypress, are so subject to false-ring layering as to be unusable in chronology building.

Errors of Interpretation. It has long been accepted as a botanical axiom that the absolute growth of plants is subject to a matrix of environmental and transmitted factors, whose interactions are extremely complex [15]. Even when the problem is simplified by the use of growth departures, it appears that only on sites critically limited with respect to an important growth element, such as rainfall, can such departures be highly representative of that element. There may be hidden effects in the growth record of past fires, pest outbreaks, mechanical injury, changes in associated flora or microorganisms, and other factors. Some species are characterized by an inherent tendency to nonuniform cambial activity, so that the ring series along any radius seems to have no rational explanation [24].

Some meteorological and hydrological factors, too, tend to destroy any simple relation between seasonal rainfall and tree growth: variation in distribution of rainfall within the season significant for ring growth, varying influence of temperature and other meteorological elements, real differences in total rainfall between the tree sites and the correlated meteorological station, and the numerous influences affecting the rainfall-runoff relationships from basin to basin.

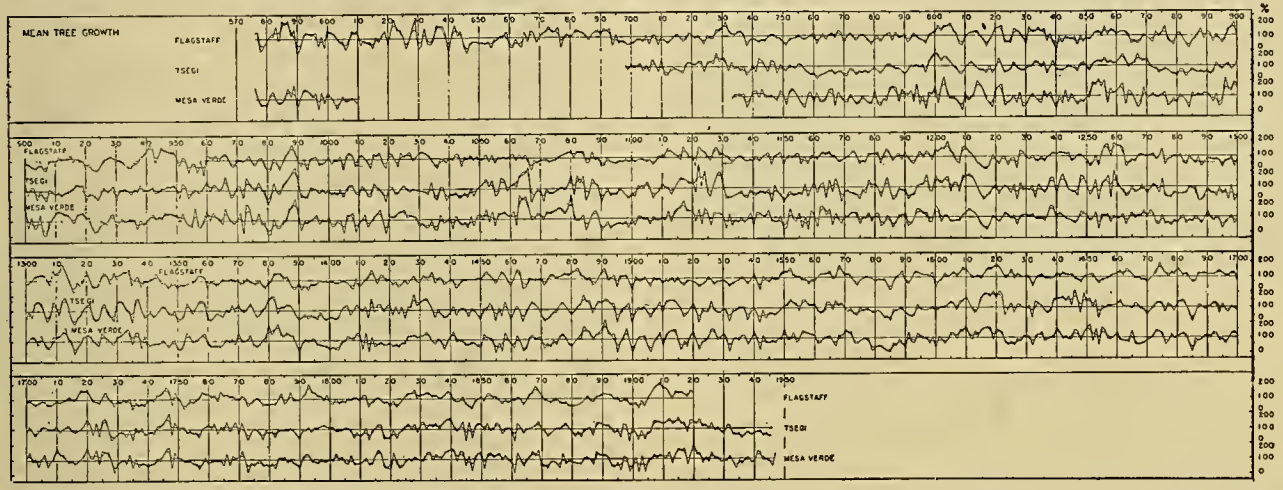


FIG. 1.—Comparison of the mean departures in tree-ring width for two localities in northern Arizona and Mesa Verde in southwestern Colorado. The records in living trees were extended backwards by use of overlapping series in archaeological beams. Smoothed curves were superposed to permit greater ease in visual comparison of the longer fluctuations. (*From Tree-Ring Bulletin.*)

It is thus a fundamental requirement for the production of significant dendroclimatic series that the ring record of any tree be checked against sufficient associated trees, and that the group mean, in which random errors have been permitted to cancel as far as possible, in turn be checked against and averaged with a sufficient number of associated groups.

Relation to Neighboring Sciences. Like most work outside the classical disciplines, dendrochronologic research is closely related to a number of fields: botany, ecology, forestry, meteorology, hydrology, archaeology, geology, and astronomy. The analyzed variable, tree-ring width, is in the domain of the first three of these fields. Its application to archaeology has

perhaps been the most spectacularly successful one to date, and has called forth very great cooperation from archaeologists in supplying specimens for analysis. Though tree-ring dating of ancient ruins in the Southwest is continuing and may well be extended to entirely new regions, it seems probable that the greatest potential contributions of dendrochronology will be in the fields of meteorology and hydrology.

TREE HISTORIES

Rainfall and River Flow in Western North America. Series which contain significant indications as to past rainfall and runoff have now been derived for a number of localities; for three of these, the Colorado River Basin as a whole [25], southern California [27], and eastern Oregon [20], the ring data are sufficiently sensitive and numerous to permit quantitative analysis in terms of climate. It must be recognized that no ring indices are yet available, nor is it likely that any will be derived, which are perfectly representative of the rainfall of any particular locality. Even the best of the present indices contain enough "random" variation to require caution in such quantitative conclusions. Yet these indices do show a fidelity to the march of rainfall which, in view of the potential refinements, leads to the expectation that a vast body of significant climatic data will eventually be obtained from selected trees. The more comprehensive drought chronologies now published are noted in Table I, which also lists the longer arctic chronologies of temperature.

TABLE I. TREE-RING INDICES*

Location and reference number	Number of trees	Rings measured (000)	Maximum length (yr)
Rainfall Chronologies			
Central California [5–I]	15	22	3219
Central Pueblo area (estimated) [7]			(to A.D. 150)
Colorado River Basin [25]	240	60	696
Eastern Oregon [20]	340	±35	668
Northern Arizona [5–I]	86	14	519
Oregon-California [3]	25	6	476
Pacific Slope, U. S. [27]	121	22	593
Southwestern Canada [26]	66	15	525
Southwestern U. S. ruins [28]	168	32	1551
Western Nevada [16]	±200	±13	685
Western United States [5-II]	305	52	371
Temperature Chronologies			
American Arctic [13, 14]	±250	±30	970
Finland [18]	214	11	55
Finland [11]	2641		50
Norway [1, 23]	+300	±50	537
Norway [10]	91	21	477
Sweden [12]	+300	±50	468

* Tabulations, in some cases of a number of series, are to be found in almost all cited references, supplemented by plotted curves. In some reports the total uumber of rings studied was several times the number used in the indices.

Throughout the West, sensitive timber conifers are found to provide histories primarily of winter precipitation, usually of the October–June interval. For several decades it has been generally accepted that the spotty convection storms of summer in the Rocky Mountains had no apparent systematic influence on ring-width fluctuations in these trees, though they were recognized as of occasional importance in producing false rings.[1]

Most fortunately for chronology building, it was found that the drought conifers show a tendency to extraordinary longevity, high sensitivity, and slow growth on particularly adverse sites. The extremely long unbroken records derivable from these trees provide the means for estimating the noncyclical trend in rainfall, if any, from one century to the next.

In developing a 658-year index for the Colorado River Basin the chronologies in a number of smaller drainage basins throughout the West were analyzed, the growth relationships in some of them being displayed in Fig. 2. Panel 1 of this figure is of special interest here, as it

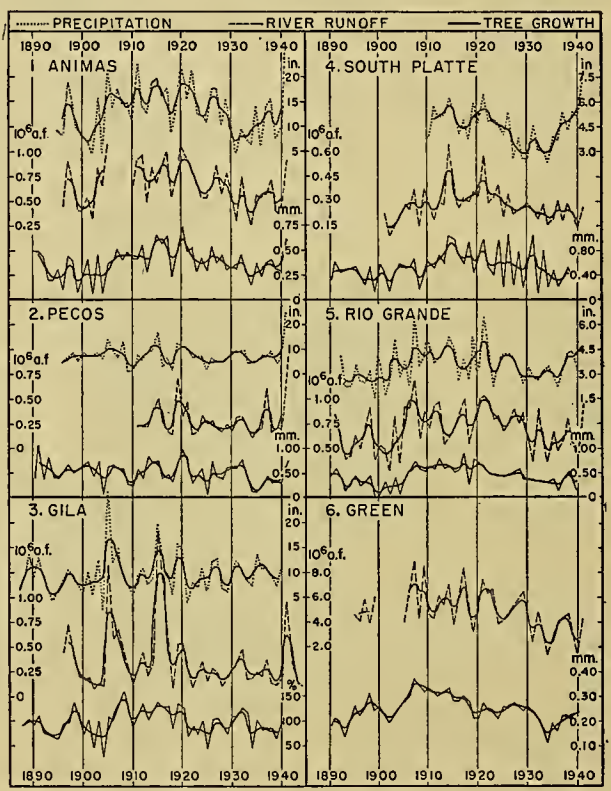


FIG. 2.—October–June rainfall and water-year runoff (in acre-feet) are compared with preliminary growth indices in six drainage basins of the Rocky Mountains. Superposed smooth curves emphasize the longer fluctuations.

compares the Mesa Verde growth index, plotted in Fig. 1 over its entire length of more than 1200 years, with the October–June rainfall at nearby Durango and the water-year runoff of the Animas River. The degree of fidelity and the limitations in the rainfall and runoff record in the best tree indices thus far derived are perhaps best

1. An extensive quantitative analysis now (June, 1951) in progress indicates that a small but significant additive influence of July–September rainfall on the growth of the *following* year is present in many ring series in this region. The reinterpretation of the numerous centuries-long indices of rainfall, now developed for the West, in terms of the total annual rainfall of July–June, rather than the more limited winter interval, may prove far more important climatically than the fact of a few per cent improvement in the tree-growth–rainfall correlations.

appreciated by a careful study of such comparisons as those in Fig. 2. Differences in growth response in the Gila area of southern Arizona and New Mexico, subject to flash floods, as compared with the conservative Green River area of Wyoming, may be specially noted.

The Douglas fir index for the Colorado River Basin above Hoover (Boulder) Dam is seen (Fig. 3) to provide a fair first approximation to the winter rainfall chronology of that region and to the river runoff. Independent 342-year indices for two types of pines support in general the main index, but show detailed differences in mean growth, which are probably only partly traceable to the relatively small number of component specimens in the pines. Correlation coefficients are shown in Table II.

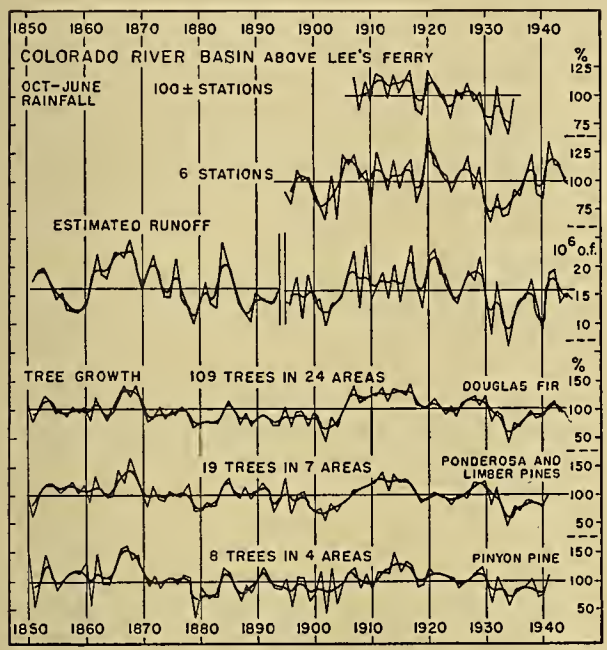


Fig. 3.—Regional tree-ring indices in the Colorado River Basin compared with gage data. Rainfall is expressed in per cent of the mean; tree growth in per cent of the mean trend; runoff in acre-feet × 10^6. The runoff of the Colorado River preceding 1895 is an approximation by E. C. LaRue (U.S.G.S. Water-Supply Paper 395, 1916) on the basis of Salt Lake levels. (*From University of Arizona Bulletin.*)

The giant sequoia of California has yielded by far the longest ring sequences, four of the trees analyzed by Douglass being over 3000 years of age [5]. Unfortunately, the relatively moist habitat of this species permits a sensitivity to rainfall fluctuations which is at best on a much lower level than that of drought-site trees.

The big-cone spruce of southern California has, however, been found to show a sensitivity rivalling that of the best trees of the Colorado Plateau; a 560-year record has been constructed. Its value as an index of rainfall and runoff is indicated in Fig. 4.

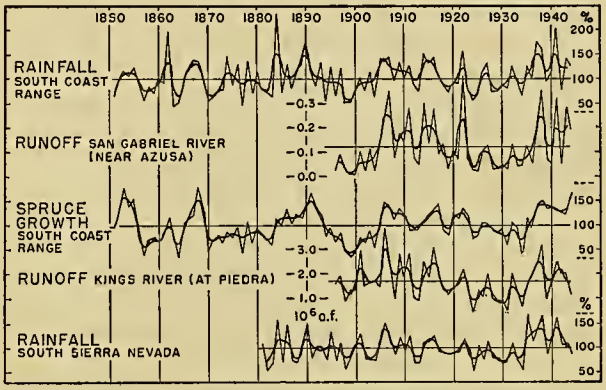


Fig. 4.—Regional tree-ring indices for southern California compared with gage data. Rainfall is expressed in per cent of the mean; tree growth in per cent of the mean trend; runoff in acre-feet × 10^6. (*From University of Arizona Bulletin.*)

The third regional survey of broad character, that of Keen [20] in the ponderosa pines of semiarid eastern Oregon, makes possible some centuries-long comparisons with southern California and the Colorado River Basin. The correlation coefficients of the Oregon index with rainfall and runoff are seen in Table II to be of the same order of magnitude as those in the other two regions.

The analysis by Antevs [3] in the general area studied by Keen is particularly useful in its detailed discussion of the complex of factors influencing the observed growth departures in recent decades. Among other analyses, the study by Hardman and Reil [16] in the Truckee River Basin provides extensive data.

Table II. Correlation Coefficients Between Gage Records and Regional Tree-Ring Indices

Tree index	Correlated with	Period	Interval (yr)	r unsmoothed	r smoothed*
Colorado River Basin	Colorado River runoff	Oct.–Sept.	49	+0.66	+0.81
Eastern Oregon	Eastern Oregon rain	Sept.–Aug.	66	+0.50	—
" "	Columbia River runoff	Jan.–Dec.	57	+0.56	—
Southern California	Coastal Rain	July–June	94	+0.65	+0.75
" "	" "	" "	49	+0.71	+0.85
" "	King's River runoff	Oct.–Sept.	49	+0.52	+0.64
" "	San Gabriel River runoff	Oct.–Sept.	49	+0.61	+0.86

* By formula $b' = \frac{a + 2b + c}{4}$.

Temperature Indices in the Arctic. Tree-ring chronologies in Scandinavia exceed in volume those for any other region outside of the southwestern United States. Two factors seem largely responsible: the major economic importance of forests, and the stimulus of the work begun in 1879 by de Geer [4] on the annual clay layers or varves.

Problems associated with the recognition and dating of false annual or locally absent rings, so important in

the Southwest, are almost nonexistent for the simple though narrow-ringed arctic series discussed by Erlandsson [12], Ording [23], and others. On the other hand, the fluctuations in ring-widths in arctic series tend to be relatively small, and the chronologies vary from site to site more than those in the Southwest.

Almost all arctic studies have noted the relation of ring-width fluctuations to the fluctuations in mean or average maximum summer temperature, particularly that of the June–July interval, the highest correlation coefficients being of the order of +0.7. As one would expect, precipitation is in general not an important limiting factor in growth. The longer tree indices of arctic temperatures are noted in Table I.

Good series for numerous areas in northern Alaska have been developed by Giddings [13, 14], who finds that the more sensitive tree records represent June–July mean temperatures with some fidelity. Extending living-tree records by the use of archaeological beams, he has developed one series almost a thousand years in length.

Indices of Physiological Drought in the Eastern United States. In the absence of a dominant climatic variable, it does not appear possible to derive a simple meteorological index from the ring-growth in the nonmarginal areas, although some correlations of larger growth departures with rainfall or temperature of various spring and summer months are reported by Lyon [21, 22] and others. An example of a fair degree of cross-dating over a considerable area is furnished by the analysis of hemlock growth in New England—an indication that trees in relatively moist areas can provide a significant history representing the resultant effect on plant growth of a complex of climatic conditions.

LONG-TERM CLIMATIC FLUCTUATIONS

Changing Areas of Drought. A sufficient number of ring chronologies of rainfall in the western United States have been developed to permit a preliminary view of this phenomenon over the centuries. The tendency towards a variability in the area affected by drought in different years, a striking characteristic of gage records of rainfall and runoff, applies also to longer drought intervals and is evident over the entire range of the data. Centuries-long chronology transects for 1500 miles or more of the Pacific Slope and the front ranges of the Rocky Mountains reveal no systematic time displacements in maxima or minima from one area to another of sufficient significance to permit their use in long-term weather forecasting. Perhaps the most important element emphasized by this analysis is the need for caution in generalizing, either in time or areal extent, any local indications of cyclic recurrence, drought frequency, trends, and similar phenomena.

Dry and Wet Years and Intervals. The three regional chronologies noted above, for the Colorado River Basin, southern California, and eastern Oregon, each on a broad statistical base of sensitive ring records, may be used to derive long-term statistical parameters of interest to the climatologist. It appears that specially dry winters and low runoff in the Colorado River Basin, of the severity of the "dust-bowl year" 1934, occurred about once in twenty years during the 658 years of ring record (not *every* twenty years!); instances of two successive years of such critically small runoff, of special importance in the management of the reservoir level at Hoover Dam, seem to have occurred only three times during that record. In the Colorado River index, after simple three-term smoothing of the type noted in Table II, the median length of an interval of general excess or deficit is about six years, intervals of deficiency or excess more than three times this length occurring once in about 200 years. In southern California the median length is about 8 years in the 560 years of similarly smoothed record. Intervals of excess and deficit are in general somewhat longer in Keen's Oregon chronology, an effect in part, at least, of the heavier smoothing.

The Colorado Basin index indicates that the interval of greatest deficiency in rainfall and runoff in this region since A.D. 1300 occurred in A.D. 1573–1593, when the average deficiency was well below that in the most severe intervals recorded by modern gages. It is significant that the greatest minimum in the southern California index occurred in 1571–1597, when the accumulated deficit approached twice that of any interval in the century or so of rain-gage records. The eastern Oregon index shows only a weak general deficiency from 1565 to 1599, broken by a short interval of excess near 1587; the most pronounced minimum in that index began in 1917 and was followed by "gradual encroachment of desert conditions into what were once thriving pine stands" [20] during the next two decades.

Relevant to this is the interregional correlation coefficient, computed for the interval 1650–1935, of +0.46 for the Colorado Basin versus southern California, which compares with a coefficient of +0.21 for the same interval between the Colorado Basin and eastern Oregon; very similar results are obtained when winter rainfall data in the three regions are correlated. No significant correlation between British Columbia growth and that in the Colorado River Basin has been noted for the last three centuries or so of data.

No long-term trend has yet been found in any of the major ring indices which could be reliably interpreted as a secular change in climate, though temporary fluctuations of considerable magnitude and duration are evident throughout the records. Beams in Southwestern ruins, which were cut as much as 1700 years ago, show about the same ring sensitivity as living trees on the same sites, a strong hint that no decidedly more moist or dry conditions prevailed at those times.

Cycle Analysis. Douglass has pointed out the persistent tendency of many tree-ring series to show fluctuations of 22–23 years and fractions and multiples of this cycle, which seem related to sunspot or possibly other extraterrestrial phenomena. Cycle lengths of this order are to be found in the ring series in fossil woods and have been reported in the Scandinavian ring chronologies of temperature. The cycles found in tree growth seem, however, to be characterized by variability in length, amplitude, and form, to tend to appear and disappear according to no discernible law, and to occur in almost any combination. A satisfactory physical

explanation of these characteristics has not yet been developed.

LINES OF FUTURE DEVELOPMENT

It is probable that the contributions of tree-ring analysis to meteorological and hydrological knowledge have only just begun to be manifest. Advances will depend on two related developments: the construction of growth indices for new areas and, in areas already studied, the replacement or amplification of indices by others of greater fidelity to the limiting climatic variable. The profound importance of the latter property of ring chronologies cannot be overemphasized.

Reconnaissance has shown that rainfall chronologies of good sensitivity may be found in selected trees in the headwaters areas of the Missouri, Snake, and other major rivers of the western United States; indices for these basins comparable to that already developed for the Colorado River Basin are now being constructed.

Potential sources of rainfall chronologies appear to exist in the mid-latitude Andes Mountains, the Mediterranean area, the northwest provinces of India, and elsewhere. The vast Siberian Arctic is, as far as is known, totally untouched as a source of temperature chronologies.

The refinement of tree-ring data is primarily related to chronologies in semiarid regions. Even the most elaborate index of this type thus far available, that for the Colorado River Basin developed in 1945, is subject to considerable improvement. The last century of this index is based on some 100 Douglas fir trees from 24 stations within the basin, supplemented by two indices in other species from seven and four stations respectively; earlier portions of the index are progressively weaker, the main index at A.D. 1500, for example, representing 10 stations and 34 trees, supplemented by one minor index for one station. Field work since 1945 has shown that at least 50 stations within this basin can yield sensitive records from 500 to 800 years old, in both Douglas fir and pinyon pine, from which it is possible to derive indices of substantially higher fidelity to rainfall and runoff, particularly in the earlier centuries. Such an extension is now in progress.

As the world map of significant tree-ring chronologies of rainfall or temperature becomes more complete, some estimate may become possible of the large-scale fluctuations in the weather year by year for many centuries. Long-term statistical frequencies found in the climatic records in trees of any area should have greater significance in the light of similar data from many other areas. The application of such knowledge to practical problems in such domains as reclamation, conservation, and water power would seem to be immediate. Of great potential importance is the possible bearing of climatic research in dendrochronology on the problems of long-range climatic forecasting.

REFERENCES

1. Aandstad, S., "Die Jahresbestimmung der Kiefer und die Zeitbestimmungen älterer Gebäude in Solör im östlichen Norwegen." *Nyt Mag. Naturv.*, 78: 201–268 (1938).
2. Antevs, E., "Die Jahresringe der Holzgewächse und die Bedeutung derselben als klimatischer Indikator." *Progr. Rei bot.*, 5: 285–386 (1917).
3. —— "Rainfall and Tree Growth in the Great Basin." *Spec. Publ. Amer. geogr. Soc.*, No. 21: *Carneg. Instn. Wash. Publ.* 469 (1938).
4. de Geer, G., "Geochronologia Suecica Principles." *K. svenska VetenskAkad. Handl.*, Ser. 3, Bd. 18, No. 6 (1940).
5. Douglass, A. E., "Climatic Cycles and Tree-Growth." *Carneg. Instn. Wash. Publ.* 289: I, 127 pp. (1919); II, 166 pp. (1928); III, 171 pp. (1936).
6. —— "The Secret of the Southwest Solved by Talkative Tree Rings." *Nat. geogr. Mag.*, 56: 736–770 (1929).
7. —— "Estimated Ring Chronology, 150–1934 A. D." *Tree-Ring Bull.*, 6: 39 (1940).
8. —— "Notes on the Technique of Tree-Ring Analysis." *Tree-Ring Bull.*, 7: 2–8 (1940); 7: 28–34 (1941); 8: 10–16 (1941); 10: 2-8 (1943); 10: 10–16 (1943).
9. ——"Precision of Ring Dating in Tree-Ring Chronologies." *Bull. Univ. Ariz.*, Vol. 17, No. 3 (1946).
10. Eidem, P., "Über Schwankungen im Dickenwachstum der Fichte (Picea Abies) in Selbu, Norwegen." *Nyt Mag. Naturv.*, 83: 145–189 (1943).
11. Eklund, B., "Ett försök att numeriskt fastställa klimatets inflytande på tallens och granens radietillväxt." *Norrlands SkogsvFörb. Tidskr.*, 3: 193–226 (1944).
12. Erlandsson, S., *Dendrochronological Studies.* Stockholms Högskolas Geokronol. Inst., Data 23, 1936.
13. Giddings, J. L., Jr., "Dendrochronology in Northern Alaska." *Bull. Univ. Ariz.*, Vol. 12, No. 4: *Publ. Univ. Alaska* No. 4 (1941). See also "Chronology of the Kobuk-Kotzebue Sites." *Tree-Ring Bull.*, 14: 26–32 (1948).
14. —— "Mackenzie River Delta Chronology." *Tree-Ring Bull.*, 13: 26–29 (1947).
15. Glock, W. S., "Growth Rings and Climate." *Bot. Rev.*, 7: 649–713 (1941).
16. Hardman, G., and Reil, O. E., "Relationship between Tree Growth and Stream Runoff in the Truckee River Basin, California-Nevada." *Bull. Nev. agric. Exp. Sta.*, No. 141 (1936).
17. Huber, B., und Holdheide, W., "Jahrringchronologische Untersuchungen an Hölzern der bronzezeitlichen Wasserburg Buchau am Federsee." *Ber. dtsch. bot. Ges.*, 50: 261–283 (1942).
18. Hustich, I., "The Scotch Pine in Northernmost Finland and Its Dependence on the Climate in the Last Decades." *Acta bot. fenn.*, 42 (1948).
19. Kapteyn, J. C., "Tree Growth and Meteorological Factors." *Rec. Trav. bot. néerland.*, 11: 71–93 (1914).
20. Keen, F. P., "Climatic Cycles in Eastern Oregon as Indicated by Tree-Rings." *Mon. Wea. Rev. Wash.*, 65: 175–188 (1937).
21. Lyon, C. J., "Water Supply and the Growth Rates of Conifers around Boston." *Ecology*, 24: 329–344 (1943).
22. —— "Hemlock Chronology in New England." *Tree-Ring Bull.*, 13: 2–4 (1946).
23. Ording, Å., "Årringanalyser på Gran og Furu." *Medd. norske Skogsforsøksv.*, 7: 105–354 (1941).
24. Schulman, E., "Centuries-Long Tree Indices of Precipitation in the Southwest." *Bull. Amer. metcor. Soc.*, 23: I, 148–161; II, 204–217 (1942).
25. ——"Tree-Ring Hydrology of the Colorado River Basin." *Bull. Univ. Ariz.*, Vol. 16, No. 4 (1945).
26. —— "Dendrochronologies in Southwestern Canada." *Tree-Ring Bull.*, 13: 10–24 (1947).
27. —— "Tree-Ring Hydrology in Southern California." *Bull. Univ. Ariz.*, Vol. 18, No. 3 (1947).
28. —— (Reports on Climatic-Archaeological Indices in the Southwest.) *Tree-Ring Bull.*, 12: 18–24 (1946); 14: 2–8 (1947); 14: 18–24 (1948); 15: 2–14 (1948); 15: 24–32 (1949); 16: 10–14 (1949).

HYDROMETEOROLOGY

Hydrometeorology in the United States *by Robert D. Fletcher* 1033

The Hydrologic Cycle and Its Relation to Meteorology—River Forecasting *by Ray K. Linsley* 1048

HYDROMETEOROLOGY IN THE UNITED STATES

By ROBERT D. FLETCHER

U. S. Weather Bureau, Washington, D. C.

INTRODUCTION

Scope of Hydrometeorology. Meteorological forecasts and advices, though varying greatly as to type, tend to fall into more or less identifiable groups. One such group has come to be known as *hydrometeorology*, that branch of meteorology which deals with the rainfall statistics of storms in order to determine, theoretically or empirically, relationships between meteorological factors and the precipitated moisture which reaches the ground. It is especially concerned with those atmospheric processes which yield the water dealt with in hydrology.[1] In practice, the hydrometeorologist develops estimates—usually in terms of the "maximum possible"—of the space-time distribution of rain over specified drainage areas. The hydrologist uses these values in estimating the flood hydrographs which would result.

The boundaries of hydrometeorology are not sharply defined. The hydrometeorologist, like the hydrologist, is concerned with snow melt, since it is intimately involved with meteorological processes. The cloud physicist and hydrometeorologist, on microscopic and macroscopic scales, respectively, deal with the formation of rain and snow in clouds. The weather forecaster and hydrometeorologist both prepare quantitative predictions of precipitation, although the hydrometeorologist usually does not forecast when precipitation of a given magnitude will occur, but rather that certain maximum intensities can occur over particular areas.

Wherever problems concerned with the accumulation of water from rainfall or snow melt are found, there is a need for hydrometeorological advices or forecasts. In recent years, however, systematic hydrometeorological services have been established largely to provide bases for the design of flood-control and water-usage structures. Among the better known of the hydrometeorological services in the United States are those of the U. S. Weather Bureau. That agency's Hydrometeorological Section was organized to provide hydrometeorological advice for the Department of the Army, Corps of Engineers, which is concerned with the design of flood-control structures. The Weather Bureau's Cooperative Studies Section was established primarily to perform a similar function for the Department of the Interior, Bureau of Reclamation, which is concerned with water conservation and irrigation. The regular duties of the Weather Bureau forecast centers include the quantitative prediction of precipitation intensities over specified areas. Such information is applicable to the operations of flood-control and water-usage structures.

1. Consult "The Hydrologic Cycle and Its Relation to Meteorology—River Forecasting" by R. K. Linsley, pp. 1048–1054 in this Compendium.

Maximum Possible Precipitation. For a given area and duration, the depth of precipitation which can be reached, but not exceeded, is defined as the *maximum possible precipitation*. Since the laws governing precipitation rates are not completely known, this figure is an estimate which implies a range of tolerance, the extent of which depends upon limitations in technical knowledge, data, and thoroughness of analysis. The values computed from the analysis are considered to be the maximum possible since, within the confines of current theory and available data, they are derived from the most productive combinations of the factors controlling rainfall intensities.

In general, three steps are followed in the derivation of maximum possible precipitation values. First a storm model is established—usually in the form of an equation —which expresses rainfall as a function of independent meteorological variables which are not only measurable but also well represented as to quantity and length of record. Theoretical meteorology is basic in the establishment of the equation form: empiricism is utilized in the selection of proportionality factors, constants of integration, and the like. Second, the model is tested through the introduction of observed values for the independent variables. Comparison between observed and computed rainfall rates provides a basis for the estimation of the reliability of the model. Third, the independent variables are maximized statistically. Introduction of the maximum values of these variables, or the most effective combination of them, into the successful rainfall equation produces the estimate of maximum possible precipitation.

BASIC DATA AND GRAPHICAL PRESENTATIONS

Rainfall Intensity, Depth, and Volume. The history of hydrometeorological investigation has demonstrated the need by the field, at one time or another, for nearly every element of the standard meteorological observation. Of considerable importance are (1) humidity, to determine the quantity of moisture which can be precipitated, (2) pressure, to determine the mass of air which can enter into the precipitation process, and (3) wind, to determine the rate at which precipitation can take place. The fundamental element, however, is the precipitation itself. This measurement is one of *intensity* of precipitation, or rate of accumulation of precipitation in the measuring device. It is thus a measurement of the average *depth* of precipitation through duration, the minimum length of duration obtainable being a function of the time-sensitivity of the measurement. It is also a measure of the average *volume* of precipitation through duration, the volume equalling the product of the

average depth and the catchment area of the measuring equipment.

Depth-Duration-Area Data. The user of hydrometeorological advices primarily requires information in the form of average depths of precipitation, duration by duration, over the areas of specified drainage basins. The areas may be as large as, for example, the 145,500 square miles of the Colorado River drainage basin above the Bridge Canyon dam site [26] or they may be small as in the case of the design of storm sewers over a few city blocks. For large areas, the total amount of water which can collect over a period of days is the dominant factor in most design problems. For small areas, design is controlled by high intensities for durations measured in hours or even minutes. The hydrometeorologist is, accordingly, required to determine values of average depth over duration for areas of a larger order of magnitude than those of the catchments of rain gages. He deals with what can be considered as the uniquely hydrometeorological "depth-duration-area" (DDA) values derived from the processing of the essentially "point-rainfall" measurements of the rain gage. Because of the peculiarities of hydrologic needs, most DDA data and estimates are either for large areas and long durations or for small areas and short durations. Because the other two types of combinations are rarely needed, few storms critical as to such combinations have been processed.

In each storm of record, for each standard area (*e.g.*, 10, 100, or 10,000 square miles) and for each duration (*e.g.*, 6, 12, 48, or 120 hr), there exists a maximum depth of rainfall. An array of such data is called the set of "maximum DDA values" for the storm. The Corps of Engineers, in cooperation with the Weather Bureau, is conducting a continuing program of developing maximum DDA values for major rainstorms of record in the United States. Approximately 600 storms have been processed up to the present time. Table I contains the greatest depths together with a list of the storms which produced them. Glasspoole [7] has presented similar data for the United Kingdom, processed by a method slightly different from that used in this country.

Mass Curves of Rainfall. Graphs of accumulated precipitation versus time, called "mass curves" of rainfall, may be prepared in considerable detail in the case of data from recording gages. From 24-hr observation stations, or from stations recording only total-storm amounts, mass curves can be synthesized by referring to the records of nearby recording-gage stations, to synoptic meteorological analysis, and to "double-mass analysis" [11]. The set of mass curves for a single storm is basic to the development of the maximum DDA values, and for the construction of isohyetal maps—for increments of six hours, say, as well as for the total storm. Obviously, the more mass curves available, the more accurate the results will be. DDA values may be derived by planimetering of the isohyetal maps, or by employment of area-weighting procedures such as the "Thiessen polygon" method [24]. A combination of the two methods is now in standard use in the United States.

Conversion of point-rainfall to DDA values involves the assumption that each rain-gage measurement is representative of a more or less large increment of area surrounding it. In its report on thunderstorm rainfall, the Hydrometeorological Section [27] has discussed in detail the reliability of areal rainfall determination.

TABLE I. MAXIMUM OBSERVED RAINFALL DEPTHS IN THE UNITED STATES*

Area (sq mi)	Duration (hr)						
	6	12	18	24	36	48	72
	in.	*in.*	*in.*	*in.*	*in.*	*in.*	*in.*
10	24.7*a*	29.8*b*	35.0*b*	36.5*b*	37.6*b*	37.6*b*	37.6*b*
100	19.6*b*	26.2*b*	30.7*b*	31.9*b*	32.9*b*	32.9*b*	35.2*c*
200	17.9*b*	24.3*b*	28.7*b*	29.7*b*	30.7*b*	31.9*c*	34.5*c*
500	15.4*b*	21.4*b*	25.6*b*	26.6*b*	27.6*b*	30.3*c*	33.6*c*
1,000	13.4*b*	18.8*b*	22.9*b*	24.0*b*	25.6*d*	28.8*c*	32.2*c*
2,000	11.2*b*	15.7*b*	19.5*b*	20.6*b*	23.1*d*	26.3*c*	29.5*c*
5,000	8.1*b*,*j*	11.1*b*	14.1*b*	15.0*b*	18.7*d*	20.7*d*	24.4*d*
10,000	5.7*j*	7.9*k*	10.1*e*	12.1*e*	15.1*d*	17.4*d*	21.3*d*
20,000	4.0*j*	6.0*k*	7.9*e*	9.6*e*	11.6*d*	13.8*d*	17.6*d*
50,000	2.5*e*,*h*	4.2*g*	5.3*e*	6.3*e*	7.9*e*	8.9*e*	11.5*f*
100,000	1.7*h*	2.5*h*,*i*	3.5*e*	4.3*e*	5.6*e*	6.6*f*	8.9*f*

* The letter after the rainfall depth identifies the particular storm which gave that depth according to the following key:

Storm	*Date*	*Location of rainfall center*
a	July 17–18, 1942	Smethport, Pa.
b	Sept. 8–10, 1921	Thrall, Tex.
c	Aug. 6–9, 1940	Miller Island, La.
d	June 27–July 1, 1899	Hearne, Tex.
e	March 13–15, 1929	Elba, Ala.
f	July 5–10, 1916	Bonifay, Fla.
g	April 15–18, 1900	Eutaw, Ala.
h	May 22–26, 1908	Chattanooga, Okla.
i	Nov. 19–22, 1934	Millry, Ala.
j	June 27–July 4, 1936	Bebe, Tex.
k	April 12–16, 1927	Jeff. Plaq. Sta., La.

Figure 1, reproduced from this report, illustrates the variation in isohyetal-pattern analysis which can result from a variation in the number of reporting stations. For large-area and/or long-duration storms, existing rain-gage networks are usually adequate for computations of DDA values. But only rarely can detailed isohyetal patterns—and thus, accurate DDA values—be determined for the intense, local, short-duration bursts of rainfall which sometimes occur in thunderstorms.

PHYSICAL BACKGROUND OF HYDROMETEOROLOGY

Empiricism and Theory. It has long been recognized that maximum intensities of point rainfall decrease as the durations through which they are computed increase. A number of empirical formulas expressing relations between duration and maximum intensity have been developed, differences between them being ascribable to the availability and type of data used in the formulation of the relationships, geographical regions of application, and similar factors. Since it is also recognized that maximum average depths of rainfall decrease as the area over which they are averaged increases, other empirical relations have been derived to include the factor of area.

Empirical formulas can represent, with an accuracy

adjustable by variation of the complexity of the equations, the envelopes of rainfall occurrences. Their extensions to include the maximum possible precipitation require adoption of arbitrary measures—the use of

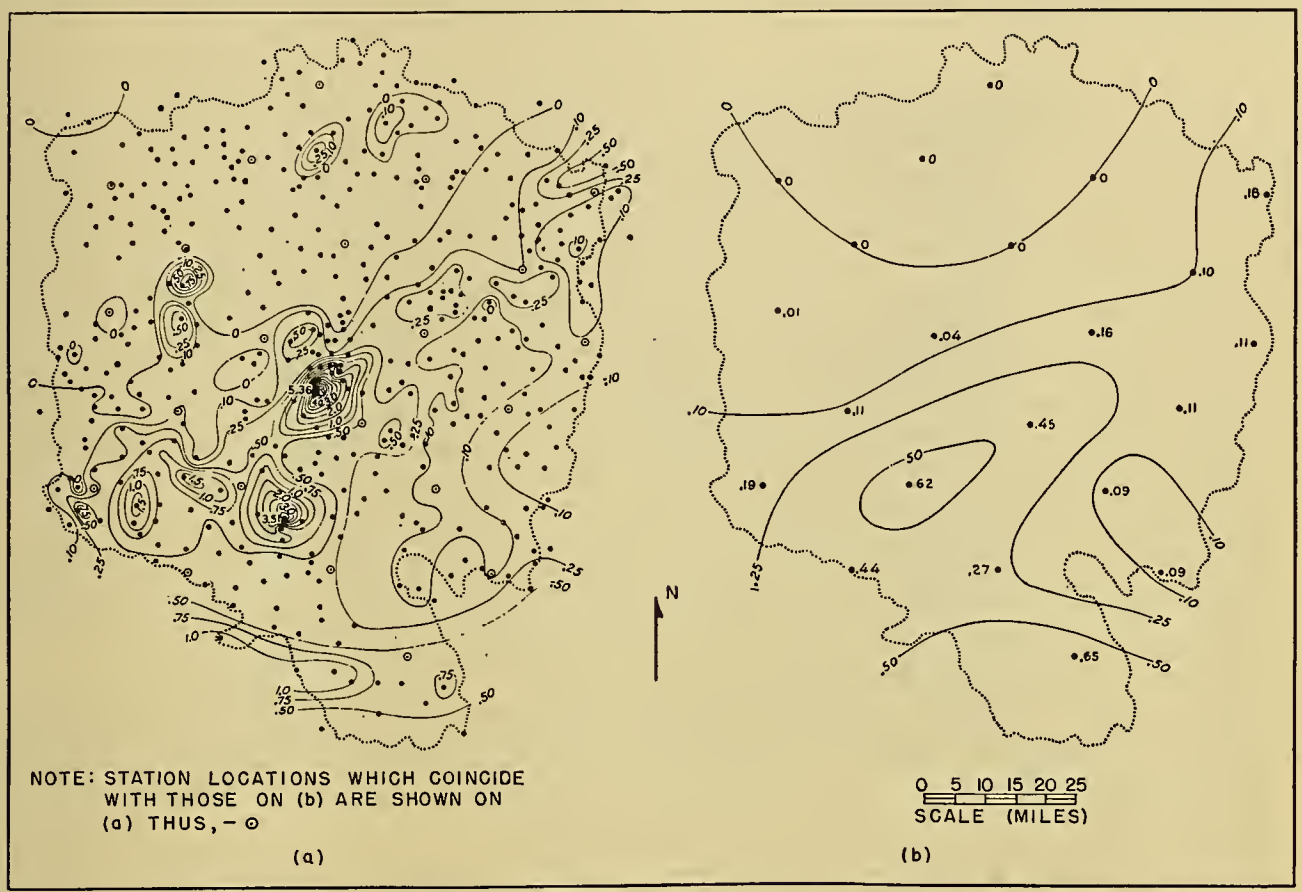


FIG. 1.—Isohyetal maps of storm of August 3, 1939, Muskingum Basin, Ohio; (*a*) 449 rain gages (1 gage per 18 sq mi); (*b*) 22 rain gages (1 gage per 375 sq mi).

safety factors, for example. With methods based upon meteorological theory, it is possible to derive enveloping relationships through the use of meteorological data other than those of precipitation, and to develop the forms of the extensions of these relationships beyond the highest observed depths of precipitation. Such methods have been found to be most suitable for the majority of hydrometeorological problems. Nevertheless, empirical or statistical approaches are still followed whenever theoretical knowledge is not sufficiently complete or whenever available data are inadequate for the test and usage of known theoretical relations. Future advances in hydrometeorology therefore hinge upon further development of basic theory and upon the acquisition of necessary and sufficient data.

Precipitable Water. The immediate source of precipitation is the water-vapor content of the air. For analytical purposes, this quantity is customarily expressed in terms of the depth of "precipitable water" in a column of air,

$$W = \int_0^{p_0} (q/g\rho_w)\, dp, \qquad (1)$$

where g is the acceleration of gravity, ρ_w is the density of water, q is specific humidity, and p_0 is the pressure at the bottom of the air column in which W is the precipitable water. As has been pointed out by Solot [21], when pressure appears in units of millibars, W (in inches) is closely approximated by

$$W = 0.4 \int_0^{p_0} q\, dp. \qquad (2)$$

Increments of W can be computed when the variation of q with pressure is known.

The basic obstacle in the way of general use of the theoretical expression for W is the almost complete absence of upper-air sampling in the major rainstorms of record. Accordingly, hydrometeorologists have been forced to make certain assumptions regarding the relation between W and other meteorological elements, well represented areally and historically. The most common such assumption is that, in major rainstorms, the air in which the rain is formed is saturated throughout its depth and is characterized by a pseudoadiabatic lapse rate. Here, the depth of precipitable water in an air column above a specified pressure surface is a single-valued function of the dew point (or temperature) at that surface. Figure 2 shows the relation, based on this assumption, between W and surface dew point when

the pressure at the surface is 1000 mb. From the graph the depth of precipitable water between any two pressure levels aloft can also be obtained when the potential dew point is known. Precipitable-water tables [32] for combinations of variables involving either pressure or are appreciably reduced in most practical applications. To achieve greater realism, however, refined assumptions can be formulated from synoptic studies of temperature and moisture stratifications, storm type by storm type. For instance, it is apparent that a saturated,

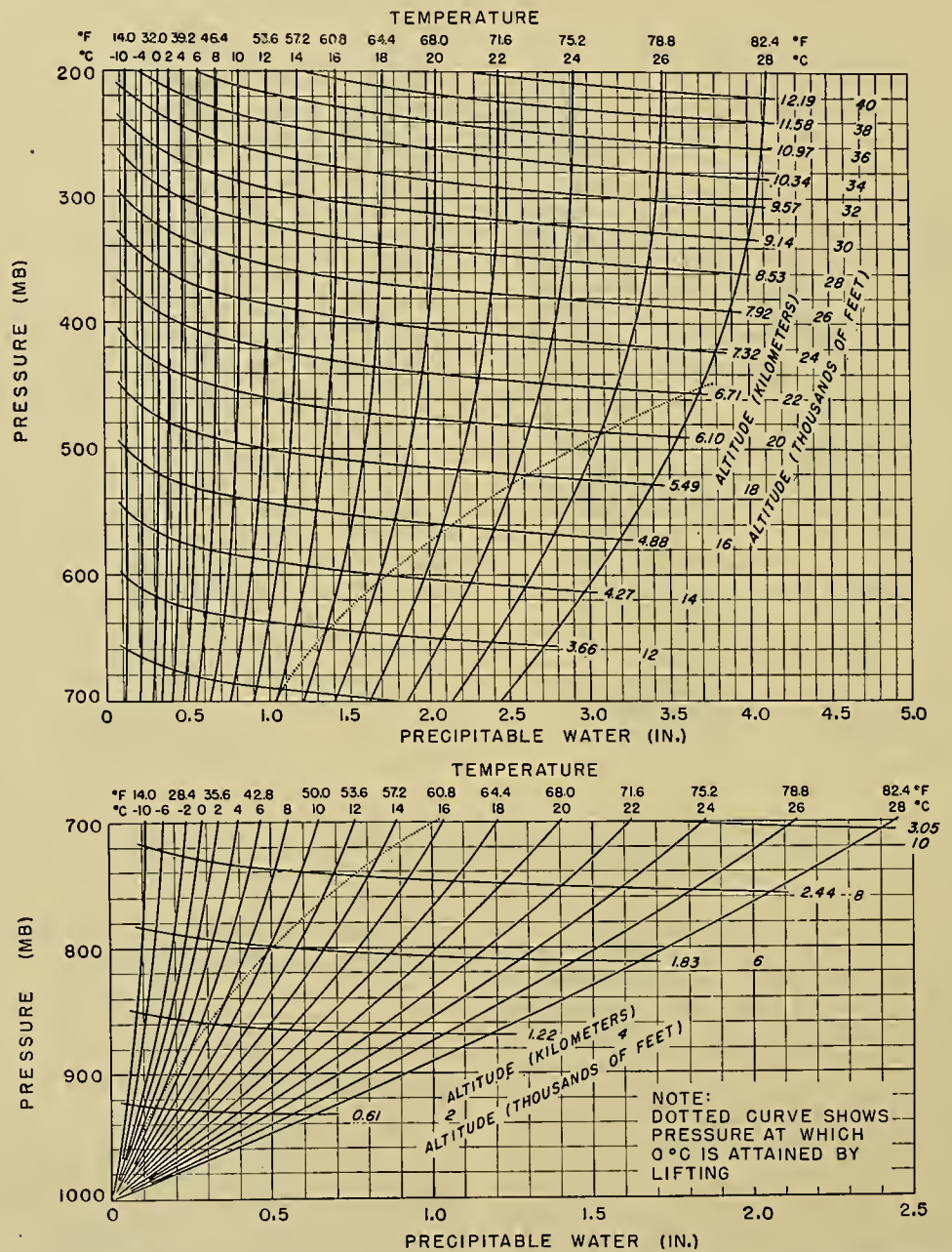


FIG. 2.—Depths of precipitable water in a column of air of given height above 1000 mb, under the assumption of saturation with a pseudoadiabatic lapse rate for the indicated surface temperatures.

height have recently been prepared for hydrometeorological use.

The few tests of this assumption which have been possible in heavy rainstorms indicate that significant deviations may sometimes occur. Since, as will be discussed later, the most common use of precipitable-water values in hydrometeorology involves ratios of W at various dew points, errors inherent in the assumption pseudoadiabatic atmosphere is not particularly close to reality in the case of convective instability, a circumstance which frequently characterizes sizable rainstorms. Whether the maximum possible storm possesses a pseudoadiabatic lapse rate in saturated air has not been established, but this is usually accepted as a reasonable assumption. Investigation is needed to determine whether it is possible to maximize theoretical

relationships to develop optimum distributions of moisture, temperature, and wind.

Vertical Motion. The precipitable-water depth is the amount of moisture available for precipitation in an air column. By itself, however, it does not provide sufficient information for computations of precipitation intensities. Lifting of a saturated column produces a cooling according to the pseudoadiabatic lapse rate, and therefore a decreasing water-vapor content. Thus, if the lift-speed of the column is known in detail, the rate of decrease of W—and hence the rate of condensation—can be computed. If the cloud-particle content of the air column remains fixed, the condensation rate becomes the precipitation intensity. For saturated air with constant cloud-particle content, precipitation intensities can be determined by means of a diagram prepared by Fulks [6], which shows the condensation rate, per unit vertical lift, in 100-m layers of air for various values of pressure and temperature. The equation for intensity of precipitation I, under the same assumptions, is

$$I = \int_{z_0}^{z} V_z \rho \frac{\partial q}{\partial z}\, dz, \tag{3}$$

where V_z is upward velocity and ρ is air density. An integrated form of this equation has been discussed by Showalter [19]:

$$I = V_{z0}\rho_0(x_0 - x_1)/7 \text{ inches per hour}, \tag{4}$$

where V_{z0} is the vertical velocity at the bottom of the air column in which the rain is produced, ρ_0 is the density at the bottom, and x_0 and x_1 are the mixing ratios at the bottom and top, respectively, of the column.

Use of equation (3) requires detailed knowledge of both q and V_z. The volumetric distribution of humidity can be estimated from available aerological soundings or, indirectly, from surface observations in cases of heavy rain. Only under very special conditions, however, can the space distribution of vertical velocity be estimated with an accuracy sufficient for hydrometeorology, and direct measurements of this factor are almost completely lacking. Miller [14] has discussed a number of ways of expressing vertical velocities in terms of quantities more amenable to measurement. Combinations of such expressions with equation (3) can produce various rainfall-intensity equations.

Equations of Continuity. Basic to all theoretically derived rainfall relations used in hydrometeorology is the equation of moisture continuity. This relation states that the average depth R of rainfall through duration D and over area A is equal to the moisture flowing into the space above A, minus that flowing out, plus the moisture above A at time $t = 0$, minus the moisture at $t = D$. The moisture q, to be exact, should be defined as the total number of grams of water in the air per gram of air, whether it be in the form of water vapor, cloud droplets, rain, or snow. Lacking sufficient data as to atmospheric content of liquid- or solid-water particles, hydrometeorologists have assumed that consideration of the water-vapor content, alone, leads to conclusions that are satisfactorily accurate. The assumption may involve important errors in the case of short-duration, small-area rainfall, the errors probably diminishing, however, with an increase of duration and/or area.

The equation of continuity for moisture may be written

$$R = (1/\rho_w A)\left[\int_0^D\int_0^\infty\oint q\rho V_1\, ds\, dz\, dt - \int_0^D\int_0^\infty\oint q\rho V_2\, ds\, dz\, dt + \left(\int_0^A\int_0^\infty q\rho\, dz\, dA\right)_{t=0} - \left(\int_0^A\int_0^\infty q\rho\, dz\, dA\right)_{t=D}\right], \tag{5}$$

where V_1 and V_2 are the horizontal inflow and outflow components, respectively, of the wind normal to a horizontal linear element ds of the vertical wall of the cylinder extending upward from the boundary of area A. The quantities V, q, and ρ must all be forecast if the equation is to be used for the quantitative prediction of precipitation. For purposes of estimating maximum possible precipitation, the most critical combination and values of the independent variables must be determined.

For the horizontal wind, unlike the vertical velocity, measurements are available. Indeed, successful attempts have been made to compute rainfall by direct application of equation (5), although q has had to be considered as a measure of water vapor only, and time and space interpolations of wind between pilot-balloon observations have been necessary. Unfortunately, while several meteorological projects have been undertaken, incorporating dense rain-gage networks and dense coverage of upper-air observations, a near-maximum rainstorm has never occurred when they were in simultaneous operation. As a consequence, investigations of major rainstorms have required approximations of equation (5), with substitutions of available measurements. Furthermore, the choice of elements to be used has necessarily been restricted to those with considerable abundance and length of record in order that methods and results could be comparable.

In some approaches to the problem, the continuity equation for air is used:

$$0 = \int_0^D\int_0^\infty\oint \rho V_1\, ds\, dz\, dt - \int_0^D\int_0^\infty\oint \rho V_2\, ds\, dz\, dt + \left(\int_0^A\int_0^\infty \rho\, dz\, dA\right)_{t=0} - \left(\int_0^A\int_0^\infty \rho\, dz\, dA\right)_{t=D} \tag{6}$$

in which the variables are the same as those appearing in equation (5). When simplifying assumptions are made, the use of equation (6) can bring about elimination of one of the more-difficult-to-measure variables of equation (5).

Storage Equation. Relationships (5) and (6) are quite general. When applied to certain simple flow models, they may be transformed into a rainfall equation which has demonstrated its ability to produce results comparable with observed rainfall amounts. In one such flow model—a two-dimensional one—the wind blows

across a rectangular area of size A and length Y. The inflow wind V_1 blows into the model at low levels, through a depth Δp_1, above which zero wind movement is defined. The outflow wind V_2 blows out of the model at high levels, through a depth Δp_2, above and below which there is no wind movement. There is assumed to be no net accumulation of air or atmospheric water content through duration D. Values used are all assumed to be averages through D. The quantities V_1 and V_2 are taken as averages through the respective Δp's, and the averages with respect to dp are assumed to be equal to the averages with respect to qdp. With these assumptions, equations (5) and (6) can be integrated. Elimination of the outflow wind results in the so-called "storage equation,"

$$R = \frac{V_1 D}{Y}\left(W_1 - \frac{\Delta p_1}{\Delta p_2} W_2\right), \tag{7}$$

in which W_1 and W_2 are the precipitable-water depths at the inflow and outflow ends, respectively. The quantity within parentheses is frequently referred to as W_E, the "effective precipitable water." For each flow model chosen, it represents the amount of rain falling from each unit column processed in the model. In the practical use of the storage equation, the precipitable water is estimated from the surface dew points, and the Δp's are assumed according to the flow model being studied; Y is a function of wind direction and drainage-basin dimensions, and V_1 is estimated from either surface or pilot-balloon observations. For estimation of the maximum possible precipitation that can fall from a given model over a project basin, envelopments of wind-speed and dew-point statistics, for each wind direction, are used in the equation.

OROGRAPHIC RAINFALL

Empirical Approaches. Much of the rain which falls in mountainous regions is due directly to the lifting of moist air up the mountain slopes. For purposes of the present discussion, this percentage of the total rainfall in a storm is defined as the *orographic component*, while the remainder is called the *nonorographic component.*

The orographic component is amenable to fairly simple theoretical derivation since the slope and the elevation of the ground surface—the factors which control the lift of the air per unit wind speed—are fixed and known. For certain noncomplex mountain systems, it is possible to calculate the rainfall with considerable success under the principal assumptions that (1) all of the rain is orographically produced, and (2) the upper-air flow pattern is deducible from surface observations. As the topography becomes more complex, however, not only does the problem become harder to handle mechanically, but also the vertical and horizontal distributions of winds become more difficult to estimate with assurance. Present deficiency in theoretical knowledge of the effects of topography upon the space distribution of wind leads to the conclusion that an empirical approach must serve for many problems involving orographic rainfall until adequate theory is developed. Such an approach may be based upon a rational selection of topographic parameters, for example, slope, exposure, and elevation. Use of multiple-correlation techniques may then produce a regression equation in which rainfall is expressed as a function of the topographic parameters. The equation would also contain a residual term, partly resulting from incomplete selection of parameters and partly giving the nonorographic component of rainfall.

The approach described above is not adaptable to computation of single-storm rainfall unless the meteorological parameters of wind and humidity are also considered. It is well suited, on the other hand, for derivation of long-duration rainfalls over orographic regions of meteorological homogeneity. The method has been used by Russler and Spreen [17] to estimate average seasonal precipitation at points between the relatively sparse network of recording gages in the Colorado River drainage area. Excellent correlations were obtained in this study, in which a graphical multiple-correlation technique was employed.

In a study of rainfall potentialities in the Willamette River Basin, Oregon, Platzman [16] employed an essentially empirical method. Lack of suitable data, in this case, precluded a theoretical investigation. Platzman reasoned that the speed and moisture content of the air passing over the basin from the Pacific were the dominant meteorological factors and, further, that these parameters were representable by observations made at a surface weather-reporting station. He made statistical studies which led to selection of an index station, and from available DDA data from five of the greatest Willamette Basin storms of record, he determined the coefficients in the empirical equation

$$R = (a + bT_d)(c + dV^2), \tag{8}$$

where T_d was dew point and V was wind speed at the index station. Envelopes of the rainy-season wind and dew-point observations at the index station provided values to be placed in equation (8) for estimation of the maximum possible rainfall.

Where theory is inadequate, as in problems associated with highly complex topographic regions, empiricism probably offers the only approach which will produce the required answers. This approach is usually based upon rain which fell in important storms of record and assumes that the model of the average large storm is the same as that of the maximum possible one.

Theoretical Barrier Models. For special, simple topographic regions it is possible to derive expressions based largely upon physical reasoning and yielding computed values which agree well with observed ones. When the outstanding feature of the basin is the existence of one or two mountain barriers which intercept rain-bearing winds, the storage equation (7) may be adapted to the problem. It is usually necessary to make assumptions regarding the vertical distributions of temperature and humidity, since radiosonde observations are inevitably missing in the great storms of record which offer the only checks on the suitability of models to be chosen. Although the assumption of a saturated, pseudo-

adiabatic atmosphere, represented by carefully selected surface index stations, probably does not cause important errors, investigations into the actual structures in major storms may result in significant refinements in the rainfall equations.

The assumption probably open to the most serious question concerns the space-time distributions of wind in the major storms of record and in the maximum possible storm. Observations of upper winds made during periods of both fair and rainy weather are required from stations located near the tops of mountain ranges. Theoretical studies of the structure of air flowing over mountain barriers should continue, special effort being directed toward the theoretical determination of wind structures over barriers during conditions prevalent in major storms. The effects of variations in the geometry of a basin on the wind flow also deserve much study. In other words, the validity of the common assumption of two-dimensional flow should be determined.

The effects of a mountain barrier upon the wind extend not only vertically but also horizontally away from the barrier. In the case of very simple obstacles to wind flow, aerodynamic theory can accurately predict distortions of straight flow. In meteorology, however, the obstacles are irregular mountain peaks, the undisturbed flow pattern is in itself complex, and the fluid is characterized by stratifications of density and humidity. Present meteorological theory is not adequate for the determination of distortions in the wind pattern both upwind and downwind from a barrier except in the qualitative sense that the air begins to rise, on the average, before the barrier is reached, and descends for some time after it has been passed. It is evident that some of the orographic rain must be produced, and must fall, at some distance upwind from the barrier. If a considerable body of rainfall data is available, the "upwind effect" can be treated empirically. Otherwise, assumptions must be made as to the structure of the wind system, whereupon theoretical computations can be made. The upwind-effect problem is only part of the general problem pertaining to flow of air over barriers upon which both statistical and theoretical investigations should be continued. With appropriate modifications of the Reynolds number, laboratory studies with models patterned after typical orographic regions could also lead to valuable results.

Spillover. Since vertical motions which produce rain are inevitably associated with horizontal motions, it is evident that rain formed at a point in the atmosphere will, in general, reach the ground at a spot not vertically below the point of formation. In the case of orographic rainfall, where the wind systems which carry the raindrops along are usually fixed with respect to topographic features, the effect is called *spillover*. The importance of spillover relative to other types of orographically produced rainfall is illustrated in Fig. 3, which contains profiles of rainfall depths in three major storms across the San Joaquin Valley [29]. Theoretical computations of spillover are based upon the following principal assumptions. For each basin two critical raindrop paths can be defined, one at the inflow and the other at the outflow end of the basin, such that all rain formed between them will fall within the basin. Rain formed outside of the space between the two paths will fall outside of the basin. The paths can be based on the assumption of an average raindrop terminal velocity corresponding to that observed in heavy rains, and the horizontal speed of each drop can be assumed to be that of its environment. As with the total orographically produced rainfall, the reliability of the spillover

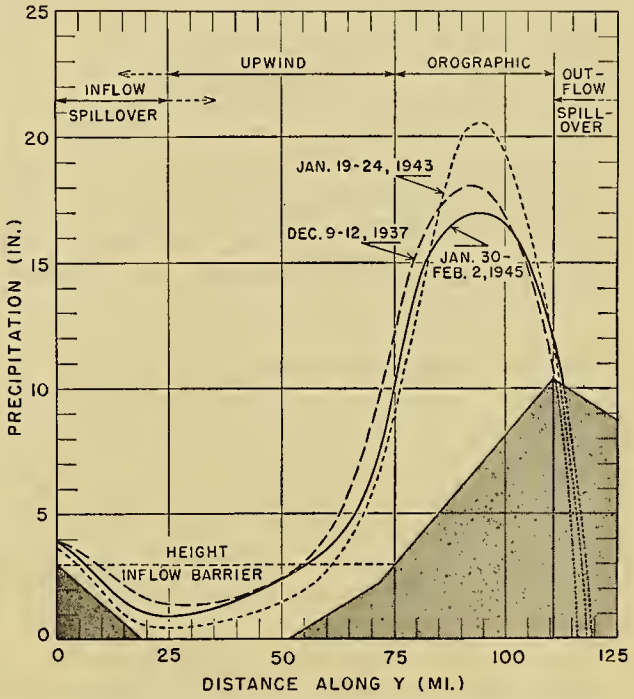


Fig. 3.—Storm profiles, San Joaquin River Basin, California

computations depends upon the accuracy with which the upper-air flow patterns can be reproduced. The importance of spillover increases as the drainage area decreases; for the average small basin of about 100 square miles, spillover may produce a component as high as about 10 per cent of the total rain caused by orography. It is probable, therefore, that errors in the assumptions produce computational inaccuracies of only a few per cent. Spillover can become highly significant for very small basins, however, hence research on this problem is desirable. Refinements in the theory can undoubtedly be attained through application of physical and statistical-mechanical approaches and by collection of data relative to raindrop-size spectra and fall velocities in rainstorms of large magnitudes.

NONOROGRAPHIC RAINFALL

Storm Transposition. In regions of rugged terrain, computations of rainfall are facilitated by an accurate knowledge of the dimensions of the mountain barriers which play such a dominant role in air flow of the lower atmosphere. On the other hand, tests of rainfall formulas for a particular basin may nearly always be carried out through use of precipitation data collected within that

basin only, since the production of rain over other basins is largely controlled by different orographic features. For nonorographic regions as, for instance, the Great Plains of the United States, the fixed controls are absent. Distribution of vertical velocities is, instead, a function of storm type, but the storm type is probably only remotely a function of geographical location within the limits of, say, a half-dozen degrees of latitude or longitude. In nonorographic regions, consequently, while fixed barrier heights are not at hand for use in computation, present meteorological knowledge permits use of data from storms that have occurred not only within the confines of a project drainage basin but also within a sizable area surrounding the basin.

The first estimates of maximum possible precipitation over a drainage basin utilized the highest depths of record within that basin; if no large storms had happened to occur within the period of meteorological history of the basin, use of factors of safety was often adopted. A degree of refinement was then added when the rainfall records of adjoining basins were employed. With the introduction of the process of "transposition" of storm-rainfall values from one region to another came the need for meteorological studies of the major storms of record. Decisions—largely subjective, and based upon the weather-map-analysis experience of trained meteorologists—had to be made as to the area of transposability of each of the great historical storms. In conjunction with other procedures, storm transposition is still in wide use. It will undoubtedly continue to be so until accurate and reliable theoretical rainfall equations can be developed for general usage in nonorographic regions. Use of this procedure obviously requires DDA processing of large numbers of recorded storms.

Moisture Adjustment. The equation of continuity of moisture indicates that rainfall depths increase with water-vapor content of the air, other factors remaining equal. Thus, it can be reasoned that a given storm of record would have produced greater rainfall values had its water-vapor charge been greater, or, in terms of the end product, were a "moisture adjustment" to be applied to the DDA values. Wide usage is made of such a procedure, in combination with that of storm transposition, for regions in which theoretical rainfall equations have not been developed.

Since nearly all of the great historical storms have occurred without available samplings of upper-air humidities, the use of surface indices has had to be adopted. For reasons already pointed out, the surface dew point—reduced to 1000 mb—is usually chosen as the index of the moisture charge of the air in which the rain is formed. In practice, for each storm of record, the dew point is selected at a location, nearest the rainfall center, where tropical air is to be found at the surface. The method of selecting these values is described in detail in a report [30] which also lists the representative storm dew points of about 500 major storms in the United States.

The so-called *maximum possible dew points*, to which the representative storm dew points are adjusted, are derived statistically from the records of first-order Weather Bureau Stations. A close envelopment of the highest values—reduced to the 1000-mb surface—recorded at these stations yields, for each month, a map of maximum possible dew-point isotherms. A detailed description of the process is given in a recent hydrometeorological report [28]. Both the maximum possible and the representative storm dew points are 12-hr values; actually, they are the highest minimum values occurring through any 12-hr duration of the period of record of a station or during the period of a storm, respectively.

Through storm transposition the DDA values of all transposable storms are applied to a project basin and each storm dew point is adjusted to the maximum possible for the season and new location. Envelopment of the transposed and adjusted values of all the storms yields a preliminary estimate of the maximum possible precipitation for the project basin.

Various methods of performing the moisture adjustment have been developed. Of these, only the three which seem to be closest to physical reality will be outlined here. The first, called the *cyclonic-adjustment method*, has frequently been applied to large-area storms. Here, the adjustment is simply multiplication of the rainfall values by the ratio of W at maximum possible dew point to W at the representative-storm dew point, under the assumption of a saturated, pseudo-adiabatic atmosphere. Precipitable-water depths for a layer of air between 1000 and 300 mb are used in the calculations.

The second, the *thunderstorm-adjustment method*, differs from the cyclonic-adjustment method in that the top of the layer for which precipitable-water values are used varies from 100 to 300 mb when the 1000-mb dew point varies from 78F to 50F, respectively. Choice of the range was based upon available data relating to the variation of cumulonimbus cloud tops with surface temperatures and dew points. This method, discussed at length in a hydrometeorological report [28] on generalized estimates of maximum possible precipitation, is intended for use in adjusting storms in which rainfall was mostly from cells of convective action.

The third, or *q-adjustment method*, is intended to apply to the moisture adjustment of storms in which only the magnitude of the moisture charge is to be varied, and all other factors, such as moisture stratification, spatial dimensions of the storm cells, and magnitude and distribution of wind, are to remain unchanged. As described by Fletcher [3], it is possible to factor a mean value of specific humidity from the equation of moisture continuity, and to introduce a factor of proportionality between the mean value and the value of specific humidity corresponding to the surface dew point representative of the storm's moisture charge. For purposes of adjusting storms for moisture only, with all other characteristics to remain unchanged, use of ratios of surface specific humidity is more rigorously correct than use of precipitable-water ratios. However, it has not been established whether the concept that moisture charge can vary without any appreciable change in the rest of a storm's characteristics is physi-

cally tenable. This concept requires considerable theoretical investigation.

Adjustments for Other Factors. In order to reach a project drainage area where rain is to fall, air often must first be lifted over a rise in the ground surface. As in the case of the single-barrier, orographic-rain problem, moisture is removed from the air during its passage over the higher ground. Thus, in treating situations where either a mountain barrier or a long, gentle rise in ground is in the way of air flow toward a basin, it is assumed that a depth of precipitable water contained in a saturated, pseudoadiabatic column of height equal to that of the barrier is removed from the total quantity of moisture initially available for precipitation. The further assumption is made that the barrier—although it may be effective in producing more frequent storms—does not act to alter the wind structure of a transposed storm. It is evident that use cannot be made of the barrier adjustment when a project basin is so close to the downwind side of a barrier that spillover is effective in increasing the precipitation within the basin. The approximate reduction effect of a barrier is one per cent for each 100 ft of barrier height. Statistical tests of the propriety of the barrier adjustment indicate that the adjustment is of the right order of magnitude.

Wind adjustments have been attempted, but the results lack the consistency which marks the moisture and barrier adjustments. Two principal difficulties arise in this connection. First, because minor topographic effects have such a profound influence on the surface wind, it has not been possible to determine reliable wind indices for either the strengths of the currents in storms or the maximum strengths which can be expected in a given locality. Second, it is likely that changes in index-station wind would produce fundamental changes in the nature of a storm by causing important alterations in the space and time distributions of winds.

McCormick [13] has recently shown that there is a significant downward trend of peak rainfall depths as latitude increases. The effect increases as the area over which the depths are averaged increases; for point-rainfall it is so small as not to be statistically significant, but for depths over an area of 2000 square miles the reduction is in the neighborhood of 2 per cent per degree of latitude between latitudes 30°N and 60°N. Variation with rainfall duration was less significant, although there was a tendency for the importance to increase with duration. It is possible that there is a simple physical explanation for the latitude effect. For instance, if the Coriolis acceleration is dominant in the balance of forces in maximum storms, then, since wind appears to the first power in the equation of continuity, rainfall depths should vary inversely with the sine of the latitude.

McCormick also found the quite unexpected, yet statistically significant, fact that the maximum observed depths of rainfall in the United States decrease sharply south of latitude 30°. It is possible that existence of the northern Gulf of Mexico coast—also at latitude 30°—is responsible for this behavior. On the other hand, it is possible that the 30°-peak of rainfall is associated with extreme southward displacements of the meandering jet stream of middle latitudes. Rainfall variations with latitude appear to be important. Thus, with northward or southward transpositions of storms, a latitude adjustment may be in order. The proper nature of the adjustment, however, must await further physical and statistical research.

When allowance is made for latitude and atmospheric moisture content of the major storms of record in the United States, a map of the peak values discloses several meridional maxima, the chief of which extends north-south approximately along the 98° meridian. It may be more than coincidence that the western extremity of the Gulf of Mexico is at about the same meridian. There also may be a certain critical wave length—eastward from the Rockies—of the zonal west winds which, if attained, can be associated with storms which produce greater rainfalls than do storms associated with other wave lengths. The problem of longitudinal variation of maximum rainfall deserves critical attention in the light of its possible use in connection with storm transposition.

Statistical treatments of many other variables should be carried through to determine the existence and probable nature of consistent variations. Hand in hand with the statistics, however, theoretical research should attempt to place the variations on sound physical bases and to determine the degree to which one type of adjustment is dependent upon another. The ultimate objective of the adjustment theory is the expression of rainfall as the product of a number of independent factors, each a function of a particular (measurable) variable, such that all storms can be reduced to a common denominator.

Isobar-Isotherm Relationships. For storms in which the rain-producing air enters at one side of the rainfall area and leaves at the other, it follows that the rainfall depths will be greater the more the inflow winds are confined to low levels and the outflow winds to high levels. Under the assumption that the geostrophic wind-shear equation holds approximately, the wind-stratification system of a heavy rainstorm is such that air entering the system does so with higher temperatures to the left of the current axis and lower temperatures to the right and leaves with the reverse distribution of temperature. A different sort of rainfall-favoring pattern, in which mass convergence takes place at low levels and mass divergence aloft, is characterized, on the inflow side, by anticyclonically curved isobars at low levels and cyclonically curved isobars aloft and, on the outflow side, by anticyclonically curved isobars aloft and cyclonically curved isobars below.

The patterns described above may be combined to form an isobar-isotherm configuration which is theoretically an important rain producer and which has been observed on synoptic maps in varying degrees. For example, combination of a warm, Λ-shaped trough in Texas and a cold cyclone in more northern states meets most of the requirements. With such a configuration, the heaviest rain should fall to the east of the line

joining the warm trough and cold low. It is of interest to note that this example is a storm type which can produce large rainfall depths for long durations. Existence of the warm continental surface (late spring through fall) south of Texas tends to retard movement of a warm trough; in addition, deep cold lows farther to the north are frequently observed to move very slowly. Synoptic studies of variations in this and other types of major storms will without question throw much light on the relationships between rainfall and other measurable variables.

Statistical Relationships. The problem of objective forecasting of rainfall has been attacked—sometimes quantitatively and sometimes qualitatively—by many writers. In the main, the approaches have been statistical, although the choice of independent variables has usually been made on physical grounds. Jorgensen [9] has related upper-air flow patterns to rainfall in California. By means of "composite maps," Solot [22] has developed a method of preparing monthly precipitation forecasts for the Hawaiian Islands. Brier [1] and Penn [15] have made use of upper-air flow patterns in developing objective methods of quantitative precipitation forecasting. Through use of mean maps, Klein [10] has related five-day precipitation amounts to characteristics observed at the 700-mb surface. The statistical approaches, of which the above are a selection, yield results which show appreciable amounts of skill. Especially for the task of quantitative precipitation forecasting, objective, statistical methods display much promise.

As far as hydrometeorological needs are concerned, approaches such as those listed in the previous paragraph constitute only the first step. They tend to deal with average rather than enveloping values of the variables. Forecasts are based on relations derived from rainfall depths of lower orders of magnitude than those with which the hydrometeorologist is concerned. Furthermore, the approaches make no provision for extrapolation of observed to maximum possible values. They have, however, disclosed variables which are important as regards rainfall formation and, consequently, suggest certain directions for future hydrometeorological research.

Enveloping Depth-Duration-Area Relationships. It has recently been shown by Fletcher [4] that maximum observed point-rainfall depths tend to vary with a power of duration so near to 0.50 that the assumption that depth varies with the square root of duration produces a very close envelopment of all recorded values of point rainfall. It was found, in addition to the square root of duration variation, that maximum observed areal rainfall depths vary hyperbolically with area. The equation

$$R = \sqrt{D}\left(0.5 + \frac{266}{19.2 + \sqrt{A}}\right) \qquad (9)$$

was found to be a close envelope of all observed rainfall depths for durations ranging from one minute to one year, and for areas between a point and 200,000 square miles. In the equation, R is depth of rainfall in inches, D is duration in hours, and A is area in square miles.

Equation (9) was derived empirically, and it is entirely possible that future rainfall records may bring about revisions in the constants. On the other hand, theoretical investigations now under way suggest that the form of the equation is correct for maximum possible precipitation.

SPECIAL HYDROMETEOROLOGICAL PROBLEMS

Rainfall Probabilities. Certain design problems require estimates not of the maximum possible precipitation but rather of the greatest precipitation to be expected within a selected number of years. Economic factors might dictate, for example, that the design criteria for a certain structure should be based upon the greatest 6-hr rainfall to be expected once in ten years. For some needs point-rainfall estimates suffice, for others, estimates of average depths over specified areas are required.

From the Weather Bureau's tabulations of excessive rainfall occurrences, Yarnell [34] has published charts of rainfall expectancies in periods of from 2 to 100 years. Such charts can answer many questions as to probabilities of heavy rains. They are confined, however, to point-rainfall depths for durations from 5 min to only 24 hr, and are based upon data which have some variation as to criteria of selection. There is great practical need for an extension of Yarnell's work to longer durations and to averaging of depths over areas.

Other writers have made studies of frequencies of point-rainfall depths and, occasionally, of averages over various sizes of area. When the highest values of rainfall are considered, however, the difficulty arises that such values occur only once in the period of record of the observing station. Thus, the reliability of a "once in a hundred years" value, and similar values, is open to question. Development in statistical theory will undoubtedly assist as regards the probability aspects of high rainfall rates. It is obvious, nevertheless, that accumulation of data through the years is necessary for the development of the probabilities of extremely high rates of rainfall.

Space and Time Distributions of Rainfall. Estimates of maximum possible DDA values do not meet all of the needs of the hydrologist, especially for larger drainage basins. It is also necessary that estimates be made of the isohyetal patterns and their variations with time. As far as river stages at the lower end of a basin are concerned, a heavy burst of rainfall occurring upstream and followed by another downstream is a more significant sequence than one in which the order is reversed.

Heavy storms of record exhibit the characteristic that the highest ratios of observed to maximum recorded DDA values for the United States tend to cluster about only one combination of area and duration, that is, each storm is important only within a restricted range of the durations and areas over which rain actually fell in the storm. Maximum recorded values for the United States are therefore determined by a series of "controlling" storms. Since the greatest storms of record all

possess this characteristic, the assumption has come to be made that all of the maximum possible DDA values will not necessarily fall in one great storm but, rather, that each of several great storms will yield a limited number of maximum DDA values. Until meteorological theory can demonstrate otherwise, such an assumption —based upon data collected for maximum storms of record—appears to be logical.

In practice, the space-time rainfall distributions observed in the greatest storms transposable to a given basin are used to estimate the distribution of maximum possible precipitation for that basin. If major-storm experience for the basin is large, a good selection of distributions becomes available. Regardless of the number of such distributions which have been observed, however, it is apparent that the most critical distribution possible may not have occurred. Accordingly, the dependability of the method necessarily hinges upon the number of data which has been collected during a relatively short number of years. While the method is objective and reasonable within limitations, investigations should be conducted to develop other methods based upon assumptions physically more acceptable.

Studies of intense, short-duration and small-area rainfall such as occurs within a severe thunderstorm have disclosed a very wide variety of isohyetal patterns. For working purposes it is therefore assumed that there is no limit to the kinds of patterns which can result from such localized storms (very roughly, for areas up to one hundred or two hundred square miles and durations up to one or two days). It is possible, nevertheless, that detailed analyses of large numbers of thunderstorm-type isohyetal patterns may reveal certain wide limits as to space and time distributions of rainfall.

Where production of rain is primarily controlled by orographic features, the problem becomes simpler. In an analysis of rainfall over a small drainage basin in Venezuela, for example, Fletcher [5] found a high correlation between patterns of major storms and between the patterns of a major storm and of the mean annual precipitation. Here, meteorological reasoning strongly suggests that the most effective rain-bearing winds come only from one direction, thus that the maximum possible storm would closely resemble the mean annual precipitation as far as the pattern in space is concerned. For other orographic regions, therefore, the analysis in Venezuela suggests that design patterns may be determined for each wind direction.

The assumption is sometimes made that the total rainfall in a storm occurring in a mountainous region may be separated into two mutually independent parts, the orographic and nonorographic components. The orographic rain is assumed to result only from orographic lifting; the nonorographic component is assumed to be that which is due only to the passing meteorological storm. If the occurrences of nonorographic rain are random, the mean seasonal precipitation pattern can be assumed to be produced purely by the topographic configurations of the region. What is known as the *isopercental method* of storm transposition is based on these assumptions. For a major recorded storm, the ratio (expressed in per cent) of the observed precipitation at each station to the corresponding mean seasonal value is plotted on a map. The resulting pattern of isopercentals is transposed from the region of storm occurrence to the project basin, and the percentages, applied to the mean seasonal values of the basin, produce the transposed isohyetal pattern and the magnitude of each isohyet. There is considerable logic to this method. However, there are certain objections which indicate that research should be conducted to estimate the degree to which it may be in error. For instance, there are probably significant inaccuracies in the assumption that the orographic and nonorographic components are mutually independent. It is conceivable that many storms occur in mountainous regions simply because the mountains are there; if the ground had been level, the synoptic situation might have existed as a nonproducer of rain. Again, the mean seasonal pattern may deviate significantly from a pattern derived from only the major rainstorms of record. In the many ordinary rainstorms which are so important in forming the mean seasonal pattern, variation of the condensation level alone could produce a pattern appreciably different from that produced by storms with a uniform condensation level. Furthermore, effects of orography upon rain-bearing winds from various directions can produce profound changes in the isohyetal pattern, such that the orographic-component pattern with one wind direction might only remotely resemble that with another.

Recurrence Interval. In some drainage basins of large area, it is a matter of days for runoff from the upper reaches to arrive at proposed construction sites. It becomes important, for a basin of this sort, to prepare an estimate of the minimum number of rainless days that can occur between two storms of either maximum possible or near maximum magnitudes. Fundamentally, this is the same kind of problem brought up in the previous paragraphs, since it is concerned with space-time distributions of rainfall. The customary DDA analysis of rainfall on the basis of individual storms requires, however, adoption of somewhat different approaches.

The technique usually followed relies upon precedent established in the approximately half a century of United States weather-map analysis. The synoptic type of the storm yielding the critical rainfall depths is first determined. The weather maps of record are then searched for recurrences of storms of the critical type, or types, and the minimum duration between the occurrences is chosen as the recurrence interval.

Application of the form of equation (9) to durations of several days can provide answers for recurrence-interval problems, although it must first be established that this equation is applicable in specific regions. The relation has been developed, basically, from highest available depths of record from storms occurring throughout the world.

Seasonal Distribution of Rainfall. Design and operational problems often require estimates of the way in which maximum possible precipitation varies with

season. Preliminary estimates of the variations may be determined by the processes of storm transposition and adjustment, month by month. Unfortunately, such an approach suffers from a decimation of the number of storms available for processing. Other difficulties with storm samplings arise from the fact that storms are most frequently chosen for DDA analysis because of unusually large rainfall depths which have been observed, and that some seasons are characterized by maximum observed amounts much lower than those of other seasons. The method will become more useful as more storms are analyzed for their DDA values, and as the sampling becomes more nearly equal, season by season.

The estimates will become more reliable when physical bases are established for seasonal variations in the factors appearing in theoretical rainfall equations. It must be emphasized, however, that extreme rainfall depths are being considered, and thus that seasonal variations of extreme rather than average values of meteorological parameters must be developed. Hydrometeorology necessarily deals with envelopes rather than with means.

Snow Melt. When estimates of maximum possible precipitation are made for drainage basins which are located in the more northerly latitudes of the United States, or parts of which lie at high elevations, consideration must be given to the contribution of snow melt to the total runoff. In the spring of 1948, the Columbia River Basin experienced an excellent example of the conversion of a deep blanket of snow, in combination with heavy rains, into a disastrous flood [25].

Basically, determination of potential runoff is amenable to analysis by means of thermodynamic and turbulent-exchange theories. In a detailed treatment of the problem, Light [12] has defined the "effective snow melt" as a combination of (1) melt due to the direct heat exchange which exists when the air is warmer than the snow mantle, (2) melt resulting from the latent heat released when atmospheric water vapor is condensed on the snow surface, and (3) the condensation of such atmospheric water vapor. He gives the simple relationship for the effective snow melt D in centimeters depth per second:

$$D = (Q + 600F)/80 + F = (Q + 680F)/80, \quad (10)$$

where Q is the heat transfer by convection in calories per square centimeter per second and F is the water transfer in cubic centimeters per second. Here, the latent heat of condensation amounts to 600 cal per cc of water deposited, and the latent heat of fusion is 80 cal per cc. It can be seen that the moisture condensed on the snow surface melts 7.5 times its own weight of snow.

The quantities Q and F, in the form given by Sverdrup [23], are expressible in terms of surface wind velocity, temperature difference between air and snow surface, air density and pressure, the surface roughness parameter, and the elevations of the anemometer and hygrothermograph above the snow surface. These expressions may be placed in equation (10) for the derivation of Light's snow-melt formula:

$$D = U_m[0.00184(T - 32)10^{-0.0156h} + 0.00578(e - 6.11)], \quad (11)$$

where D is the effective snow melt in inches per six hours, U_m is the average wind speed in miles per hour, T is air temperature in degrees Fahrenheit, e is vapor pressure in millibars, and h is station elevation in thousands of feet above sea level. The formula assumes a snow-surface temperature of 32F, and reference elevations of 50 and 100 ft for the anemometer and hygrothermograph, respectively. The roughness parameter is assumed to be 0.25 cm.

The snow-melt formula can be refined to take air stability and wind shear into account more accurately through the use of meteorological measurements at more than one level. The net melting effect of incoming and outgoing radiation may also be estimated by a method given by Wilson [33]. Furthermore, the relatively unimportant melting effect of rain may be estimated when assumptions are made as to raindrop temperatures.

The theoretical formula represents the rate of melt of a smooth snow surface over which the wind blows without appreciable retardation due to such obstructions as trees. Project basins usually have sizable tree-covered areas, however, and sometimes are located in mountainous terrain. To account for errors in the assumption of surface roughness, an empirical factor of proportionality—the *basin factor*—is introduced into the formula. Comparisons of observed with computed values of melt are the bases for determination of the basin factor. The latter is normally found to be less than unity since, in most basins, forests exist to such an extent that the turbulent heat exchange occurring at a well-exposed index anemometer station is greater than the average heat exchange over the basin as a whole.

When the basin factor has been determined, synoptic meteorological studies lead to conclusions regarding magnitudes of snowstorms antecedent to the maximum possible rainstorms, and the antecedent temperature regime and its effect upon the antecedent snow cover. The space and time distributions of wind and dew point immediately preceding and during the maximum rainstorm then form the bases for the determination of the wind and temperature parameters appearing in the snow-melt formula. It must be established that a certain critical depth of snow can exist at the onset of the rainstorm. This depth is such that the winds and temperatures of the rain period will melt all of the snow cover. A snow cover which is greater than that which can be melted during the rain period will store some of the snow melt and act to reduce the runoff. Thus it is evident that the critical snow depth is not necessarily the maximum possible.

The snow-melt problem is being systematically attacked by the collection and processing of data at three

field laboratories,[2] operated cooperatively by the Weather Bureau and the Corps of Engineers, in the mountainous regions of the western United States. Studies are currently under way on such features as the effects upon snow melt of varying degrees of forest cover, of elevation and character of the ground surface, and of meteorological factors. Study is also being made of the structure, density, water-storage capacity, and other variable characteristics of snow covers. Data available through the work done at these laboratories will undoubtedly lead to a great increase in knowledge of the processes involved in the melting of snow and, indirectly, of depths of snow critical from the standpoint of runoff.

Wind Tides and Waves in Reservoirs. The hydrometeorologist, because he is primarily a meteorologist, is called upon at times to investigate meteorological problems only indirectly connected with calculations of rates of precipitation. The height of a flood-control structure depends upon the maximum expected water level behind it; such a level will be caused primarily by maximum rainfall occurring with the most critical space-time distribution. If winds are blowing, however, the surface of the water cannot remain horizontal. Instead, a tilting of the reservoir surface results from the stress of the wind blowing across it. On the tilted surface, furthermore, waves develop which may have certain destructive effects.

To computations of maximum possible precipitation, then, must sometimes be added estimates of maximum wind-duration values for critical directions. Information of this sort is usually obtained through a study of long series of synoptic weather charts and through statistical analyses of post-rainstorm wind records of weather stations.

When the surface-wind system of a storm is known in considerable detail, as was the case with the Florida hurricane of August 1949 [31], the process of storm transposition can often be used. The process requires statistical studies of relations between winds at stations whose observations determine the surface-wind system, and winds over the project reservoir. The consistency in such relationships has generally been found to be quite significant. However, the assumption that a windstorm of record could occur without structural change over a project reservoir limits the kinds of transpositions to be made. Theoretical investigations are needed to improve present knowledge of modifications in a wind system (*e.g.*, of a hurricane) due to changes in place of occurrence, in direction of storm travel, and in character of the underlying surface.

A "design" windstorm derived, for instance, by the transposition method, must be converted into terms of water levels and wave heights for the project reservoir. The field of oceanography at present does not provide much information for the solution of such a problem. Relationships between surface winds, wind tides and waves in shallow, inland reservoirs of limited extent, and the effects of a steadily changing wind system, must be known before many important hydrologic design problems can be successfully attacked.

Water Resources. Recent years have seen an increasing need for evaluation of water supply, especially in the western states. To appraise this all-important resource requires the combined efforts of the hydrometeorologist, climatologist, and hydrologist. The problem is multifold, in that it embraces studies of rainfall, evaporation, transpiration, snow melt, runoff, and surface and underground storage. The hydrometeorologist is directly concerned with the association of wind, humidity, and other meteorological elements with factors involved in the water budget of an area.

The process of evaporation exercises much control over the amount of water available for use. However, conversion of evaporation-pan measurements to actual ground- and reservoir-surface evaporation has not yet met with a satisfactory solution. Pan evaporation has been empirically related to such meteorological parameters as wind and humidity, and there are theoretical equations expressing evaporation as a function of vertical gradients of wind and humidity. A great deal of work, theoretical and statistical, remains to be done, however, before statistics of meteorological factors can be combined to determine, with assurance, long-term evaporation losses.

THE FUTURE OF HYDROMETEOROLOGY

Systematic Organization and Use of Existing Data. There exists in the weather archives a wealth of information requiring organization for hydrometeorological appraisal and research. Maximum recorded point-rainfall depths have been extracted from some of the records for some durations. For example, Shands and Ammerman [18] have tabulated maximum values for 207 first-order Weather Bureau stations for a selection of durations from five minutes to one day. It would be desirable for many more durations to be considered. A maximum observed depth-duration curve for each station would be especially well received. Records of stations other than first-order Weather Bureau stations, as well as those of foreign stations, should be similarly processed.

Maximum observed values should not be restricted to those falling between fixed clock hours, days, or months. According to Jennings [8], the world's record 24-hr value was recorded at Baguio, Philippine Islands, on July 14–15, 1911. An accumulation of about 46 in. was observed from noon of the 14th to noon of the next day. It is almost a certainty that a still greater depth could be selected from some other 24-hr interval during the storm period. Unfortunately, records needed for determination of such a greater depth were lost during World War II.

The selection of storms hydrometeorologically analyzed for DDA values by the Corps of Engineers and the Weather Bureau should be greatly augmented. For both theoretical and statistical research, there is

2. The Central Sierra Snow Laboratory at Soda Springs, California; the Upper Columbia Snow Laboratory near Marias Pass, Montana; and the Willamette Snow Laboratory, Oregon.

need for such analyses of storms covering a wide variety of magnitudes, durations, areas, storm types, seasons, and locations. The DDA values, were they placed on punch cards, could be speedily correlated with many factors which have so far not been tested as to their significances to areal rainfall intensities.

The use of punch-card techniques would facilitate extension of the work of Yarnell [34] and other writers, as regards rainfall probabilities. Addition of data from large numbers of DDA-analyzed storms would eventually permit establishment of probabilities of occurrence of areal rainfall intensities.

Augmentation of Existing Rainfall-Reporting Networks. Statistical studies as to correlations between point and areal rainfall depths should be undertaken with the objective, among others, of determining the proper density for rain-gage networks. It is likely that existing networks are acceptable for most purposes in some sections of the United States. In mountainous regions the network is still far too meager for determination of effects of topography on rainfall. Over oceans and other large bodies of water—where special instrumentation is required—there are essentially no observations available for research.

As suggested by Smith and Fletcher [20] in 1946, the use of radar for quantitative determination of precipitation intensities between observing stations may have practical possibilities. Later investigations, such as those of Byers [2], tend to substantiate the idea. Radar has already proved its value in qualitative, short-range precipitation forecasting. Development of electronic equipment to measure areal values of precipitation intensities would be of extreme value for both forecasting and planning.

Physical Research. In theoretical hydrometeorology, the greatest need is for knowledge of the behavior of wind in space and time, over orographic and nonorographic terrain, in synoptic situations favorable for peak rainfall intensities. Wind observations are required in great quantity in this connection, and physical reasoning regarding limiting conditions must be developed.

The equations of continuity and hydrostatics have been the principal theoretical relationships used in hydrometeorology. Exploration into the applicability of others, such as the equations of motion and of energy relationships, is needed.

Results of research in the field of atmospheric turbulence can be turned to great practical use in hydrometeorology. Intimately connected with the theory of turbulence are the problems of snow melt, evaporation, behavior of wind over rugged terrain, and the aggregate behavior of raindrops in convective cloud currents. Some of these and other problems are amenable to treatment by the methods of statistical mechanics.

As far as the field of hydrometeorology is concerned, there are two main goals toward which research should be directed. The first is the development of improved theoretical equations relating rainfall depths, averaged over area and through duration, with measurable independent variables. The other is the establishment of physically reliable methods of maximizing the rainfall through maximization of the variables, in combination, to which it is related.

REFERENCES

1. Brier, G. W., "A Study of Quantitative Precipitation Forecasting in the TVA Basin." *U. S. Wea. Bur. Res. Pap.* No. 26 (1946).
2. Byers, H. R., and Collaborators, "The Use of Radar in Determining the Amount of Rain Falling over a Small Area." *Trans. Amer. geophys. Un.*, 29:187–196 (1948).
3. Fletcher, R. D., "Computation of Thunderstorm Rainfall." *Trans. Amer. geophys. Un.*, 29:41–50 (1948).
4. —— "A Relation between Maximum Observed Point and Areal Rainfall Values." *Trans. Amer. geophys. Un.*, 31:344–348 (1950).
5. —— "A Hydrometeorological Analysis of Venezuelan Rainfall." *Bull. Amer. meteor. Soc.*, 30:1–9 (1949).
6. Fulks, J. R., "Rate of Precipitation from Adiabatically Ascending Air." *Mon. Wea. Rev. Wash.*, 63:291–294 (1935).
7. Glasspoole, J., "The Areas Covered by Intense and Widespread Falls of Rain." *Proc. Instn. civ. Engrs.*, 229:137–166 (1930).
8. Jennings, A. H., "World's Greatest Observed Point Rainfalls." *Mon. Wea. Rev. Wash.*, 78:4–5 (1950).
9. Jorgensen, D. L., "An Objective Method of Forecasting Rain in Central California during the Raisin-Drying Season." *Mon. Wea. Rev. Wash.*, 77:31–46 (1949).
10. Klein, W. H., "An Objective Method of Forecasting 5-Day Precipitation for the Tennessee Valley." *U. S. Wea. Bur. Res. Pap.* No. 29 (1949).
11. Kohler, M. A., "On the Use of Double-Mass Analysis for Testing the Consistency of Meteorological Records and for Making Required Adjustments." *Bull. Amer. meteor. Soc.*, 30:188–189 (1949).
12. Light, P., "Analysis of High Rates of Snow Melting." *U. S. Wea. Bur., Hydrometeor. Sect. Tech. Pap.* No. 1 (1941).
13. McCormick, R. A., "Latitudinal Variation of Maximum Observed U. S. Rainfall East of the Rocky Mountains." *Trans. Amer. geophys. Un.*, 30:215–220 (1949).
14. Miller, J. E., "Studies of Large Scale Vertical Motions of the Atmosphere." *Meteor. Pap., N. Y. Univ.*, Vol. 1, No. 1 (1948).
15. Penn, S., "An Objective Method for Forecasting Precipitation Amounts from Winter Coastal Storms for Boston." *Mon. Wea. Rev. Wash.*, 76:149–161 (1948).
16. Platzman, G. W., "Computation of Maximum Rainfall in the Willamette Basin." *Trans. Amer. geophys. Un.*, 29:467–472 (1948).
17. Russler, B. H., and Spreen, W. C., "Topographically Adjusted Normal Isohyetal Maps for Western Colorado." *U. S. Wea. Bur. Tech. Pap.* No. 4 (1947).
18. Shands, A. L., and Ammerman, D., "Maximum Recorded U. S. Point Rainfall." *U. S. Wea. Bur. Tech. Pap.* No. 2 (1947).
19. Showalter, A. K., "Rates of Precipitation from Pseudo-adiabatically Ascending Air." *Mon. Wea. Rev. Wash.*, 72:1 (1944).
20. Smith, E. D., and Fletcher, R. D., "A Summary of the Uses of Radar in Meteorology." *Trans. Amer. geophys. Un.*, 28:713–714 (1947).
21. Solot, S. B., "Computation of Depth of Precipitable Water in a Column of Air." *Mon. Wea. Rev. Wash.*, 67:100–103 (1939).

22. —— "Further Studies in Hawaiian Rainfall." *U. S. Wea. Bur. Res. Pap.* No. 32 (1949).
23. Sverdrup, H. U., "The Eddy Conductivity of the Air over a Smooth Snow Field." *Geofys. Publ.*, Vol. 11, No. 7 (1934).
24. Thiessen, A. H., "Precipitation Averages for Large Areas." *Mon. Wea. Rev. Wash.*, 39:1082–1084 (1911).
25. U. S. Geological Survey, Department of Interior, "Floods of May–June 1948 in Columbia River Basin." *Water-Supply Pap.* No. 1080 (1949).
26. U. S. Weather Bureau, *Cooperative Studies Report* No. 9 (1949).
27. —— "Thunderstorm Rainfall." *Hydrometeor. Rep.* No. 5 (1947).
28. —— "Generalized Estimates of Maximum Possible Precipitation over the U. S. East of the 105th Meridian." *Hydrometeor. Rep.* No. 23 (1947).
29. —— "Maximum Possible Precipitation over the San Joaquin Basin, California." *Hydrometeor. Rep.* No. 24 (1947).
30. —— "Representative 12-Hour Dewpoints in Major U. S. Storms East of the Continental Divide." *Hydrometeor. Rep.* No. 25A, 2nd ed. (1949).
31. —— "Analysis of Winds over Lake Okeechobee, Florida, during the Tropical Storm of August 26–27, 1949." *Hydrometeor. Rep.* No. 26 (1951).
32. —— Hydrometeorological Section, "Tables of Precipitable Water and Other Factors for a Saturated, Pseudoadiabatic Atmosphere." *U. S. Wea. Bur. Tech. Pap.* No. 14 (1951).
33. Wilson, W. T., "An Outline of the Thermodynamics of Snow-Melt." *Trans. Amer. geophys. Un.*, 22:182–195 (1941).
34. Yarnell, D. L., "Rainfall Intensity-Frequency Data." *U. S. Dept. Agric. Misc. Publ.* No. 204 (1935).

THE HYDROLOGIC CYCLE AND ITS RELATION TO METEOROLOGY—RIVER FORECASTING

By RAY K. LINSLEY

U. S. Weather Bureau, Washington, D. C.

THE HYDROLOGIC CYCLE AND ITS RELATION TO METEOROLOGY

Hydrology is that branch of physical geography which deals with the waters of the earth exclusive of the oceans. As a central focus for their efforts, hydrologists have adopted what is known as the *hydrologic cycle* (Fig. 1). This concept describes the circulation of water as it evaporates from the oceans and enters the atmosphere, is precipitated to the earth, and ultimately returns to the oceans by surface and underground channels. The interest of meteorology in the atmospheric phase of this cycle is self-evident.

No schematic diagram can do justice to the complexities of the hydrologic cycle. Instead of a simple step-by-step cycle, all phases may occur simultaneously. Large volumes of water may remain in storage at the earth's surface as snow or soil moisture. Some of this stored moisture is evaporated and may be reprecipitated over land. Many writers [4] have attempted to evaluate the quantities of water involved in the meteorological phases of the cycle. From the viewpoint of the hydrologist such evaluations are largely academic, for he is interested in the disposition of the water which reaches a specific area, and its ultimate source is of little concern. There seems little doubt that moisture evaporated from the land is an unimportant source of continental precipitation and, hence, that such activities as land drainage or reservoir construction can have little effect on the precipitation regime of an area.

Because of the importance of water in human life, hydrology in a simple form has been practiced since the dawn of history. Archeological evidence indicates that many successful water-supply and drainage projects were constructed at early dates. Scientific hydrology got its start in the 15th century A.D., when it was demonstrated that precipitation over land areas was adequate to supply the flow of streams. Relatively little further progress was made in the science until the current century, when engineering requirements forced the development of techniques for the design of major water-control projects.

Fields of Hydrology. O. E. Meinzer has suggested the division of hydrology into four fields representing stages in the surface phase of the hydrologic cycle: *potamology*, study of surface streams; *limnology*, study of lakes; *cryology*, study of snow and ice; and *geohydrology*, study of ground water.

Lakes are an important influence on the climate of their immediate vicinity and may contribute substantial amounts of water vapor to the atmosphere. The limnologist is vitally interested in long-term trends in precipitation and evaporative potential, as these factors affect the progressive rise and fall of lake levels. A study of lake-level trends may ultimately prove valuable in interpreting long-period changes in climate.

While the field of cryology is, in the strict sense, limited to snow and ice on the earth's surface, the meteorological conditions during snowfall greatly influence the physical characteristics of the snow which accumulates on the ground. This snow subsequently undergoes a continuous metamorphosis in response to the effects of temperature, humidity, wind, and precipitation. Studies of the influence of meteorological factors on snow may eventually contribute to our knowledge of the changes in air-mass characteristics resulting from passage over extensive snow-covered areas.

Of the four branches of hydrology, geohydrology is probably least closely related to meteorology. The ground-water specialist is interested in precipitation and precipitation trends in the intake areas which feed the major water-bearing formations. Beyond this point, however, the geological features of the ground-water problem are paramount. In turn there appears to be little of interest to the meteorologist in the study of ground water.

Potamology is the largest of the branches of hydrology and to some extent embraces the other three branches. The surface-water hydrologist is interested in lakes as sources of streams or as factors in the characteristics of streams which flow through them. He is greatly interested in snow and ice as an important source of water for stream flow. Finally, the dry-weather flow of most streams is derived largely from ground water and the potamologist must therefore concern himself with the ground-water features of his area. He is more interested in meteorological conditions than are his colleagues, for the flow of surface streams is directly related to the prevailing weather situation. In the material to follow, some of the more important problems in surface-water hydrology in which the meteorologist may be interested, or to which he may ultimately contribute a solution, are discussed briefly.

Precipitation. Since precipitation is the source of all stream flow (and ground water) with the exception of small quantities of water of internal origin, a basic problem in hydrology is the evaluation of the precipitation regime over an area. The hydrologist finds it necessary to deal with areas defined in terms of natural drainage boundaries, that is, river basins. His aim is the quantitative solution of the equation defining the hydrologic balance of the area,

$$P - E \pm \Delta S = R, \qquad (1)$$

where P is average precipitation, E is the water returned to the atmosphere by evapotranspiration, ΔS is the increment of water added to or removed from storage in the basin, and R is the runoff volume leaving the basin via the streams.

To balance equation (1) it must first be possible to determine the true average precipitation over the basin. This immediately involves problems in sampling, unless radar techniques can measure the integrated total over the basin. In level terrain, where the occurrence and distribution of precipitation are more or less random, the planning of an adequate precipitation-gage network depends on an understanding of the structure of storms and particularly of the variations of depth with distance from the storm center. In mountainous regions the effect of topography must also be considered. Spreen [13] has suggested the use of empirical correlations between average annual precipitation and topographic parameters. Maps prepared from such correlations would provide a guide to the proper location of stations for better evaluation of annual or seasonal precipitation. Similar studies of individual storms in rugged terrain are still needed.

Evaporation. The study of evaporation has presented more practical difficulties than most other items in the hydrologic balance. Storage reservoirs are designed to conserve water which would otherwise go to waste during the rainy season and to make it available when needed. There have been numerous instances in which the increased evaporation from the enlarged water surface created by the reservoir has exceeded the volume of usable water gained by storage, thus making the reservoir a liability rather than an asset. Less critical in many respects, but nevertheless preventing the quantitative solution of the water balance, is the unsolved problem of evaporation from land surfaces.

For many years the only attempt to solve these problems at the practical level has been by the use of evaporation pans. Obviously, there is a vast difference between the evaporation from a pan 4 ft in diameter and 10 in. deep and that from a large reservoir. It has been generally accepted, on the basis of experiments by Rohwer [11], that the evaporation from reservoirs averages about 0.7 of that from the U. S. Weather Bureau Class A pan. It is known, however, that there are considerable variations in this coefficient depending on size, depth and exposure of the reservoir, and time of year. Any quantitative application of pan data to soil evaporation is made difficult by the varying evaporation opportunity with variations in soil moisture.

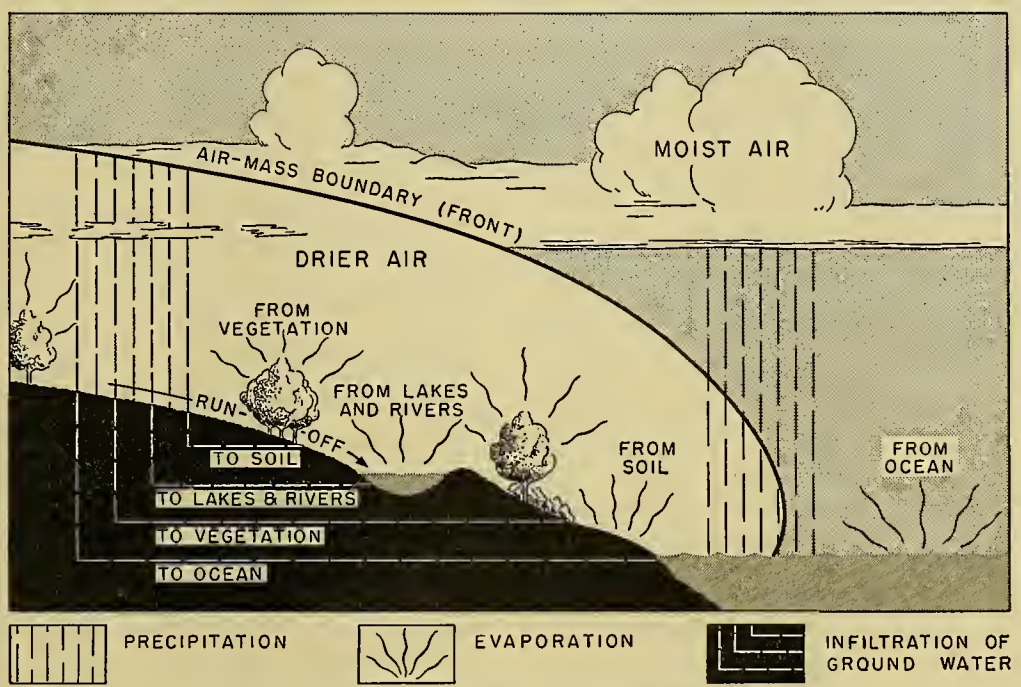


Fig. 1.—Schematic diagram of the hydrologic cycle.

A technique for determining evaporation either by direct measurement of outgoing moisture or by computation from the energy balance [10], turbulent transfer [14, 15], or other theory is urgently needed. However, evaporimeter data cannot be ignored, for they are all that will be available for many years. It has been shown that a good correlation exists between pan evaporation and readily available meteorological data—temperature, dew point, and wind [7, 9]. Through such correlations, available evaporimeter data may serve to guide the extrapolation and application of data collected by other means. A more rational method for adjusting pan data to reservoir conditions may be found if water temperature, wind, and humidity are observed at both the pan and the reservoir site. One difficulty hampering a study of this problem is the fact that almost no completely watertight reservoir

exists—leakage through gates and valves and seepage through the banks accounting for losses of the same order of magnitude as evaporation. In addition, the accurate measurement of all inflow to the larger reservoirs is exceedingly difficult.

In the spring of 1950 the U. S. Navy, Geological Survey, Bureau of Reclamation, and Weather Bureau jointly undertook a study of evaporation at Lake Hefner near Oklahoma City, Oklahoma. This reservoir, selected after a nationwide survey, is considered to be most nearly the ideal experimental site available. Instruments have been installed with a view to testing the turbulent transfer and energy balance concepts and to refine the relationship between evaporation pans and lake evaporation. This study should prove to be one of the most important steps toward the solution of the evaporation problem undertaken in recent years. It is to be noted, however, that almost any relationship developed will involve an empirical coefficient which must be selected largely on the basis of judgment before the method can be applied to any other lake or reservoir.

Snow. The snow pack which accumulates each winter in the mountains of the western United States constitutes a reservoir of such proportions as to dwarf all man-made lakes in this country. With the coming of the spring thaws, this water is delivered to the valleys when it is most urgently needed for irrigation. Occasionally, under the influence of extremely abnormal meteorological conditions, the melt water is released so rapidly that damaging floods result. It is natural therefore that the problem of ice and snow should receive considerable attention from the hydrologist.

The problem of snow melt is not unlike that of evaporation. Melting of snow is the result of heat brought to it from several sources—conduction and convection from air, radiation, condensation of water vapor, rainfall, and the soil. Continued research into the mechanics of turbulent transfer will materially aid studies of snow melt. However, the snow-melt problem involves extensive areas, usually with a wide range of elevation and rugged topography. It is doubtful, therefore, that a theoretical approach will find practical application in most hydrologic work. It is to be hoped that theoretical studies will suggest a more adequate empirical approach to the job.

Design Problems. Another phase of hydrologic work in which meteorology can play an important role is the field of design. Every structure, large or small, designed to control the flow of water, be it a culvert on a secondary road or a major reservoir, must be planned to withstand some maximum flow of water known as the *design flood*. For the smaller structures, economics dictates that the design flood be the probable maximum flood to be expected during the estimated useful life of the structure. For larger structures, where failure may take a large toll of life or cause damage far beyond the value of the structure itself, the design may be based on the maximum flood which can possibly be expected to occur. Cost normally increases with the magnitude of the design flood. Hence, overdesign is costly in terms of wasted material and labor, while underdesign may be equally costly as a result of failure and resulting replacement. Any knowledge which helps to refine estimates of design-flood magnitude is of definite economic value [1].

Generally speaking, the design of smaller structures is based on frequency analysis, and for structures having a useful life of thirty years or less the design criteria may be assumed to be reasonably adequate. For larger structures the situation is not so favorable. Reliable records of stream flow in excess of thirty years are rare. The situation is somewhat better as far as precipitation records are concerned but, even for these data, records in excess of fifty years are scarce. Some writers have proposed the so-called *station-year* technique [3] in which records from a number of stations within a homogeneous area are assumed to be independent and can thus be combined to give a record equivalent in length to the total length of the individual station records. Even if the validity of this approach were beyond question, it fails to satisfy the design problems for intermediate and larger projects whose drainage areas are so large that single-station records cannot be considered representative. A study of the frequency of occurrence of various average depths of rainfall over areas from 100 to 10,000 square miles would be extremely useful. Because of the lack of processed storm data, such a study would involve an immense amount of work. A study of storm morphology with particular reference to the relation between maximum point and average depths would also be of considerable value.

Another problem often encountered involves joint frequency, that is, the question, What is the frequency of rainfalls of various magnitudes with snow cover on the ground? or, What is the probability of a small, intense storm which would overload the drainage works of a leveed area at the same time that a major flood is occurring as the result of general heavy rains several days earlier? [16] Probably these problems cannot be solved by either meteorologists or hydrologists until much longer records exist. However, the considered judgment of the meteorologist can be of material aid to a hydrologist forced to make a decision and adopt a design value, no matter how poor its basis.

The situation reaches an extreme in the case of those major structures whose failure cannot be permitted under any condition. Here it is necessary to adopt the maximum possible flood as a design criterion. Since records are far too short to define such a physical extreme, it is necessary to synthesize it on a theoretical basis. It is therefore logical to turn to the meteorologist for an expression of the magnitude of the maximum possible storm for the project basin. The joint venture deriving from this need—hydrometeorology—is discussed elsewhere in this Compendium.[1]

It would be an easy matter for the hydrometeorologist to protect himself against an underestimate by using

1. Consult "Hydrometeorology in the United States" by R. D. Fletcher, pp. 1033–1047.

a liberal factor of safety. However, such a procedure would result in excessive project costs, and it is necessary to adhere to estimates made on the basis of sound meteorological theory supported by professional judgment. Obviously, the answer to the question, What is the maximum amount of rain which can fall over a given drainage basin in a specified time? requires the entire resources of the science and a considerable body of fact and theory not yet available. Consequently, any fundamental advance into the problems of meteorology will ultimately contribute to the solution of the problems of hydrometeorology.

RIVER FORECASTING

Of all the applications of hydrology, perhaps the one of most interest to the meteorologist is river forecasting. In flood forecasting particularly, meteorology is the outpost which provides a warning of the earliest evidence of possible floods. Many countries have recognized this fact by combining their hydrological and meteorological activities in a single hydrometeorological service. In the United States, where most hydrologic work is assigned to other agencies, the U. S. Weather Bureau bears the responsibility for all types of river forecasts.

The Organization of a River-Forecasting Service. The first essential of a river-forecasting service is a reporting network bringing current river stage and weather data to the forecast office. In terms of instrumentation and types of observations, this network may be much simpler than that used for weather forecasting. River stages, precipitation depths, and occasionally air temperature and the water equivalent of snow are the essential items of data. Because small-scale variations in weather are important in river forecasting, the network of stations must be far denser than is normally required for weather forecasting. The controlling factors for the frequency of reports are the size and the hydrologic characteristics of the river basin. Forecasts at the point of outflow from basins of 20,000 square miles or more can usually be made on the basis of daily reports, while for basins under 100 square miles hourly reports may be inadequate because of the rapid concentration of flow.

In addition to the reporting network, the service must, of course, have forecast offices staffed with an adequate number of trained personnel and equipped with properly developed forecasting relationships. Because of the high peak work load during floods, the staff must be well-organized to accomplish its job in a minimum of time.

Flood Routing. The forecasting operation naturally divides into several categories. In making forecasts along a large river it is necessary to estimate the speed of movement and the changes in shape which the flood wave undergoes as it moves downstream. In terms of theoretical hydraulics, this is a problem in wave motion. However, the equations of wave motion are far too complex for application to natural channels within the time available for forecasting. Hydrologists have therefore turned to a more empirical solution known as *flood routing*.

The simplest flood-routing relation is the *crest-stage relation*, which is a plotting of historical flood crests at one station against the resulting crests at another station downstream. Used in conjunction with a *time-of-travel curve*, which shows the rate of crest travel at various stages, the crest-stage relation is a simple and often effective means of forecasting peak stages. However, considerable flow may be contributed to the flood wave by tributaries entering between the two stations and, unless this flow is always proportional to the upstream inflow (or negligible in comparison to this inflow), large errors may result. Moreover, crest forecasts alone are frequently insufficient. When operation of reservoirs is involved, it is necessary to know the time distribution of all runoff which is to enter the reservoir in order to plan its effective operation. In addition, riverbank interests are almost as concerned with the time at which the river will rise to the critical level at their particular location as they are in the ultimate crest, since this time fixes the interval available to them for evacuation.

Forecasts of the shape of the entire hydrograph may be made by use of *storage routing* based on the storage equation,

$$It - Ot = \Delta S, \tag{2}$$

where I and O are the rates of inflow to and outflow from a section of the river, t is time, and ΔS is the change in volume of water within the reach during time t. If it is assumed that the average of the flows at the beginning and end of the routing period equals the average flow for the period, this equation can be written as

$$\frac{I_1 + I_2}{2}\, t - \frac{O_1 + O_2}{2}\, t = S_2 - S_1, \tag{3}$$

where the subscripts 1 and 2 refer to the beginning and end of the period, respectively. Equation (3) contains two unknowns, O_2 and S_2, but by expressing S as a function of O (and sometimes I) a second equation is available for the solution. One of the more straightforward, analytical solutions of the routing equation is the Muskingum method. This method assumes that the relation between storage and flow in a stream can be written as

$$S = K[xI + (1 - x)O], \tag{4}$$

where K is a storage constant having the dimension of time and x is a dimensionless ratio expressing the relative importance of O and I in determining the storage within the reach.

Introducing equation (4) into equation (3) and solving for O_2, we find

$$O_2 = C_0 I_1 + C_1 O_1 + C_2 I_2, \tag{5}$$

where the several C's are functions of K, x, and t. If values of S are plotted against values of the bracketed term in (4), the best value of x is that which reduces the plotted points most nearly to a straight line, and

K is the slope of this line. More interesting, however, is the possibility of solving (5) by means of least squares in which the C's become regression constants, or by multiple graphical correlation yielding a chart such as Fig. 2. The latter method has the advantage of permitting curvilinear and joint functions without the complexities of the analytical solution.

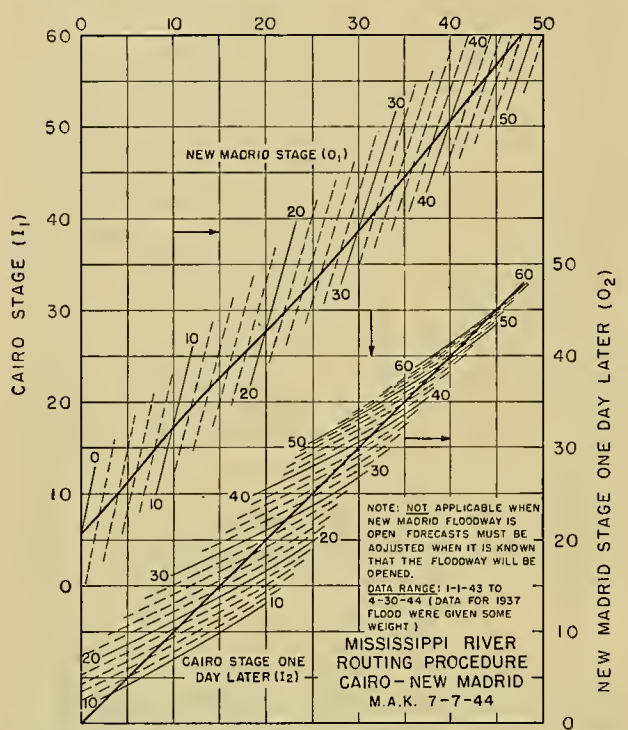


Fig. 2.—Stage routing relation between Cairo, Ill. and New Madrid, Mo. developed by multiple graphical correlation.

Recently an electronic circuit has been devised for the solution of the differential form of the storage equation ($I - O = dS/dt$) by analogy with the flow of electricity in a circuit containing capacitors [6].

Forecasting the Runoff from Rainfall. The forecaster dealing with small headwater basins cannot make use of upstream flows as a basis for forecasts, and he must turn to the ultimate inflow to the basin—precipitation. The first step in headwater forecasting is therefore the determination of the quantity of water which will actually reach the stream. This step is a detailed evaluation of one small portion of the hydrologic cycle. The forecaster must determine the portion of the precipitation which will be held in the basin as interception on vegetation, in puddles on the ground surface, and as subsurface storage (soil-moisture and ground water). The amount of this "basin recharge" in any storm is a function of (1) the moisture conditions in the basin prior to the storm, and (2) the intensity, amount, duration, and distribution of the precipitation. Much has been written on the theoretical considerations which may ultimately lead to a completely rational solution of this problem [7]. For the time being at least, the complexities introduced by superimposing the varying characteristics of a storm with respect to area over a basin which itself has varying soil types, vegetal cover, and antecedent moisture conditions have forced the adoption of empirical techniques.

Many methods have been employed, ranging upward in complexity from a simple correlation between rainfall and runoff, but the procedure which has proven most successful in the experience of the U. S. Weather Bureau is a multiple graphical correlation such as that shown in Fig. 3. Here the moisture condition of the basin is introduced in terms of an antecedent precipitation index, a weighted summation of precipitation for approximately thirty days prior to the storm with the highest weights assigned to the most recent days. The significance of this antecedent index is modified by the introduction of the week of the year to reflect at least the normal evapotranspiration characteristics and the extent of interception by foliage. Storm characteristics are expressed in terms of duration and amount of precipitation. If the variation in depth of precipitation is great, the average over the basin may not reflect the true runoff because of the curvilinear relations involved. In this case runoff may be computed for subareas of the basin and averaged to get the best estimate of average runoff.

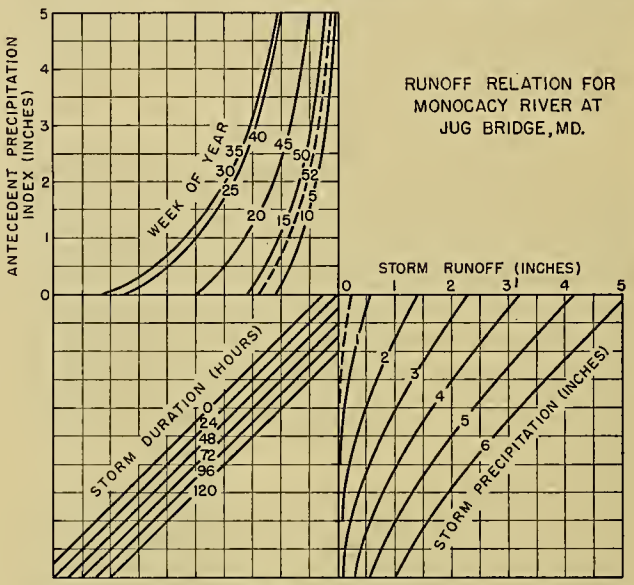


Fig. 3.—Rainfall-runoff relation for the Monocacy River above Frederick, Md.

Runoff Distribution After the total volume of runoff is computed, usually in terms of the depth in inches over the basin, the final step is the determination of the time distribution of this runoff in terms of the stream flow at the outlet of the basin. Sherman [12] observed that the hydrograph shapes for storms of like distribution and equal duration were characteristically similar on any basin. This led to the concept of the *unit hydrograph* which is almost universally used today for determining hydrograph shape. The unit hydrograph (Fig. 4) is a composite of the hydrographs of storms of equal duration and similar areal distribution of rainfall, all reduced to a volume of one inch of runoff by dividing their ordinates by the runoff depth in inches. A series of such graphs must be available for

various durations and patterns of rainfall distribution on each basin. The forecaster then selects the one most nearly comparable to the conditions of the current storm. By multiplying its ordinates by the predicted volume of runoff for the storm, he determines the probable outflow hydrograph. For storms extending over several forecast periods, the predicted hydrographs for each period may be added to arrive at the total flow.

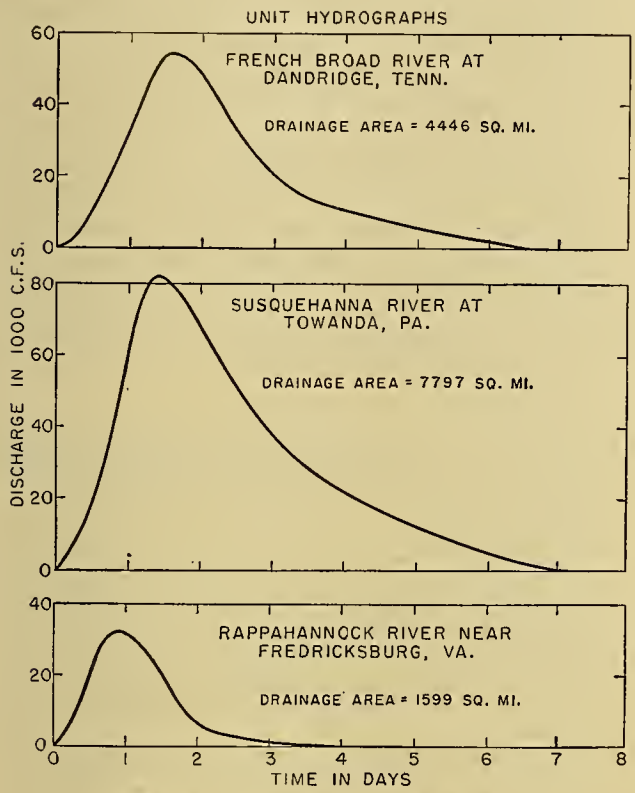


Fig. 4.—Typical unit hydrographs.

The problem of predicting the basin-outflow hydrograph is quite similar in its theoretical aspects to that of routing a flood wave downriver. The ordinary routing procedure makes it necessary to adopt routing periods so short that the computations for a headwater basin would be too time-consuming for forecasting purposes. However, the electronic routing machine [6] makes this approach a practical reality. In fact, the use of headwater forecasting techniques may be extended to areas much larger than can now be successfully treated with the unit hydrograph.

Snow-Melt Forecasting. Forecasting the runoff from melting snow poses many problems which have not yet been solved satisfactorily. In level terrain, the rate of melting of snow has been forecast with reasonable accuracy by use of factors expressing melt per degree day above 32F. If melting is concurrent with rainfall, the water equivalent of the snow may be added to the rainfall and a relation such as Fig. 3 used to compute runoff.

In mountainous terrain the situation is much more difficult. Here the snow pack accumulates in the form of a wedge with its greatest depths at high elevations, tapering to zero depth at intermediate and low levels. Owing to the variation of temperature with elevation, melting normally occurs over a fairly narrow range of elevation immediately above the snow line. The term "snow line" is used to describe the lower limit of the snow pack and does not imply a contour of constant elevation. As this melting zone moves upslope with the annual rise of temperature, the distance to the outlet of the basin increases and the effect of a unit of melt on the outflow changes.

Although it is believed that our understanding of the problem is sufficiently complete to permit a solution by the use of the multiple graphical correlation so effectively used in other phases of forecasting, records of snow-line elevation or area covered by snow have never been kept on a systematic basis and it has been exceedingly difficult to procure data which are adequate for such analysis.

Water-Supply Forecasting. The time lag of several months between the accumulation of snow in mountain areas and its subsequent melting has led to the development of methods for forecasting the volume of stream flow to be expected from a winter's accumulation of snow. Such forecasts make possible the scheduling of hydroelectric operations and the planning of agricultural work in terms of the water which will be available for irrigation.

Two different methods have been used in the preparation of such forecasts. One method uses relations derived by correlation of accumulated winter precipitation with runoff. Both precipitation and stream-flow data must be adjusted to eliminate time trends, and weights are assigned to the various precipitation stations and months of the year in proportion to their significance in the correlation [5]. The other technique makes use of a direct relation between the water equivalent of snow on the ground at the end of the accumulation season and subsequent runoff [2]. Both approaches seem to have about the same level of accuracy, with some advantage in reliability going to the first because of the longer record lengths which permit more careful statistical analysis. In all probability, a technique combining the two types of basic data would yield the most accurate forecasts.

Closely related to the problem of long-range volume forecasts is that of long-range estimates of the seasonal peak flow. Such forecasts have been attempted for the larger basins with moderate success. However, meteorological conditions during the late spring and early summer months play a far greater role in determining the peak flow than in determining the total volume. With the same amount of snow available, continued high temperatures for two or three weeks may cause a flood, while the same temperatures interspersed with short periods of cool weather will result in only moderate flows. Hence, large errors in long-range peak-flow forecasts must be expected in some years.

The Role of Meteorology in River Forecasting. Both short-range flood forecasts and long-range estimates of water-supply volume or seasonal peak flow may be seriously in error if weather conditions differ materially

from those anticipated by the forecaster. Hence, improvement of quantitative weather-forecasting techniques will materially aid the river forecaster, for he can treat weather forecasts in the same manner as observed data to synthesize the stream-flow situation which will develop. The U. S. Weather Bureau uses a statistical approach to this problem in its water-supply forecasts, in which estimates of probable flow volume are given for maximum and minimum of record, quartile, and median precipitation during the late spring and early summer months. A system which would interpret the meteorological situation in terms of an extended-range weather forecast would be much superior to such a statistical evaluation.

Broadly speaking, the river forecaster needs two types of weather forecasts. For flood forecasting he needs a detailed forecast extending from 1 to 10 days into the future, giving the amount, time of occurrence, and areal distribution of precipitation and, if snow melt is involved, temperature. For water-supply forecasting a less detailed outlook for 30 to 90 days ahead, indicating in general terms the precipitation anomalies and temperature trends, is necessary. Such forecasts would, in almost every case, give sufficient advance warning to permit carefully planned action, thus avoiding the waste and inefficiency of emergency methods. This need is highlighted by the existence of flood-control reservoirs which require a foreknowledge of the probable stream flow as much as thirty days in advance for successful operation.

REFERENCES

1. Bernard, M., "The Primary Role of Meteorology in Flood Flow Estimating." *Trans. Amer. Soc. civ. Engrs.*, 109:311–382 (1944).
2. Church, J. E., "Principles of Snow Surveying as Applied to Forecasting Stream Flow." *J. agric. Res.*, 51:97–130 (1935).
3. Hafstad, K. C., "Reliability of Station-Year Rainfall Frequency Determinations." *Trans. Amer. Soc. civ. Engrs.*, 107:633–683 (1942).
4. Holzman, B., "Sources of Moisture for Precipitation in the United States." *U. S. Dept. Agric. Tech. Bull.* No. 589, Washington, D. C. (1937).
5. Kohler, M. A., and Linsley, R. K., "Recent Developments in Water Supply Forecasting from Precipitation." *Trans. Amer. geophys. Un.*, 30:427–436 (1949).
6. Linsley, R. K., Foskett, L. W., and Kohler, M. A., "Electronic Device Speeds Flood Forecasting." *Engng. News Rec.*, Vol. 141, No. 26, pp. 64–66 (1948).
7. Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H., *Applied Hydrology*. New York, McGraw, 1949.
8. Lundquist, R. E., and Richards, M. M., "Flood Forecast Centers—What Makes Them Tick." *Engng. News Rec.*, Vol. 141, No. 24, pp. 98–100 (1948).
9. Meyer, A. F., *Evaporation from Lakes and Reservoirs. A Study Based on Fifty Years' Weather Bureau Records*, 56 pp. Minnesota Resources Commission, St. Paul, 1942.
10. Richardson, B., "Evaporation as a Function of Insolation." *Trans. Amer. Soc. civ. Engrs.*, 95:996–1019 (1931).
11. Rohwer, C., "Evaporation from Free Water Surfaces." *U. S. Dept. Agric. Tech. Bull.* No. 271, Washington, D. C. (1931).
12. Sherman, L. K., "Streamflow from Rainfall by Unit-Graph Method." *Engng. News Rec.*, 108:501–505 (1932).
13. Spreen, W. C., "A Determination of the Effect of Topography on Precipitation." *Trans. Amer. geophys. Un.*, 28:285–290 (1947).
14. Sutton, O. G., "Wind Structure and Evaporation in a Turbulent Atmosphere." *Proc. roy. Soc.*, (A) 146:701–722 (1934).
15. Thornthwaite, C. W., and Holzman, B., "Measurement of Evaporation from Land and Water Surfaces." *U. S. Dept. Agric. Tech. Bull.* No. 817, Washington, D. C. (1942).
16. Williams, G. R., "Drainage of Leveed Areas in Mountainous Valleys." *Trans. Amer. Soc. civ. Engrs.* 108:83–114 (1943).

MARINE METEOROLOGY

Large-Scale Aspects of Energy Transformation over the Oceans *by Woodrow C. Jacobs* 1057

Evaporation from the Oceans *by H. U. Sverdrup* 1071

Forecasting Ocean Waves *by W. H. Munk and R. S. Arthur* 1082

Ocean Waves as a Meteorological Tool *by W. H. Munk* 1090

LARGE-SCALE ASPECTS OF ENERGY TRANSFORMATION OVER THE OCEANS

By WOODROW C. JACOBS

Headquarters, Air Weather Service, Washington, D. C.

INTRODUCTION

There has been a recent and encouraging re-emphasis on investigations which involve some phase of the terrestrial heat budget, and the steadily accumulating number of investigations on the subject has brought proof that many meteorological and climatological problems can be wholly or partially solved through a consideration of energy transformation processes. In nearly all cases the purpose of the inquiries has been to ascertain the order of magnitude of one or more of the following components of the heat budget:

1. Radiative balance between the sun, earth, space, clouds, and atmosphere. This includes separate considerations of the solar constant and of the reflection, scattering, absorption, and emission of radiant energy by the earth's surface, clouds, and turbid (moist) atmosphere.
2. Exchange of sensible heat between the surface and the atmosphere.
3. Heat used for evaporation (or melting).
4. Heat realized through condensation (or freezing).
5. Heat produced through dissipation of kinetic and gravitational potential energy in the atmosphere (and ocean).
6. Transport of heat by ocean currents.
7. Lateral and vertical transport of real heat and latent heat (as water vapor) by the atmosphere.
8. Transfer of heat through conduction in the earth's surface (or through vertical mixing and layer transmission in the case of water bodies).
9. Changes in the internal energy of the atmosphere and of the surface layers of land and oceans.

Unfortunately, all of these factors in the energy budget of the earth and atmosphere are highly interdependent, a fact which renders their numerical computation exceedingly laborious. This is especially true if, instead of applying to average conditions, the evaluations are carried out for particular large-scale situations. It is because of this complexity that most of the general solutions to the large-scale problems which have been presented to date are greatly simplified and apply to average annual conditions and most often to global zones rather than to specific geographic areas. Nevertheless, a knowledge of the basic causes of space and time variations in temperature, humidity, pressure, winds, precipitation, and nearly everything else that contributes to the sum total of both weather and climate is obviously of fundamental importance in meteorology. Hence it is imperative that some consideration be given to the heat energy necessary to these conditions, its sources, when and where it is delivered, and how it is distributed.

Because additional details on energy transformation processes are covered elsewhere in this volume, the following discussion is limited to several narrow but important phases of the problem. The field of application covers only the ocean areas of the globe. The discussion is further restricted to cover only the large-scale and more or less long-term (average) aspects of the *convective* transfer of energy between sea surface and atmosphere. Space does not permit consideration of the radiative flux of heat energy between sea and atmosphere, nor are the short-term or "synoptic" aspects covered except as they are mentioned among the recommendations for future research.

Classical meteorological literature is replete with references to the profound effect of the oceans on world weather and climate. The moderating influence of oceans on the climates at high latitudes and particularly on the west coasts of continents has long been recognized, but steps to evaluate the "effects" have consisted largely of attempts to compare or classify the local climates on the basis of "maritimeness" or "continentality." Not until recently have serious efforts been made to evaluate the sea surface as a heat and moisture source and to demonstrate the role of the oceans in supplying energy for initiating and maintaining the general circulation of the atmosphere. When it is considered that the oceans comprise seventy-one per cent of the earth's surface—the surface from which the atmosphere receives by conduction and radiation all but a small fraction of its total heat energy—and that the oceans are the prime source of moisture for all atmospheric processes induced by water vapor, it is not unreasonable to expect that a solution to the problem of ocean-atmosphere energy relationships will carry the meteorologist a long way toward achieving a full understanding of the terrestrial energy budget. Unfortunately, the investigators who have dealt with the large-scale aspects of the problem are few in number, and the meager results of their studies have been presented in something less than a dozen brief papers. For this reason the author hopes he will be forgiven if in the following discussion he appears to make excessive reference to some of his own work along this line.

RATE OF EXCHANGE OF SENSIBLE HEAT BETWEEN SEA SURFACE AND ATMOSPHERE

Most of the earlier attempts to evaluate the rate at which sensible heat is conducted to the atmosphere

from the sea surface were incidental to computing oceanic evaporation on the basis of energy considerations.[1] In these cases very simple assumptions were made as to the magnitude of the ratio R between the amount of surplus heat given off directly to the atmosphere as sensible heat Q_h and the amount used for evaporation Q_e. Schmidt [26], in his determinations of the annual latitudinal evaporation from the oceans, accepted values for R which ranged from about 1.67 to 0.28, but subsequent analysis by Ångström [2] indicated that these values were much too high to represent average conditions over the oceans. Ångström concluded that only about 10 per cent of the heat surplus is given off to the atmosphere by conduction and that about 90 per cent is used for evaporation. Mosby [18], in recomputing the annual latitudinal evaporation over all oceans, accepted Ångström's supposition and considered R to be constant at 0.10 at all latitudes. McEwen [13], on the other hand, assumed R to be constant at 0.20 when determining the annual latitudinal evaporation over the North Pacific.

Up to 1940 none of the investigators had made any attempt to evaluate separately the sea-surface and time variations in Q_h or in the ratio R. Nevertheless, Bowen [3] in 1926 had derived a formula which proposed to establish the relation between evaporation and heat exchange at a water surface and this formula had been applied by Cummings and Richardson [5] in determining the evaporation from lakes. In considering the rates at which heat and water vapor are transported across the lower and upper surfaces of a volume of air in contact with a water surface, Bowen had concluded that the ratio between the heat loss by conduction and that by evaporation can be obtained from the expression

$$R = 0.64 \frac{p}{1000} \left(\frac{t_w - t_a}{e_w - e_a} \right), \qquad (1)$$

where t_w and t_a are the water and air temperatures respectively, e_w is the vapor pressure of the water surface, e_a is the vapor pressure of the air, and p is atmospheric pressure (all pressures in millibars). Bowen's derivation of this formula is quite involved but the same equation has been derived quite simply by Sverdrup [34] and the method will not be repeated here.

Sverdrup [35] subsequently pointed out that the application of the Bowen ratio might be invalidated if it proved necessary to consider the effects of the radiative transfer of heat through the laminar layer next to the sea surface, and if the evaporation from the sea surface is greatly increased at wind velocities high enough to carry spray into the air. However, Sverdrup [36] has more recently presented data to indicate that near the boundary surface the radiative transfer of heat is relatively unimportant. In addition, through a comparison between the observed humidity gradients existing above the sea surface and those derived on the basis of theoretical considerations of the transfer of water vapor within the boundary layer, he has also concluded that the effect of the evaporation from spray is relatively unimportant. He states that "The successful application of the theory of turbulence to the problem of air mass transformation indicates that there exists no large difference in the processes by which heat and water vapor are diffused."

Although the Bowen formula has suffered extensive criticism from meteorologists, the available observations to date indicate that it is capable of giving values for the ratio Q_h/Q_e of the approximate order of magnitude, at least if reasonably correct temperature and humidity observations are obtained near the sea surface and if the computations involve use of data obtained over extensive water bodies where the effects of lateral mixing can be neglected.

The author [6], following a suggestion given by Sverdrup, applied the Bowen ratio to seasonal evaporation values computed for five-degree squares over the North Atlantic and North Pacific and demonstrated that this ratio is a highly variable quantity, both seasonally and with respect to the regional distribution. The ratio proves to be low, or even negative, in regions overlain by dry air and where the temperature differential between sea surface and atmosphere is small (as in the trade-wind areas). High values are found in regions where warm water is overlain by relatively moist (often cool) air (as over the Gulf Stream and Kuroshio during winter). The average seasonal five-degree zonal values are shown to range from -1.5 to $+0.6$, although most values were within the range from -0.4 to $+0.5$.

The seasonal and annual values for the rate of exchange of sensible heat with each five-degree square over the North Atlantic were then obtained by applying the Bowen ratio as follows:

$$Q_h = RL_tE \text{ cal cm}^{-2} \text{ day}^{-1}, \qquad (2)$$

where L_t is the latent heat of vaporization at average sea-surface temperature t, and E is the evaporation rate in centimeters per day. The results of the computations on the basis of the annual data only are given in Fig. 1; the full set of seasonal charts is to appear in a later publication [10]. During all seasons the isolines for Q_h show, roughly, the same configuration as those for E (or Q_e); these values being at their maximum in winter and along the western sides of the oceans in mid-latitudes. However, one important difference is shown. No tropical areas of maximum Q_h appear within the trade-wind regions to correspond to the areas of maximum evaporation found in those areas.

The average seasonal values of Q_h for the various latitude zones are given in Table I. One interesting aspect of these data is the fact that the values are higher in the North Pacific than in the North Atlantic at all latitudes and during all seasons except winter for the areas north of latitude 35°N. From the data so far accumulated it appears that the atmosphere is being directly heated by the sea surface at significant rates only in the middle and high latitudes, along the eastern sides of the continents and principally during the winter

1. Consult "Evaporation from the Oceans" by H. U. Sverdrup, pp.1071–1081 in this Compendium.

season. During summer it appears that over large areas the sea is actually receiving some energy by conduction from the atmosphere.

In connection with the probable accuracy of the author's Q_h values, one phase of which is discussed by Sverdrup elsewhere in this volume,[1] it should be pointed out that all of the computations of evaporation and heat exchange were based on data published by the U. S. Weather Bureau [38] and represent ship observations recorded over the oceans at or very near Greenwich Meridian noon. As a result, most of the observations in the North Atlantic were taken during daylight hours, between 0600 and 1200 LMT, while those in the North Pacific have been obtained largely during night hours, between 2000 and 0600 LMT. Although the diurnal variations in evaporation or heat exchange over the oceans can be assumed to be small [33], the values computed from the *uncorrected* observational data obtained on shipboard might conceivably show significant differences between the two oceans, partly because of a true diurnal variation in the quantity $(t_w - t_a)$ and partly because of an apparent variation created by the daytime heating and nighttime cooling of ship and instruments. The author has previously considered this latter possibility [7] when attempting to account for the very nearly constant latitudinal differences in the annual values for R between the North Atlantic and North Pacific.[2]

The author has more recently re-examined the question of possible errors in the computed Q_h values—errors which might result from the time differences in

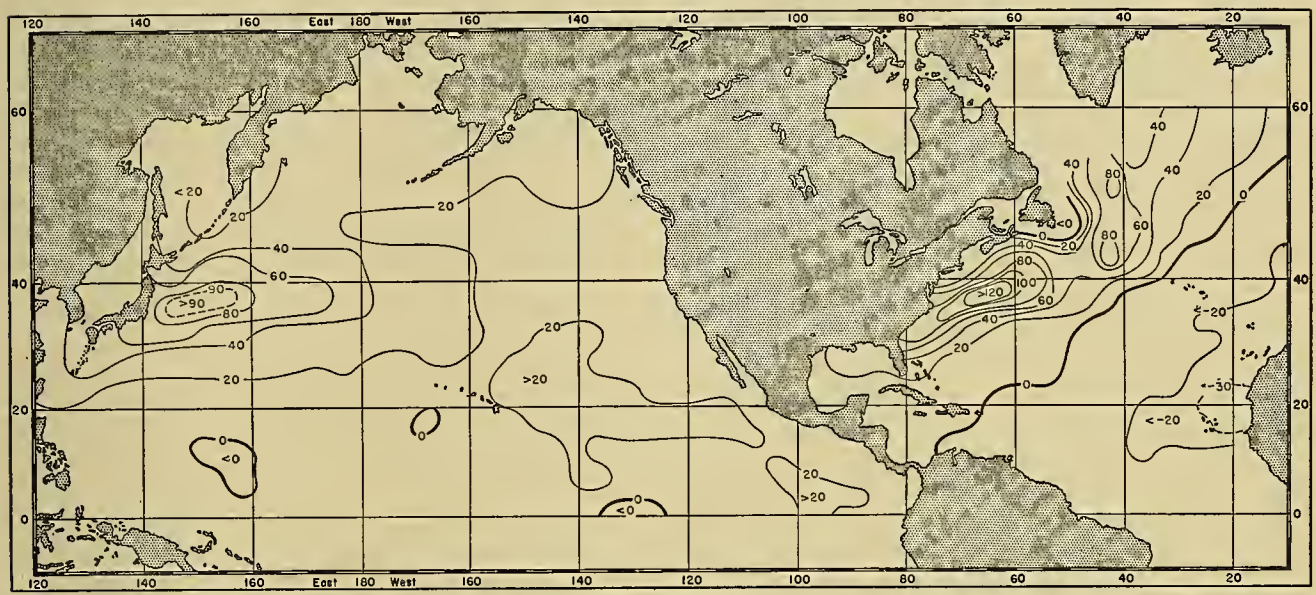


FIG. 1.—The annual values of the rate of exchange of sensible heat between ocean and atmosphere Q_h over the North Atlantic and North Pacific, expressed in calories per square centimeter per day.

TABLE I. SEASONAL VALUES OF Q_h IN DIFFERENT LATITUDE ZONES (in cal cm^{-2} day^{-1})

North latitude zone	North Atlantic*				North Pacific			
	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.
0°– 5°	−8	−3	−10	−15	6	6	12	11
5°–10°	−10	−13	−3	−4	11	10	13	13
10°–15°	−8	−16	−11	−5	10	9	10	13
15°–20°	0	−14	−16	−4	20	11	11	11
20°–25°	6	−13	−16	2	34	17	7	14
25°–30°	22	−4	−12	10	51	19	4	20
30°–35°	37	−3	−16	10	69	22	2	29
35°–40°	95	18	−14	29	106	29	0	40
40°–45°	99	14	−23	24	89	16	−8	31
45°–50°	80	−4	−38	15	49	9	−10	20
50°–55°	109	10	−33	29	59	17	−7	32
55°–60°	121	5	−28	34	70	25	−3	31

* The North Atlantic includes the Gulf of Mexico and Caribbean Sea, within the appropriate latitude ranges for all zonal energy values, in this and succeeding tables.

2. The mean annual values for R arranged by five-degree ranges of latitude show that for nearly every zone in the North Pacific the value for the ratio R is 0.10 or 0.11 greater than for the corresponding zone in the North Atlantic. The seasonal values, however, do not exhibit this relationship.

ship observations. For this purpose a sampling has been made of the extensive hourly sea and air temperature data obtained on the last *Carnegie* Expedition [11]. In order to test the diurnal variation of $(t_w - t_a)$ under conditions such that radiation effects should be at a maximum, the analysis was limited to hourly temperature data collected on 103 days when the *Carnegie* was in the tropics or, during summer months, when it was at the higher latitudes. The hourly values for $(t_w - t_a)$, when t_a was uncorrected for radiation effects (see Table II), varied from a maximum of +0.90C between the hours of 0200 and 0400 LMT to a minimum of −0.24C at 1200 LMT, giving an extreme range of 1.14C. When t_a data corrected for radiation effects were used (see [11] for discussion of correction method), the time and magnitude of the maximum values for $(t_w - t_a)$ remained unchanged, but the hour of occurrence of the minimum was shifted to 1000 LMT and the previous negative value raised to +0.21C, giving a corrected range of 0.69C. If it is assumed that the method of correcting the *Carnegie* data for radiational effects is valid, these differences indicate that about 40 per cent of the apparent diurnal variation in $(t_w - t_a)$ is brought about by daytime heating of ship and instruments.

From the standpoint of the effects of the diurnal variation of $(t_w - t_a)$ upon the published results for Q_h, it appears that the time error should be significant only in those regions where Q_h is small or negative and where the temperature observations were obtained during daylight hours. Fortunately this effect was noticeable on the author's charts only in the eastern and southern North Atlantic where $(t_w - t_a)$ is small or negative and where the observations were recorded during the late forenoon. On the charts for Q_h this correction would serve to restrict the negative areas in the North Atlantic, giving slight positive values to the peripheral portions of those areas which at present indicate negative values. None of the values for the North Pacific should be noticeably affected. In future computation of Q_h, when Greenwich Meridian noon observations are used, particular care should be exercised to remove the diurnal effects from the $(t_w - t_a)$ data which represent ocean areas from about 60°W eastward to 90°E. (See Table II.)

One of the principal objections to the use of the Bowen ratio for establishing Q_h is that it requires both extensive humidity observations and observed or calculated evaporation rates. Neither of these two classes of data are available for the Southern Hemisphere nor do such data exist in any quantity for the higher latitude oceans. A more direct method for determining the sensible heat flux between sea and atmosphere would be desirable, preferably one which does not involve the use of either humidity or evaporation data.

In a recently published article, Miyazaki [15] presents monthly values of Q_h within five restricted areas "along the *Tusima* warm current" (Sea of Japan). The results have been obtained through use of an equation developed by Kuzmin and Saito [12] in which only the quantities $(t_w - t_a)$ and W_a, the wind velocity at height a, are needed.[3] However, the validity of the method is not established by the computation nor is any attempt made to adjust the formula to fit the type of observational materials which were used.

Burke [4] has developed a method for determining the changes in the heat content of continental polar air as it moves out over a warm sea surface. The technique requires a knowledge of the initial air temperature, the initial lapse rate in the air mass, and the distance the air mass has traveled over a sea surface exhibiting a

Table II. Mean Hourly Values of $t_w - t_a$ (°C) over the Oceans from a Sampling of the "Carnegie" Data

Hour (LMT)	0	1	2	3	4	5	6	7	8	9	10	11	12
Meridian equivalent to GM noon	180°W	165°W	150°W	135°W	120°W	105°W	90°W	75°W	60°W	45°W	30°W	15°W	0
Corrected data (*a*)	0.82	0.89	0.90	0.90	0.90	0.87	0.81	0.66	0.43	0.23	0.21	0.41	0.36
Uncorrected data (*b*)	0.82	0.89	0.90	0.90	0.90	0.87	0.81	0.65	0.40	0.16	−0.05	−0.15	−0.24
Difference (*a*) minus (*b*)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.07	0.26	0.56	0.60

Hour (LMT)	13	14	15	16	17	18	19	20	21	22	23	Mean all hours
Meridian equivalent to GM noon	15°E	30°E	45°E	60°E	75°E	90°E	105°E	120°E	135°E	150°E	165°E	
Corrected data (*a*)	0.23	0.38	0.35	0.42	0.53	0.67	0.74	0.74	0.77	0.75	0.77	0.62
Uncorrected data (*b*)	−0.23	−0.17	−0.09	0.13	0.30	0.50	0.66	0.73	0.77	0.75	0.77	0.46
Difference (*a*) minus (*b*)	0.46	0.55	0.44	0.29	0.23	0.17	0.08	0.01	0.00	0.00	0.00	0.16

(*a*) t_a corrected for radiation effects.
(*b*) t_a uncorrected for radiational heating and cooling of ship and instruments.

3. The final equation which was used for the computations is

$$Q_h = \frac{4.15\,(t_w - t_a)W_a}{5.742\left[\ln\left(\frac{a + z_0}{z_0}\right)\right]^2 + 0.16W_a} \text{ cal cm}^{-2}\text{ day}^{-1}, \quad (3)$$

where z_0 is the roughness parameter (given as a function of wind speed after Krümmel and increasing from 0.25 cm at W_a = 100 cm sec^{-1} to 27.00 cm at W_a = 2000 cm sec^{-1}). The thickness of the laminar boundary layer is assumed constant at 0.16 cm.

known temperature gradient. For these reasons the equations are not in convenient form for application in the large-scale climatological sense. This statement is not to be construed as a criticism of the method, however, since it has been designed for use by forecasters and not for the evaluation of large-scale energy transformation processes over the oceans. Nevertheless, the method might prove invaluable for investigations, on a limited areal scale, of the synoptic aspects of heat transfer between sea surface and atmosphere. It might also be applied in the climatological sense within those peripheral ocean areas where, during one or more of the seasons, air trajectories remain fairly constant and the initial continental air-mass characteristics do not show great variation (as, perhaps, over the Sea of Japan in winter).

RATE OF EXCHANGE OF ENERGY IN THE LATENT FORM OF WATER VAPOR

Of the total solar energy absorbed at the sea surface during the course of a year, approximately fifty per cent is used for evaporating sea water and, therefore, is made available to the atmosphere in the latent form of water vapor.[4] In view of this, and considering the relative magnitudes of land and sea areas, it would appear that the latent energy represented by water vapor derived from ocean evaporation constitutes the most important single component of the atmospheric heat budget. It is obvious that any paper which proposes to deal with atmospheric energy transformation processes should give primary attention to the problem of ocean evaporation. The fundamental details concerning this problem are so well covered by another article in this Compendium[1] that it will not be necessary here to expound the theoretical aspects of the processes governing the transfer of water vapor from sea surface to atmosphere. For this reason the present discussion will be limited to consideration of the large-scale and seasonal aspects of ocean evaporation in terms of its energy equivalence.

It has long been known that evaporation is governed by atmospheric humidity, surface water temperature, and wind speed. Nevertheless, empirical methods which have been used in the past to relate evaporation to these factors generally have not been successful, either because the equations contained unevaluated functions or because they were constructed to fit a special limited set of observations. In addition, the several evaporation equations derived on the basis of theoretical considerations have given equally divergent results, and no single theoretical method for computing evaporation has yet been devised which has general application when use is made of temperature, humidity, and wind data obtained under a variety of observational circumstances.

4. According to Mosby [18], 41 per cent of the total solar energy absorbed at the sea surface between the 70th parallels is radiated directly back to space, 53 per cent is used for evaporation and 6 per cent is conducted to the atmosphere as sensible heat. However, on the basis of the author's data, it appears that the last figure should be revised upward.

This state of affairs led Sverdrup in 1940 to suggest to the author that it might be possible to compute the average evaporation over the oceans, using available marine climatic data, by applying an equation of the type:

$$E = K\,(e_w - e_a)\,W_a, \tag{4}$$

where K is an empirical "evaporation factor" arrived at by comparing the long-term annual ocean evaporation computed through the use of the energy equations with that computed for the same period through the use of the several existing equations which involve theoretical considerations of the interchange of water vapor within the turbulent boundary layer near the sea surface. In equation (4), e_w is the vapor pressure at the sea surface, e_a is the vapor pressure at height a above the sea surface, and W_a is the wind speed at height a.

In following this suggestion, the mean annual evaporation was computed for each of four selected areas in the North Atlantic and North Pacific by two methods: first, by using an energy-budget method similar to that employed by Mosby [18] and arriving at values designated as E_1, and second, by using the evaporation equations of Sverdrup [31] and Montgomery [16] and arriving at values for K (and E). Since, as a first step in computing evaporation by the energy equations, it was necessary to disregard the oceanic heat advection term, the areas selected were those within which the latitudinal transport of surface waters is at a minimum and, at the same time, those allowing a rather wide sampling of latitudinal differences in available solar energy (Table III). The marine climatic data which entered into the computations were obtained from U. S. Weather Bureau sources [38].

Of the two theoretical equations which were used to arrive at the final values for K in equation (4), the one presented by Montgomery is in somewhat simpler form than that presented by Sverdrup and for that reason it will be exemplified in the following discussion. According to Montgomery,

$$E = \rho k_0 \gamma_a \Gamma_a\,(q_w - q_a)\,W_a, \tag{5}$$

where ρ is air density, $k_0 = 0.4$ is the Kármán constant, γ_a the resistance coefficient, Γ_a the evaporation coefficient, q_w the specific humidity at the sea surface, and q_a the specific humidity at height a. With all quantities in cgs units, the evaporation is expressed in grams per square centimeter per second. It should be pointed out that equation (5) does not deal with instantaneous values of humidity and wind speed but probably applies to these quantities averaged over a period of not less than an hour.

The coefficients γ_a and Γ_a in equation (5) depend upon the height a at which humidity and wind speed are measured, and also upon the character of the sea surface. The sea surface can be considered hydrodynamically *smooth* at wind speeds less than about 650 cm sec^{-1} (measured at a height of 600 cm) and hydrodynamically *rough* at higher wind speeds [19, 24, 25].

If averages for a smooth surface and maxima for a

rough surface are used, the numerical values of the constants are:

		γ_a	Γ_a
Smooth surface	($W_{600} < 650$ cm sec^{-1})	0.030	0.085
Rough surface	($W_{600} > 650$ cm sec^{-1})	0.059	0.148

With $q = 0.623\ e/p$, where e is vapor pressure and p is atmospheric pressure, and $\rho = 1.25 \times 10^{-3}$ and $p = 1000$ mb:

Smooth surface $E = 0.79 \times 10^{-3}\ (e_w - e_a)\ W_a$, (6)
Rough surface $E = 2.71 \times 10^{-3}\ (e_w - e_a)\ W_a$. (7)

The hourly evaporation in grams or centimeters becomes:

Smooth surface $E = 2.8 \times 10^{-6}\ (e_w - e_a)\ W_a$, (8)
Rough surface $E = 9.8 \times 10^{-6}\ (e_w - e_a)\ W_a$. (9)

However, when climatological data are used, the evaporation should be computed from the equation

$$\bar{E} = 2.8 \times 10^{-6} \{n[\overline{(e_w - e_a)W_a}]_{W_a<650} + 3.5m[\overline{(e_w - e_a)W_a}]_{W_a>650}\}, \quad (10)$$

where n and m represent the fraction of hours with wind speeds less or greater than 650 cm sec^{-1}, respectively ($n + m = 1$), and where the notation $\overline{(e_w - e_a)W_a}$ represents the average of hourly values of the product $(e_w - e_a)W_a$. When dealing with climatological data, only the average monthly or seasonal values $\overline{(e_w - e_a)}$ and $\overline{W}_a$ are known, and even if the percentage of winds of speeds less than 650 cm sec^{-1} (less than force Beaufort 4) is determined, the corresponding value of $\overline{(e_w - e_a)W_a}$ is still unknown. It remains to determine whether evaporation could be computed with a reasonable degree of accuracy, using a simple equation similar to (4), where for hourly values the factor K might be expected to lie between 2.8×10^{-6} and 9.8×10^{-6}, if E is expressed in grams (or centimeters), $\bar{e}$ in millibars, and $\overline{W}_a$ in centimeters per second.

The climatological values for q_w, q_a, W_a, and the computed values (hourly) for E_1 were substituted in equation (5), and the factor K was determined for each of the four areas. The results are given in Table III. A considerable spread in the K values among the four areas is indicated but this is the expected result because it is obvious that the advection term cannot be neglected when computing evaporation by the energy equations for any restricted ocean area. The computed quantities E_1 and E could be expected to be equal only in the absence of lateral heat transport or if the entire ocean surface could be included in their computation.

The average value of K for the four areas proves to be approximately 6×10^{-6}, that is, a value which lies about halfway between that applicable to a smooth surface and that which applies to a rough surface. Thus, if E is the 24-hr evaporation in millimeters or in decigrams per square centimeter and W_a the wind speed in meters per second, equation (4) becomes

$$E = 0.142\ (e_w - e_a)\ W_a. \quad (11)$$

It would be expected that the values for E_1, computed by means of the energy equations, would tend to be too high at the lower latitudes and too low at the higher latitudes because of the neglect of the advection term. This difference should be particularly noticeable in the North Atlantic where the dialatitudinal transport of surface waters from lower to higher latitudes is considerably greater than in the North Pacific. The data in Table III show that this is the case. Sverdrup has

Table III. Values of K, E_1, and E for Selected Areas

Latitude	Longitude	E_1* (10^{-3} cm hr^{-1})	$e_w - e_a$ (mb)	W_a (cm sec^{-1})	K	E† (10^{-3} cm hr^{-1})	E_1/E
North Atlantic							
40°N–45°N	20°W–50°W	9.59	2.5	840	4.6×10^{-6}	12.51	0.77
20°N–25°N	25°W–65°W	19.12	4.0	573	8.3×10^{-6}	13.58	1.41
North Pacific							
40°N–45°N	140°W–160°E	7.73	1.9	760	5.4×10^{-6}	8.58	0.90
20°N–25°N	130°W–140°E	16.92	5.3	540	5.9×10^{-6}	17.04	0.99

* E_1 = evaporation computed on the basis of the energy equations.
† E = evaporation computed by equation (4) with $K = 6 \times 10^{-6}$.

further analyzed the significance of the heat advection term elsewhere in this volume[1] and he points out that the zonal differences in K are of the proper order of magnitude to be accounted for by the zonal differences in the heat advection term. Moreover, the computations of E made for the North Atlantic and North Pacific through use of equation (11), and averaged for the year for the entire area, give results which are in agreement with those previously reached through energy considerations. Nevertheless, there must still remain some uncertainty as to the average value of K. However, the time and space variations in evaporation over the oceans are so large that it is believed that the remaining uncertainty is relatively unimportant.

Through the use of equation (11), which applies only to the particular marine climatic data which were used, computations of the seasonal and annual values of E have been made for each five-degree square in the North Atlantic and North Pacific. The results have been converted into their energy equivalents by taking L_t to be 585 cal g^{-1} and thus letting

$$Q_e = 585\ E \text{ cal cm}^{-2} \text{ day}^{-1}, \quad (12)$$

where E is given in centimeters per day.

The results of the computations on the basis of the annual data only are given in Fig. 2. The full set of seasonal evaporation charts upon which the annual Q_e is based will appear in the publication to which reference has previously been made [10]. The computations show that the regions of greatest energy exchange are on the western sides of the oceans and around the semipermanent high-pressure fields. The values for Q_e on the eastern sides of the oceans are generally smaller. The absolute maximum value for Q_e for all oceans occurs in the North Atlantic in winter within the Gulf Stream between latitudes 35°N and 40°N and about 700 miles east of the Virginia Capes, where the average Q_e is about 670 cal cm^{-2} day^{-1}. The maximum value in the North Pacific, 550 cal cm^{-2} day^{-1}, occurs farther south and nearer the coast within the Kuroshio between latitudes 25°N and 30°N and approximately 300 miles northeast of Formosa. As shown by the data in Table IV, except for equatorial areas, the values for Q_e are nearly everywhere greatest during winter and generally lowest during summer. However, the summer decrease is small within the tropical areas of high evaporation. The quantity Q_e remains positive during all seasons for all areas except in the North Pacific north of latitude 55° and west of 160°W, and in the North Atlantic waters immediately surrounding Labrador, where slight negative values are shown during summer months (where $e_a > e_w$). A comparison between the data in Table I and Table IV indicates that over the oceans as a whole the atmosphere is being directly heated by the sea surface most rapidly in the higher latitudes, but that it is receiving moisture principally in the middle and lower latitudes.

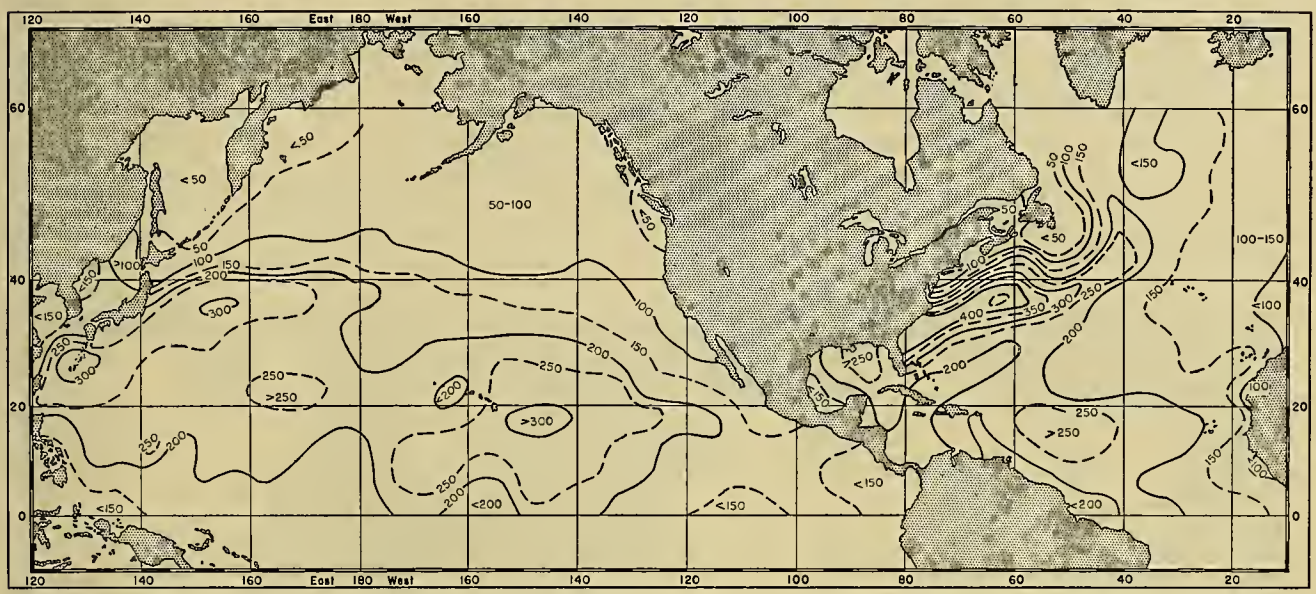


Fig. 2.—The annual values of the rate of energy loss from the sea surface through evaporation Q_e over the North Atlantic and North Pacific, expressed in calories per square centimeter per day. (The values above can be converted into rough evaporation rates by considering that the isometric interval of 50 cal cm^{-2} day^{-1} is approximately equivalent to an evaporation rate of 12 in. yr^{-1}.)

Table IV. Seasonal Values of Q_e in Different Latitude Zones (in cal cm^{-2} day^{-1})

North latitude zone	North Atlantic				North Pacific			
	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.
0°– 5°	144	162	203	181	167	127	185	186
5°–10°	219	191	149	148	211	200	181	172
10°–15°	249	220	195	170	234	242	192	180
15°–20°	250	198	211	202	240	232	224	203
20°–25°	220	159	176	197	268	212	204	232
25°–30°	234	169	136	246	276	189	153	231
30°–35°	247	171	119	232	260	145	98	216
35°–40°	342	193	108	264	268	149	74	234
40°–45°	250	123	56	193	169	81	37	148
45°–50°	207	98	25	127	100	59	22	104
50°–55°	246	112	64	142	97	69	26	111
55°–60°	260	104	43	147	109	49	30	77

Concerning additional direct methods for determining Q_e over ocean areas, Miyazaki [15] has attempted to apply a modification of the Sverdrup evaporation

equation [32] directly to determine the monthly averages of Q_e within the same five areas for which he determined Q_h.[5] He does not attempt to evaluate the equation empirically against the available observational materials and he resorts to humidity data from land stations as a substitute for observations at sea. His monthly averages for Q_e appear somewhat too large during summer (the three-month summer average for the five areas proves to be 116 cal cm^{-2} day^{-1}).

RATE OF TOTAL HEAT LOSS FROM THE OCEANS THROUGH CONVECTION

The only method for directly estimating the total convective heat loss Q_a from a water surface appears to be the one recently developed by Montgomery [17]. He presents a detailed discussion of methods for estimating the total eddy flux of heat from a moist surface by combining the effects of heat transfer by conduction and heat transfer by evaporation (or surface condensation) and, after deriving several general expressions, he arrives at a more restricted equation which yields good approximations for Q_a under certain conditions. These conditions are that the humidity and temperature observations be obtained within 10 m of the water surface and that the pressure should be within about 10 per cent of standard pressure. He shows that if the *specific enthalpy*, including the latent heat of water vapor, is designated by h_t, and if V_z' is a specially defined fluctuation of the vertical component of velocity, then the unit-area eddy flux of heat is $\langle \rho V_z' h_t \rangle$ and

$$\langle \rho V_z' h_t \rangle = \left(\int_a^b \frac{dz}{k\bar{\rho}} \right)^{-1} c_{pd}(\overline{T}_{ea} - \overline{T}_{eb}), \qquad (14)$$

where k is the *eddy diffusivity* (Richardson), c_{pd} is the isobaric specific heat of dry air, $\overline{T}_{ea}$ is the mean *equivalent temperature* at the surface, and $\overline{T}_{eb}$ is the same quantity at a chosen, short distance above. For this purpose, the *equivalent temperature* of a sample of air is defined as the temperature of dry air having the same enthalpy.

Montgomery's method, however, has not yet been made applicable to the type of observational materials generally available in the marine climatic record. Furthermore, the meteorologist generally is more interested in the separate fractions of the energy exchange, Q_h and Q_e, than he is in their sum. Nevertheless, the method could facilitate computations of the rate of total "contact" heat loss from the sea surface and thus be useful to the oceanographer when he finds it convenient to evaluate the sea-surface heat loss for special analytical purposes and when it is possible for him to obtain controlled meteorological observations.

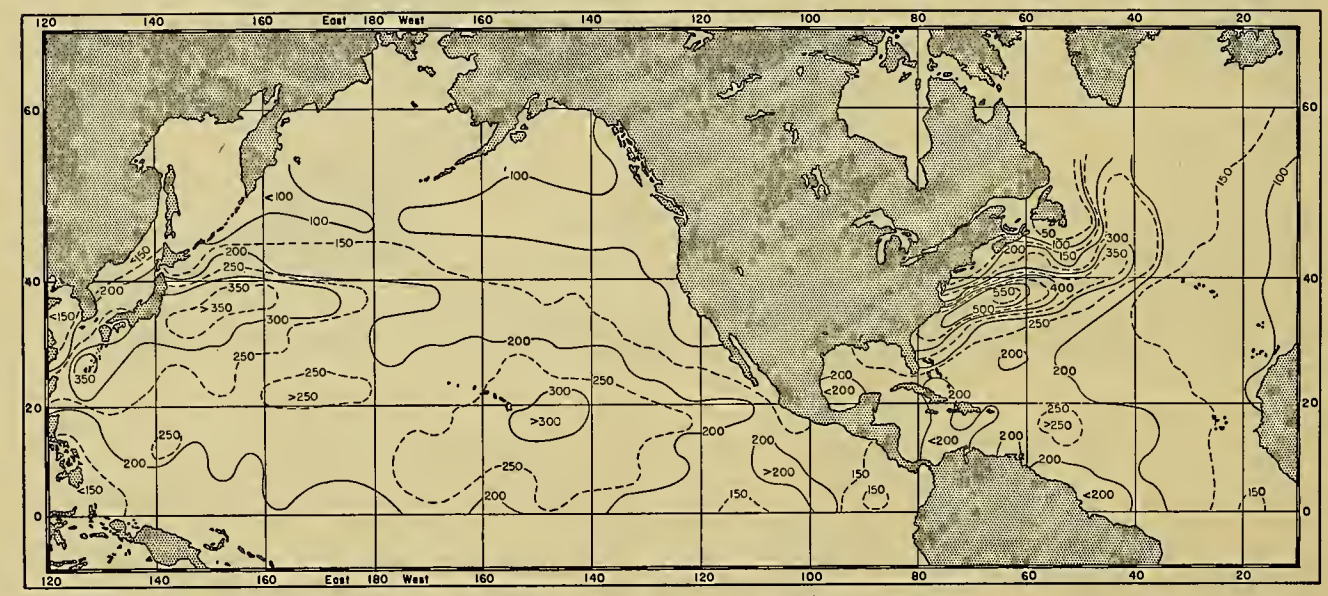


FIG. 3.—The annual values of the rate of total energy loss from the ocean surface through convection ($Q_a = Q_h + Q_e$) over the North Atlantic and North Pacific, expressed in calories per square centimeter per day.

The only available large-scale computations of the total energy exchange between the ocean and atmosphere have been presented in chart form for the North Pacific and North Atlantic [6, 10] and are based on the sum

$$Q_a = Q_h + Q_e. \qquad (15)$$

Therefore, they represent the summation of computations which have already been described in the preceding two sections.

In their general configuration the charts for Q_a are quite similar to the charts for Q_e, except that the

5. The equation used is:

$$Q_e = \frac{590 \times 2.15 \times 10^{-2}(e_w - e_a)W_a}{8.64\left[\log \frac{(a + z_0)}{z_0}\right]^2 + 0.16W_a} \text{ cal cm}^{-2} \text{ day}^{-1}, \qquad (13)$$

where $L_t = 590$ cal g^{-1}.

west-east gradient in Q_a is intensified due to the high values of Q_h over the western sides of the oceans and the low values over the eastern portions. The charts are also, in a qualitative manner, quite similar to those for Q_h except that the latter show no maximum areas of sensible heat exchange within the tropical trade-wind region corresponding to the areas of maximum Q_a in this belt. In the North Atlantic in winter, Q_a reaches values as high as 937 cal cm^{-2} day^{-1} within the Gulf Stream off the Virginia Capes. In the North Pacific the maximum value for Q_a occurs in winter within the Kuroshio northeast of Formosa, coinciding with the area of maximum Q_e, where a value of 728 cal cm^{-2} day^{-1} is computed, with a secondary area of maximum Q_a (718 cal cm^{-2} day^{-1}) between latitudes 35°N and 40°N approximately 350 miles east of Japan. The annual chart for Q_a is given as Fig. 3.

The seasonal values for Q_a arranged by latitude zones are given in Table V. These data show that, except near the equator, Q_a is at a maximum during winter and at a minimum during summer with the quantities tending to be greater during autumn than during spring. The seasonal variations are large in middle and high latitudes but small near the equator. Both oceans show secondary minima in the total energy exchange between latitudes 45°N and 50°N.

Table V. Seasonal Values of Q_a in Different Latitude Zones (in cal cm^{-2} day^{-1})

North latitude zone	North Atlantic				North Pacific			
	Dec.-Feb.	Mar.-May	June-Aug.	Sept.-Nov.	Dec.-Feb.	Mar.-May	June-Aug.	Sept.-Nov.
0°– 5°	136	159	193	166	173	133	197	197
5°–10°	209	178	146	144	222	210	194	185
10°–15°	241	204	184	165	244	251	202	193
15°–20°	250	184	195	198	260	243	235	214
20°–25°	226	146	160	199	302	229	210	246
25°–30°	256	165	124	256	327	208	157	251
30°–35°	284	168	103	242	329	167	100	245
35°–40°	437	175	94	293	374	178	74	274
40°–45°	349	109	33	217	258	97	29	179
45°–50°	287	94	−13	142	149	68	12	124
50°–55°	355	122	31	171	156	86	19	143
55°–60°	381	109	15	181	179	74	27	108

The author [6] has previously pointed out that during winter the locations of the principal Northern Hemisphere frontal zones as given by Petterssen [20] appear to correspond quite closely to the zones of maximum Q_a; in every case the zones of frontogenesis lie to the northwest of a zone of maximum Q_a with the major axes of the zones of maximum energy exchange parallel to the principal axes of the frontal zones.

RATE OF TOTAL HEAT GAIN BY THE ATMOSPHERE THROUGH CONVECTION

The quantity Q_a discussed in the preceding section represents the rate of total energy loss (or gain) from the ocean surface by conduction to the atmosphere. From the standpoint of the sea surface, this loss is entirely in the form of sensible heat; but in the case of the atmosphere only that fraction which appears as Q_h is immediately available for heating the air, the remaining energy being latent in the form of water vapor. Therefore, in order to obtain a quantity for the atmosphere which is equivalent to the quantity Q_a for the oceans, it is necessary to consider the heat supplied to the atmosphere through the condensation of the water vapor. The writer has previously shown that the latter quantity can be obtained, completely, from data concerning precipitation amounts over the oceans [9].

The sensible heat equivalent (calories per square centimeter per unit time) of the precipitated water is given by

$$Q_p = L_t P, \tag{16}$$

where L_t is the latent heat of condensation which, in this case, is correctly assumed to take place at the evaporation temperature t, and P is the rate of precipitation. One source of error which may be locally important during the winter season is introduced through the assumption in (16) that all precipitation is in the form of liquid water. However, even if it is assumed that *all* of the water precipitated at the higher latitudes occurs in the form of unmelted snow or sleet, the total error cannot exceed about 12 per cent. Actually, only a fraction of the precipitation occurring over the oceans during winter will appear as snow, even at high latitudes, so the probable error is much less.

Therefore, in order to complete the energy computations for the atmosphere it is necessary to have seasonal data concerning precipitation amounts over the oceans. However, an examination of the literature reveals not only that no seasonal data on precipitation amounts over the oceans exist, but that even the few data on the annual amounts have been subject to considerable criticism. In order to obtain approximately correct annual precipitation values, the published regional data [14, 27, 28, 29, 30] were made compatible and brought into agreement with Wüst's more reliable zonal values [39] by the assignment of simple correction factors for each

of the oceans.[6] Upon this basis an entirely new chart of the annual precipitation over all oceans has been prepared.

Then, through the use of published climatological data on the seasonal frequencies of precipitation over all oceans [38], seasonal precipitation values were computed for each ten-degree square through the use of the expression

$$P_s = \frac{F_p P_a}{\sum_{i=1}^{4} F_{pi}}, \tag{17}$$

where P_s is the seasonal precipitation rate (g cm^{-2} season^{-1}), P_a is the annual precipitation, and F_p is the seasonal frequency of Greenwich Meridian noon observations showing precipitation of any type. The complete series of seasonal (and annual) charts showing precipitation amounts over the oceans, together with details concerning the justification and verification of the method used, is contained in the publication previously referred to [10].

Since the precipitation rates over the oceans are given in centimeters per season (equivalent to grams per square centimeter per season) and since it was desired to have the energy equivalent Q_p given in calories per square centimeter per day, for purposes of convenience equation (16) was put in the form

$$Q_p = n^{-1} L_t P_s, \tag{18}$$

where n is the number of days in the season. Thus, where n is 91 and L_t is 585 cal g^{-1}, (18) becomes

$$Q_p = 6.43\ P_s \text{ cal cm}^{-2} \text{ day}^{-1}. \tag{19}$$

On the basis of formula (18) and the precipitation data arrived at through the use of (17), computations of the seasonal (and annual) values of Q_p have been made for each ten-degree square over the oceans. The seasonal averages arranged by latitude zones are given in Table VI. These data show that, in general, Q_p in the higher latitudes is greatest during the winter season in each hemisphere, while in the lower latitudes it is greatest during summer. In higher latitudes the minima occur during the corresponding summer seasons, and in the lower latitudes during the spring. With respect to time, therefore, the periods of maximum and minimum Q_p in the two hemispheres are exactly out of phase, that is, the zonal maxima in one hemisphere tend to occur during the months when the corresponding zonal minima occur in the opposite hemisphere. This result is hardly unexpected.

More interesting is a comparison between the corre-

Table VI. Seasonal Values of Q_p in Different Latitude Zones (in cal cm^{-2} day^{-1})

Latitude zone	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.
	Atlantic Ocean*				Pacific Ocean†			
50°–60°N	149	115	108	139	102	113	129	165
40°–50°N	180	134	113	145	146	137	130	159
30°–40°N	149	115	67	105	157	131	99	102
20°–30°N	90	62	68	100	115	94	104	98
10°–20°N	90	54	108	102	98	95	162	171
0°–10°N	227	237	280	234	248	236	274	255
0°–10°S	68	101	43	40	155	112	122	82
10°–20°S	34	30	50	30	165	116	123	125
20°–30°S	61	70	57	64	115	123	122	100
30°–40°S	74	99	120	99	70	99	129	95
40°–50°S	100	124	139	116	109	130	135	113
50°–60°S	133	139	125	102	100	121	89	102
	Indian Ocean				Mediterranean Sea Area‡			
40°–45°N	—	—	—	—	108	52	24	83
30°–40°N	—	—	—	—	95	73	32	86
20°–30°N	39	20	178	80	—	—	—	—
10°–20°N	60	69	274	192	—	—	—	—
0°–10°N	164	149	206	226	—	—	—	—
0°–10°S	266	219	241	256	—	—	—	—
10°–20°S	195	177	163	120	—	—	—	—
20°–30°S	82	106	100	50	—	—	—	—
30°–40°S	68	96	132	91	—	—	—	—
40°–50°S	117	132	144	123	—	—	—	—

* Includes the North Sea and ocean areas east of 10°W— areas which are not included in computations for other Q values
† Includes sea areas from 120°E to 100°E— areas which are not included in the computations for other Q values.
‡ Includes the Adriatic Sea.

6. The factors to be applied to the regional data (charts) prove to be 0.75 for the Atlantic Ocean, 0.70 for the Indian Ocean, and 0.55 for the Pacific Ocean.

sponding zonal values in each hemisphere. It may be noted that, except in the Indian Ocean, there exists no zone of high Q_p values within the Southern Hemisphere to correspond to the equatorial zone of maximum Q_p found in the Northern Hemisphere. This results in an approximate ten-degree poleward displacement of the analogous Q_p zones in the North Atlantic and North Pacific compared to the South Atlantic and South Pacific. In the case of the Indian Ocean, the situation is reversed. In this case the equatorial region of high Q_p values exists south of the equator. Also, the northern part of the Indian Ocean exhibits greater seasonal variations in Q_p than does any other portion of the oceans.

Of greater interest from an atmospheric energy point of view, however, are the data concerning the rate at which total convective energy is acquired by the atmosphere in the Northern Hemisphere, both through the condensation of water vapor and through the exchange of sensible heat with the sea surface. The total energy Q_{ph} is quite obviously the sum of these two quantities, that is,

$$Q_{ph} = Q_p + Q_h \text{ cal cm}^{-2} \text{ day}^{-1}. \quad (20)$$

The annual chart of Q_{ph} is given as Fig. 4 and the seasonal zonal averages are given in Table VII.

The results of the computations indicate that the principal sources of heat energy for maintaining the general circulation over the oceans are in the equatorial regions and at the higher latitudes. The subtropical latitudes appear as regions where very little heat energy is being supplied to the atmosphere. In these latter regions the very large rate of energy transfer which is indicated by the Q_a data does not make itself felt until it is released in the zone of equatorial convergence (or along the polar front at higher latitudes). The principal sources of heat energy for maintaining the circulation

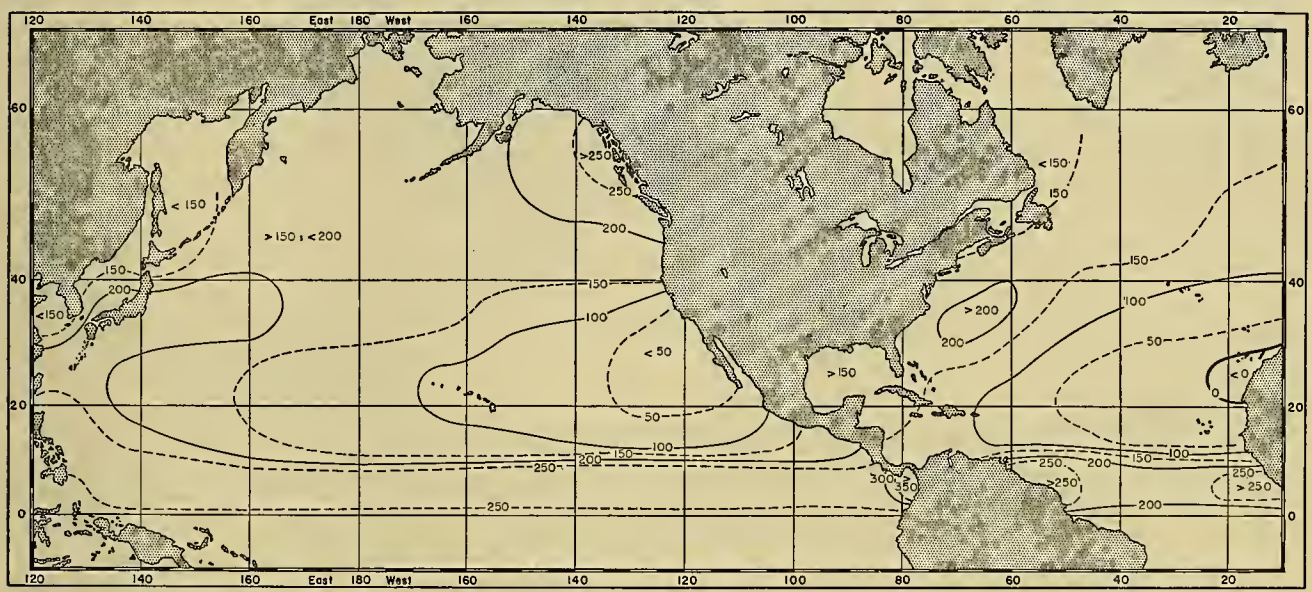


FIG. 4.—The annual values of the rate of total energy gain by the atmosphere Q_{ph}, both through the exchange of sensible heat with the sea surface Q_h and through condensation Q_p. The results are expressed in calories per square centimeter per day.

TABLE VII. SEASONAL VALUES OF Q_{ph} IN DIFFERENT LATITUDE ZONES (in cal cm^{-2} day^{-1})

North latitude zone	North Atlantic				North Pacific			
	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.
0°–10°	237	229	330	217	268	249	292	261
10°–20°	85	38	93	97	109	108	171	177
20°–30°	109	55	62	112	157	110	109	104
30°–40°	213	120	52	129	238	151	100	137
40°–50°	268	140	87	164	218	149	119	184
50°–60°	265	135	83	189	246	193	130	202

at the higher latitudes are probably the western portion of the oceans (particularly in winter) and the central and eastern portions of the oceans, poleward from latitude 40°N.

The existence of the seasonal and annual Q_a and Q_{ph} data for the North Atlantic and North Pacific allows the determination of the important quantity represented by their difference. Because

$$Q_a - Q_{ph} = (L_t E + Q_h) - (L_t P + Q_h) = L_t (E - P), \quad (21)$$

the difference $(Q_a - Q_{ph})$, when positive, represents latent energy which is locally surplus and available for transport to the continents or other portions of the oceans where it is made available as real heat through

condensation. When the quantity is negative, it indicates that an excess of latent energy is being locally transformed into sensible heat. However, since the author has already computed the seasonal and annual differences between evaporation and precipitation over the oceans [10], the difference $(Q_a - Q_{ph})$ can be determined very simply through use of these data and by applying the expression

$$Q_a - Q_{ph} = 6.43\ (E_s - P_s)\ \text{cal cm}^{-2}\ \text{day}^{-1}, \quad (22)$$

when $n = 91$ and $L_t = 585$, and E_s and P_s are the seasonal values of E and P expressed in grams per square centimeter per season.

The annual values of $(Q_a - Q_{ph})$ are given in Fig. 5 and the seasonal averages for latitude zones are given in Table VIII. These data show that the surplus latent energy is derived almost entirely from the ocean areas between 10°N and 40°N. The values are greatest in nearly every latitude zone during winter; during summer the quantity $(Q_a - Q_{ph})$, averaged for the entire North Atlantic and North Pacific, is negative, indicating that the Northern Hemisphere oceans are not sources of water vapor as surplus latent energy during this season.

Albrecht [1] has prepared a rough annual chart, showing $(E - P)$ over the oceans, which is based upon a comparison between observed surface salinities and those computed by means of Wüst's formulas [39] which give surface salinities as a function of $(E - P)$. A comparison between Albrecht's chart and Fig. 5 shows that the two agree in the gross features. However, the centers of Albrecht's areas of maximum $(E - P)$ are displaced from the positions indicated on Fig. 5. This displacement would be expected on the basis of the horizontal oceanic circulation which tends to shift the areas of maximum salinity in the direction of surface flow away from the centers of maximum $(E - P)$. Wüst's formulas apply regionally only in the absence of significant horizontal transport. Therefore, any $(E - P)$ chart based upon an examination of the distribution of surface salinities will show important regional departures from the true distribution [9].

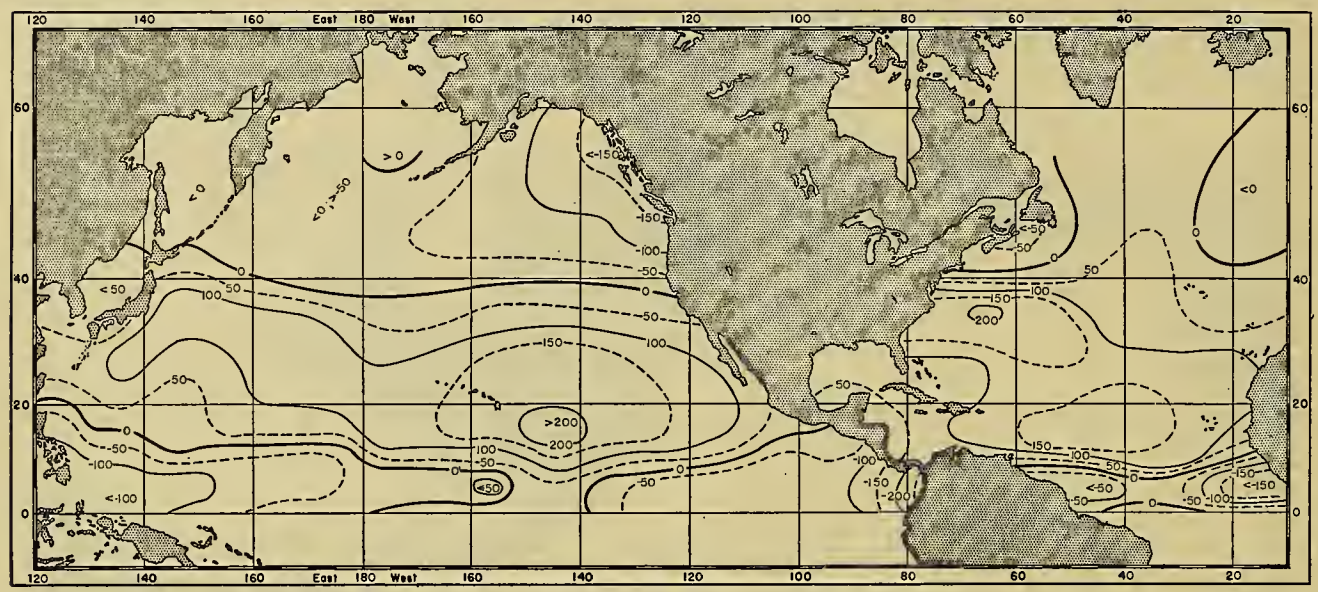


FIG. 5.—The annual values of the surplus of energy available in the latent form of water vapor $[Q_a - Q_{ph} = L_t\ (E - P)]$, expressed in calories per square centimeter per day. (The values above can be converted into rough $E - P$ rates by considering that the isometric interval of 50 cal cm^{-2} day^{-1} is approximately equivalent to an excess rate of evaporation over precipitation of 12 in. yr^{-1}. Negative values indicate a corresponding excess of precipitation over evaporation.)

TABLE VIII. SEASONAL VALUES OF $Q_a - Q_{ph}$ IN DIFFERENT LATITUDE ZONES (in cal cm^{-2} day^{-1}) WHERE $Q_a - Q_{ph} = L_t\ (E - P)$

North latitude zone	North Atlantic				North Pacific			
	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.	Dec.–Feb.	Mar.–May	June–Aug.	Sept.–Nov.
0°–10°	−62	−57	−126	−64	−34	−70	−99	−78
10°–20°	162	155	95	84	145	141	48	25
20°–30°	133	103	85	114	155	111	75	143
30°–40°	162	68	36	142	114	19	−10	125
40°–50°	57	−17	−72	24	−3	−66	−96	−28
50°–60°	78	−21	−74	−19	−13	−118	−89	−79

CONCLUSIONS

It should be emphasized that the mean seasonal and annual distributions of the several ocean-atmosphere energy components which have been discussed here are by no means representative of instantaneous conditions over the oceans. It is to be expected that large nonperiodic variations in the seasonal (and perhaps annual) Q rates will occur and that the areas of maximum and minimum values will vary in both intensity and position. It is quite probable, also, that the *relative* magnitudes of the various Q values will exhibit important regional variations of a nonperiodic nature. The daily synoptic patterns of each energy factor should be as complex and individualistic as are the synoptic patterns of other meteorological elements such as pressure and winds. If the tracks of pressure centers or the zonal circulations show a progressive shifting from decade to decade, as has been pointed out by Petterssen [21], then corresponding shifts in the normal patterns of the several Q values should be expected. If secular variations in the amount of solar energy received at the surface of the earth occur, it is to be expected that similar secular variations in Q_h, Q_e, and Q_p will occur. However, it is not suggested that such variations must be proportional, since the effects of increased humidity, cloudiness, etc., may be important.

For the purpose of filling the gaps in our knowledge of the interrelations between oceanic energy transformation processes and fluctuations in the general circulations of the atmosphere and oceans, the short-term and nonperiodic aspects of the exchange of energy between ocean and atmosphere should be investigated. Furthermore, from the climatic point of view, the seasonal and annual analyses of the exchange of sensible heat and water vapor between sea and atmosphere should be extended to the Southern Hemisphere as soon as the necessary humidity data become available.[7]

In viewing the major unsolved problems associated with the transformation of energy over the oceans it should be pointed out that a consideration of the convective transfer of energy between ocean and atmosphere does not, itself, complete the analysis of the ocean-atmosphere energy cycle. The quantity Q_a discussed on pp. 1065 ff. does not include the large fraction of stored heat that the sea surface loses through direct radiation to space nor does it take into account the smaller fraction of heat that is locally made available to (or removed from) the water mass through the dissipation (or creation) of the kinetic energy represented by current, wave, and tidal motions.

The quantity Q_{ph} discussed in the last section does not include the amount of heat the moist atmosphere receives through the absorption of long-wave sea-surface radiation or through the small absorption of direct solar energy. Neither does Q_{ph} include the regionally important fractions of heat energy locally made available to the atmosphere (or removed) through the dissipation (creation) of kinetic energy of atmospheric motion or the "layer heating" resulting from the reduction of gravitational potential energy through subsidence (or the "layer cooling" resulting from the creation of gravitational potential energy through lifting).[8] The end result, of course, is that the total energy acquired by the whole atmosphere through conduction, condensation, and radiation must exactly balance the radiative loss.

It is the author's opinion that a nonexhaustive list of the more important phases of the ocean-atmosphere energy relationships that need investigation should include:

1. The determination of the nonperiodic and short-term variations in the energy exchange between ocean and atmosphere.
2. The analysis of the seasonal and regional aspects of the radiative transfer of energy between sea surface and atmosphere.
3. The determination of the surplus of solar energy stored in the oceans and transported by ocean currents.
4. The analysis of the large-scale aspects of the transport of internal energy by the atmosphere over the oceans.
5. The analysis of the large-scale aspects of the interconversion of heat and kinetic and gravitational potential energy of atmosphere and oceans.

7. Száva-Kováts [37] has presented charts, which have been rather widely published in climatological texts, showing the January and Jnly distributions of relative humidity and vapor pressure over all oceans. However, an examination of the original paper reveals that the values have been obtained on the basis of simple assumptions as to the average relationship between sea-surface temperatures and the humidities at a standard height in the overlying atmosphere. Obviously, snch data are not suitable for the evaporation computations where no such assumptions are allowable.

8. The author has considered the latter aspects in some detail elsewhere [8].

REFERENCES

1. Albrecht, F., "Die Aktionsgebiete des Wasser- und Wärmehaushaltes der Erdoberfläche." *Z. Meteor.*, 1: 97–109 (1947).
2. Ångström, A., "Applications of Heat Radiation Measurements to the Problems of Evaporation from Lakes and the Heat Convection at Their Surfaces." *Geogr. Ann., Stockh.*, 2: 237–252 (1920).
3. Bowen, I. S., "The Ratio of Heat Losses by Conduction and by Evaporation from any Water Surface." *Phys. Rev.*, 27: 779–787 (1926).
4. Burke, C. J., "Transformation of Polar Continental Air to Polar Maritime Air." *J. Meteor.*, 2: 94–112 (1945).
5. Cummings, N. W., and Richardson, B., "Evaporation from Lakes." *Phys. Rev.*, 30: 527–534 (1927).
6. Jacobs, W. C., "On the Energy Exchange Between Sea and Atmosphere." *J. mar. Res.*, 5: 37–66 (1942). (Contains charts of winter and summer values for Q_a over the North Pacific and North Atlantic.)
7. —— "Sources of Atmospheric Heat and Moisture over the North Pacific and North Atlantic Oceans." *Ann. N. Y. Acad. Sci.*, 44: 19–40 (1943). (Contains charts of winter and summer values of Q_h and E over the North Pacific and North Atlantic.)
8. —— "The Energy Acquired by the Atmosphere over the

Oceans through Condensation and through Heating from the Sea Surface." *J. Meteor.*, 6: 266–272 (1949).

9. —— "The Distribution and Some Effects of the Seasonal Quantities $E - P$ (Evaporation minus Precipitation) over the North Atlantic and North Pacific." *Arch. Meteor. Geophys. Biokl.*, (A) Teil II, 1: 1–16 (1950).
10. —— "The Energy Exchange between Sea and Atmosphere and Some of Its Consequences." *Bull. Scripps Instn. Oceanogr. tech.* (in press). (Contains complete set of seasonal charts of Q_h, Q_a, E, Q_{ph}, and $(E - P)$ for North Pacific and North Atlantic and seasonal charts of precipitation for all oceans.)
11. —— and Clarke, Katherine B., "Meteorological Results of Cruise VII of the *Carnegie*, 1928–1929." *Carneg. Instn. Wash. Publ.* 544, 168 pp., Washington, D. C. (1943). (See pp. 35–38)
12. Kuzmin, P. P., "On the Meteorological Conditions of Heat Exchange between Sea and Adjacent Layers of Air." *Meteorologia i Gidrologia*, 4: 3–13 (1946); Saito, Y., in *Umi to Sora* (Sea and Sky), 24: 326, Japan (1944).
13. McEwen, G. F., "Some Energy Relations between the Sea Surface and the Atmosphere." *J. mar. Res.*, 1: 217–238 (1938).
14. Meinardus, W., "Die Niederschlagsverteilung auf der Erde." (*Petermanns Mitt.*, 80: 1–4 (1934).) *Meteor. Z.*, 51: 345–350 (1934).
15. Miyazaki, M., "The Incoming and Outgoing Heat at the Sea Surface along the Tusima Warm Current." *Oceanogr. Mag., Tokyo*, 1: 103–111 (1949).
16. Montgomery, R. B., "Observations of Vertical Humidity Distribution above the Ocean Surface and Their Relation to Evaporation." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 4, 30 pp. (1940).
17. —— "Vertical Eddy Flux of Heat in the Atmosphere." *J. Meteor.*, 5: 265–274 (1948).
18. Mosby, H., "Verdunstung und Strahlung auf dem Meere." *Ann. Hydrogr., Berl.*, 64: 281–286 (1936).
19. Munk, W. H., "A Critical Wind Speed for Air-Sea Boundary Processes." *J. mar. Res.*, 6: 203–218 (1947).
20. Petterssen, S., *Weather Analysis and Forecasting.* New York, McGraw, 1940. (See pp. 269–271)
21. —— "Changes in the General Circulation Associated with Recent Climatic Variation," in *Glaciers and Climate. Geogr. Ann., Stockh.*, Vol. 31, Nos. 1–2, pp. 212–221 (1949).
22. Pogosyan, K. P., and Taborovsky, N. L., "The Transformation of Air Masses and the Dynamics of Atmospheric Processes." *Meteorologia i Gidrologia*, No. 4, pp. 43–50 (1940). (Translated by USAF, Air Weather Service.)
23. Ponomarenko, G., "Evaporation and Heat Exchange in the Sea of Okhotsk." *Meteorologia i Gidrologia*, No. 10, pp. 109–113 (1940). (Translated by USAF, Air Weather Service.)
24. Rossby, C.-G., "A Generalization of the Theory of the Mixing Length with Applications to Atmospheric and Oceanic Turbulence." *Meteor. Pap., Camb., Mass.*, Vol. 1, No. 4, 36 pp. (1932).
25. —— and Montgomery, R. B., "The Layer of Frictional Influence in Wind and Ocean Currents." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 3, No. 3, 101 pp. (1935).
26. Schmidt, W., "Strahlung und Verdunstung an freien Wasserflächen." *Ann. Hydrogr., Berl.*, 43: 111–124, 169–178 (1915).
27. Schott, G., *Geographie des Atlantischen Ozeans*, 2. Aufl. Hamburg, C. Boysen, 1926.
28. —— "Die jährlichen Niederschlagsmengen auf dem Indischen und Stillen Ozean." *Ann. Hydrogr., Berl.*, 61: 1–12 (1933).
29. —— *Geographie des Indischen und Stillen Ozeans.* Hamburg, C. Boysen, 1935.
30. Soviet World Atlas, The Great, *Maps of the World*, Vol. 1, Pt. I. Moscow, 1937.
31. Sverdrup, H. U., "Das maritime Verdunstungsproblem." *Ann. Hydrogr., Berl.*, 64: 41–47 (1936).
32. —— "On the Evaporation from the Oceans." *J. mar. Res.*, 1: 3–14 (1937).
33. —— "On the Annual and Diurnal Variation of the Evaporation from the Oceans." *J. mar. Res.*, 3: 93–104 (1940).
34. —— *Oceanography for Meteorologists.* New York, Prentice-Hall, Inc., 1942. (See pp. 61–70)
35. —— "On the Ratio Between Heat Conduction from the Sea Surface and Heat Used for Evaporation." *Ann. N. Y. Acad. Sci.*, 44: 81–88 (1943).
36. —— "The Humidity Gradient over the Sea Surface." *J. Meteor.*, 3: 1–8 (1946).
37. Száva-Kováts, J., "Verteilung der Luftfeuchtigkeit auf der Erde." *Ann. Hydrogr., Berl.*, 66: 373–378 (1938). (Contains February and July charts of vapor pressure and relative humidity for all oceans.)
38. U. S. Weather Bureau, *Atlas of Climatic Charts of the Oceans.* W. B. Publ. No. 1247, Washington, D. C., 1938.
39. Wüst, G., "Oberflächensalzgehalt, Verdunstung, und Niederschlag auf dem Weltmeere." *Länderkundliche Forschung*, Festschrift Norbert Krebs, SS. 347–359 (1936).

SUPPLEMENTARY REFERENCES

Papers received subsequent to the preparation of the preceding article.

40. Albrecht, F., "Über die Wärme- und Wasserbilanz der Erde." *Ann. Meteor.*, 2: 129–143 (1949). (See also 11 insert charts)
41. Model, F., *Berechnungsmethode für die Übertragung von Wärme durch Wassermassen (Strömungen) unter verschiedenen physikalischen Bedingungen*, 96 SS. Deutsches Hydrographisches Institut, Hamburg, 1949.

EVAPORATION FROM THE OCEANS

By H. U. SVERDRUP

Norwegian Polar Institute, Oslo

INTRODUCTION

The problem of evaporation from the oceans has received little attention compared to that directed towards the study of evaporation from lakes and land surfaces. One of the reasons for this is that the results of the latter, being of great interest to the hydrological engineers, are of immediate economic importance. Furthermore, until recently, weather forecasting has been concerned primarily with forecasting over land areas, for which purpose it has been sufficient to know that air which has travelled over the oceans contains a great deal of moisture, but it has not been necessary to ask how this moisture has been added to the air. With the rapid development of air routes across the oceans, accurate weather forecasts for ocean areas are needed, and in this connection the question of how air masses are modified when passing over water becomes important. When dealing with this modification, knowledge of evaporation as a function of the meteorological conditions near the sea surface becomes indispensable.

This development is reflected in the very manner in which the question of evaporation from the oceans has been dealt with during the last seventy-five years. Up to about 1930, most efforts were directed towards establishing the annual evaporation in different latitudes in order to examine the general water budget of the atmosphere, but during the last two decades the emphasis has been placed on examinations of the evaporation under given meteorological conditions. Some advances have been made, but these studies are still in their infancy.

EVAPORATION FROM THE OCEANS DETERMINED BY EXTRAPOLATION OF VALUES FROM LAND

In his discussion of the water budget of the atmosphere, Brückner [3] tried to find the evaporation from the oceans by extrapolating values derived from observations on land. He used observations from coastal regions particularly and obtained an average evaporation from all oceans between 70°N and 70°S of 110 ± 15 cm per year. He suggests that this value may be somewhat high because he has probably assumed too large an evaporation between 10°N and 10°S (see Table I). It is of interest to observe that, besides other material, Brückner made use of Russell's map of 1888 of the evaporation in the United States, from which he obtained the following values at the Atlantic and Pacific coasts:

North Latitude (deg)	Evaporation (cm per year)	Evaporation (mm per 24 hr)
25–30	120	3.3
30–40	98	2.7
40–45	82	2.2
45–50	51	1.4

A comparison with the evaporation data in Table I shows that these values are in remarkable agreement with results that have been obtained by entirely different methods.

DIRECT MEASUREMENTS OF EVAPORATION AT SEA

History. Measurements of evaporation from pans have been carried out on board ship in limited number. The first observations of this type were made by Mohn [15] during the Norwegian North Atlantic Expedition, 1876–78. Mohn used an evaporation pan that floated in a large container and determined the evaporation in 12 hours by adding enough fresh water to the pan, at the end of the time interval, to bring the weight of pan and water back to its original amount.

In subsequent experiments the evaporation has been determined by means of the increase in salinity during the period of exposure; in the earlier ones the salinity was determined by aerometric measurements of density and in the later ones by chlorine titration. In several cases pans have been exposed simultaneously to wind and sunshine, to sunshine only, and to wind only, in order to examine the effects of exposure. However, the results, which have been extensively discussed, have all been derived from observations obtained with the pan placed in as unprotected a position as possible.

Nearly all the measurements referred to in the last paragraph were made in German investigations during the years 1892 to 1914, and most of them between 1908 and 1914. They have been summarized and discussed by Wüst [29], who gives a list of the pertinent literature. Several measurements [10] were made during the last cruise of the *Carnegie*, but it is doubtful if such observations will be continued in the future because the results are difficult to interpret and because other methods of approach have met with fair success.

Discussion. Wüst [29] showed that the evaporation from pans on board ship depended principally on two variables, the wind velocity w, and the evaporation potential defined as

$$P = (T/273)(e_s - e). \tag{1}$$

In (1), T represents the absolute temperature of the water in the pan, e_s the vapor pressure at the water surface, and e the vapor pressure in the air measured on board ship. The wind velocity, w, is the wind velocity measured or estimated on board ship.

Regarding the vapor pressure at the water surface, it should be observed that the vapor pressure over sea water is somewhat lower than that over distilled water. If the latter is represented by e_d,

$$e_s = e_d\,(1 - 0.00053S), \tag{2}$$

where S is the salinity in parts per thousand. Over the oceans the salinity differs so little from 35 per mille that with sufficient accuracy $e_s = 0.98e_d$.

The observed evaporation values could be represented with good approximation by the equation,

$$E = 0.40\,P\,(1 + 0.40w) \qquad (3)$$

where E represents evaporation in centimeters per 24 hours, and e, which enters into the term P, is measured in millibars, and w in meters per second.

reduction factor. Somewhat arbitrarily he selected a height of 0.2 m above the sea surface, calculated what the evaporation from a pan at that height should have been, and assumed that a pan placed 0.2 m above the sea surface would have given values that would represent the evaporation from the sea surface itself. The numerical values are based on only a few series of measurements of wind and humidity gradients carried out at low wind velocity and on assumptions, the validity of which is not obvious. It is therefore not

Table I. Energy Available for Evaporation in Areas Between Parallels of Latitude, and Average Evaporation Rates Determined by Different Methods

Latitude range (deg)	Ocean area (10^{16} cm^2)	$(Q_s - Q_r)$* (cal cm^{-2} day^{-1})	R†	Q_e‡ (cal cm^{-2} day^{-1})	L§ (cal)	Evaporation (mm day^{-1})			
						Extrapolation (Brückner)	Obs. at sea (Wüst)	From Q_e	From met. obs. (Jacobs)
70°–60°N	5.6	45	0.45	31	593	0.55	0.35	0.52	...
60°–50°	10.9	80	0.31	61	592	1.10	1.18	1.03	1.78
50°–40°	15.0	113	0.21	93	590	1.80	2.11	1.58	2.00
40°–30°	20.8	164	0.15	143	586	2.75	2.90	2.44	3.32
30°–20°	25.1	239	0.11	215	583	3.55	3.50	3.69	3.56
20°–10°	31.5	253	0.10	230	582	4.10	3.61	3.95	3.66
10°–0°	34.0	239	0.10	217	582	4.35	3.20	3.73	3.11
0°–10°S	33.7	218	0.10	198	582	4.35	3.36	3.40	...
10°–20°	33.4	228	0.10	207	583	4.10	3.60	3.55	...
20°–30°	30.9	229	0.11	206	585	3.55	3.36	3.52	...
30°–40°	32.3	191	0.14	168	587	2.75	2.69	2.86	...
40°–50°	30.5	117	0.17	100	591	1.80	1.76	1.69	...
50°–60°	25.4	67	0.20	56	594	1.10	0.71	0.94	...
60°–70°	17.1	35	0.23	28	597	0.55	0.22	0.47	...

* Radiation surplus according to W. Schmidt (see p. 1073).

† Bowen ratio, equation (8). Adjusted values from Jacobs (see p. 1074).

‡ $Q_e = (Q_s - Q_r)/(1 + R)$.

§ Latent heat of vaporization, taking average sea-surface temperature into account. ($L = 596 - 0.56\vartheta$).

In discussing the accuracy of the evaporation values observed on board ship, Wüst [29] arrived at the conclusion that the values represent the actual evaporation *from the pan* with an accuracy better than ±8 per cent. The pertinent question is, therefore, whether these measurements permit any conclusions as to the evaporation *from the sea surface*. This question was first dealt with by Schmidt [21], who showed that the evaporation from the sea surface was probably only half that from the pans. Later on, the matter was discussed in great detail by Wüst [29, 30] and by Cherubim [5].

In attempting to reduce the pan values to values of evaporation from the sea surface, Wüst points out that: (1) the temperature of the sea surface deviates from that at the surface of the water in the pan, (2) the vapor pressure of the air at a short distance from the sea surface differs from that observed on board ship, and (3) similarly the wind velocity at a short distance from the sea surface differs from the wind velocity recorded on board ship. By means of special observations carried out in the Baltic in 1919 Wüst established the variation of the vapor pressure and wind with height above the sea surface and used the values thus obtained for a computation of a combined

surprising that somewhat different approaches give different results. In 1920 Wüst found a factor of 0.46, in 1931 Cherubim obtained a value of 0.58, and in 1936 Wüst finally adopted the value of 0.53, which, as will be shown, gives evaporation values that agree remarkably well with those obtained on the basis of energy considerations.

Results. From the evaporation measurements which have been carried out on board ship along different routes and by means of climatological data from which the evaporation on board ship can be computed using equation (3), Wüst derived average values of the annual evaporation from all oceans between 70°N and 70°S. These, based on a reduction factor of 0.53, are entered in Table I. Wüst estimates the probable error in his values to be ±12 per cent. The average value is 96 ± 12 cm per year.

Wüst claims only that the *average* values observed on board ship can be reduced to values of evaporation from the sea surface. He does not claim that the relationship between evaporation, evaporation potential (equation (3)), and wind velocity, which was established for the pan measurements, is applicable to the evaporation from the sea surface after multiplication by the factor 0.53. It does not appear probable that the equation

applies, because it gives high evaporation rates at very low wind velocities and, by extrapolation, an evaporation of 0.21 mm per 24 hr in absolutely calm weather. When one is dealing with a pan, the evaporation may indeed be considerable even at very low wind velocities, but the evaporation from the sea surface under this condition must be negligible. The question as to the relation between the evaporation and the meteorological conditions at sea will be discussed later.

EVAPORATION FROM THE OCEANS COMPUTED FROM ENERGY CONSIDERATIONS

Energy available for evaporation. The first attempt to determine the evaporation from the oceans on the basis of energy considerations was made by Schmidt [20]. In principle the procedure is quite simple. During the year the oceans gain energy by short-wave radiation from the sun and the sky (Q_s) and lose energy by effective long-wave radiation to the atmosphere (Q_r), by evaporation (Q_e), and by conduction of heat to the atmosphere (Q_h). Other sources of energy gains or losses, such as conduction of heat from the interior of the earth, energy changes related to chemical and biochemical processes in the sea, and friction losses, are negligible compared to the former. Furthermore, it can be assumed that the average temperature of the oceans remains so nearly constant from one year to another that, to a close approximation, the average energy gain must equal the loss:

$$Q_s = Q_r + Q_e + Q_h. \tag{4}$$

Introducing

$$R = Q_h/Q_e \tag{5}$$

and

$$E = Q_e/L, \tag{6}$$

where L is the latent heat of vaporization, we obtain

$$E = \frac{Q_s - Q_r}{L\,(1 + R)}. \tag{7}$$

Schmidt determined the values of the radiation surplus ($Q_s - Q_r$), using measurements of radiation at sea combined with climatological data as to cloudiness and sea-surface temperature (see Fig. 1).

Mosby [17] undertook a new computation of the radiation surplus, using certain empirical relations between the incoming radiation and the altitude of the sun. On the whole his values agree very well with those of Schmidt (see Fig. 1). The major discrepancy is found between latitudes 20°N and 20°S where Schmidt's results show a minimum of radiation surplus near the equator, whereas Mosby obtains no such minimum. The reason for this is that Schmidt has introduced a considerably greater cloudiness at the equator than at 30°N or 30°S (5.9, 4.2, and 4.0, respectively), whereas Mosby has used nearly the same values of cloudiness in all latitudes between 30°N and 30°S (values between 5.6 and 5.2). It seems probable that Schmidt's values give a more correct representation of the variation of radiation surplus with latitude.

McEwen [13] has computed the radiation surplus between latitudes 20° and 50°N in the eastern North Pacific, using a very elaborate method. His results are in very good agreement with those of Schmidt and Mosby, as is evident from Fig. 1.

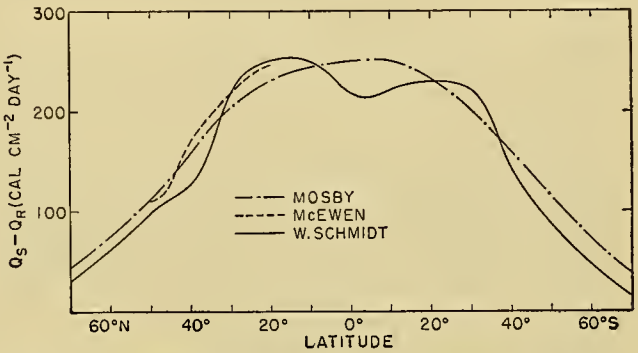


FIG. 1.—Average annual surplus of energy (in cal cm^{-2} min^{-1}) which the oceans receive in different latitudes by radiation processes.

The Bowen Ratio. From the preceding discussion and the results shown in Fig. 1 it appears that the radiation surplus received by the oceans has been determined with considerable accuracy. For computing the evaporation it is in addition necessary to know the ratio $R = Q_h/Q_s$, which is often referred to as the *Bowen ratio* because Bowen [2] has shown that it can readily be computed from meteorological observations on board ship if, in addition, the sea surface temperature has been recorded.

When computing the evaporation, Schmidt [20] did not introduce the ratio R, but a ratio $R' = Q_s/(Q_s - Q_r)$ from which R can be found from the relation $R = (1 - R')/R'$. Schmidt had no means of determining R' directly, but concluded from certain general considerations that it varied greatly with latitude. Expressed as R, the range is from a value of about 0.28 at the equator to a value of about 1.67 at 70°N or S.

Ångström [1] criticised Schmidt's determination of R and pointed out that in middle latitudes Schmidt's values were far too large. From the application of energy considerations to Swedish lakes and from measurements of actual evaporation and of meteorological conditions Ångström concluded that the ratio R was only about 0.1, meaning that, of the net energy received by radiation ($Q_s - Q_r$), only about 10 per cent was given off to the atmosphere by conduction and about 90 per cent was used for evaporation.

Mosby [17] accepted Ångström's estimate of R and used a value of 0.1 for the computation of the average evaporation from all oceans. He was not aware of Bowen's paper, according to which R can be computed if the air temperature ϑ_a and the vapor pressure e_a have been observed at the same height a above the sea surface, and if the temperature at the sea surface ϑ_s has been recorded, from which the vapor pressure at the sea surface is obtained using equation (2). If we antici-

pate part of the following discussion (p. 1075, eq. (12)), the Bowen formula can be derived in a simple manner. If the eddy coefficients for diffusion of water vapor and conduction of heat are assumed to be identical, the upward fluxes of latent energy of water vapor and of heat can be written

$$Q_e = -L \frac{0.621}{p} A \frac{de}{dz}$$

and

$$Q_h = -c_p A \frac{d\vartheta}{dz},$$

where p is the atmospheric pressure, A the eddy conductivity (or diffusivity), c_p the specific heat of the air, and ϑ the air temperature. To be exact, potential temperature should have been used, but near the sea surface, where the vertical temperature gradient is large, only a small error is introduced by using the ordinary temperature. It follows that

$$R = \frac{Q_h}{Q_e} \frac{c_p\, p}{0.621\, L} \frac{d\vartheta/dz}{de/dz}.$$

Replacing the ratio of the differentials by the ratio $(\vartheta_s - \vartheta_a)/(e_s - e_a)$ and introducing the numerical values $c_p = 0.240$, $p = 1000$ mb, and $L = 585$, one obtains R, the Bowen ratio:

$$R = 0.64 \frac{\vartheta_s - \vartheta_a}{e_s - e_a}. \qquad (8)$$

The Bowen ratio has been used extensively by Cummings and Richardson (see, for example, [6]) in the study of evaporation from lakes, and by Jacobs [8] in his examination of the energy exchange between the ocean and the atmosphere.

For the North Atlantic and the North Pacific Oceans, Jacobs determined the Bowen ratio as a function of latitude. In both cases R increased with latitude, but the values found for the North Atlantic were consistently smaller than those for the North Pacific. In the Atlantic even small negative values were found near the equator, indicating that there the average air temperature is slightly higher than the sea-surface temperature. The reality of this feature must be questioned. Careful observations from specially equipped ships have demonstrated that in the tropics of the Atlantic Ocean the air temperature is about 0.8C *lower* than the sea-surface temperature, but the values found on climatological charts show a smaller difference or even a higher air temperature. The reason for this is that the climatological charts are based on routine observations on board ship, in which errors due to the ship's heat or faulty exposure of the thermometer have not been eliminated. In his computations of R, Jacobs [8] used data from climatological charts and, therefore, his R values are probably too small in low latitudes, but in higher latitudes they are more nearly correct.

Taking the average for the two northern oceans one obtains:

North Latitude (deg)	R (Jacobs)	R (Corrected)	R (in intervals of latitude)
70	0.53	0.53	
			0.45
60	0.37	0.37	
			0.31
50	0.25	0.25	
			0.21
40	0.18	0.18	
			0.15
30	0.08	0.13	
			0.11
20	0.02	0.10	
			0.10
10	0.00	0.10	
			0.10
0	0.00	0.10	

The increase with increasing north latitude is in part an effect of the continents from which cold air flows out over the oceans in winter. In the Southern Hemisphere the effect is absent or weak, and therefore we shall assume that R increases only to 0.25 at 70°S.

Results. In revising the computation of evaporation on the basis of energy considerations we shall make use of Schmidt's values of the radiation surplus and of Jacobs' values of R in the Northern Hemisphere after application of an estimated correction in the tropics. For the Southern Hemisphere we shall introduce estimated values. For the latent heat of vaporization, values will be used corresponding to the average sea-surface temperature in the different latitudes. All values used and the results obtained are shown in Table I, which also contains the evaporation observed on board ship reduced to true evaporation (p. 1072) and average values according to Jacobs.

If we take into consideration the ocean areas between parallels of latitude, the average annual evaporation between latitudes 70°N and 70°S is found to be 99 cm per year. The error of this value is probably less than ±10 cm per year.

EVAPORATION FROM THE OCEANS COMPUTED FROM METEOROLOGICAL ELEMENTS

History. In evaporation studies one of the aims has long been the establishment of relationships that would permit computation of evaporation from the meteorological elements recorded on board ship, that is, from air temperature, humidity, wind velocity, and barometric pressure, and from knowledge of the sea-surface temperature. Prior to about 1920 several empirical equations were developed on the basis of pan measurements on shipboard or on land (see for example equation (3)), but these equations give only the empirical relation between the evaporation *from the pan* and certain meteorological elements and they cannot be expected to give correct values of the evaporation from the sea surface. This point of view had not been generally recognized [7, 11] but had been emphasized by Schmidt [21], who strongly recommended the detailed examination of humidity, temperature, and wind conditions in the very lowest layers of the air directly above the sea surface in order to establish a basis for reducing the evaporation measurements on board ship to the sea surface.

Wüst [29] carried out a limited number of such measurements (see p. 1072), but used them only for a reduction of the *averages* and not for the establishment of a new relationship. Recent advances have been made by following the entirely different procedure

of computing evaporation directly from a knowledge of humidity and wind conditions in the lowest layer. A first attempt in this direction was made by Wagner [28], who had to confine himself to general considerations and the use of empirical relations established over land because the type of studies urged by Schmidt had not been made.

Even today there are available only a few series of measurements of temperature, humidity, and wind at various levels in the vicinity of the sea surface. The interpretation of these few observations has, however, helped to throw light on the problem of evaporation, but clear-cut results have not been obtained. It is necessary to review the manner in which the problem has been dealt with by different authors and to evaluate the empirical evidence that supports or contradicts the different conclusions.

Theoretical Considerations. When one is dealing with evaporation, the moisture content of the air is best described by the specific humidity q. The moisture concentration (water vapor per unit volume) is then equal to ρq, and this concentration will be considered a conservative property, that is, a property which is altered locally (except at the boundaries) by processes of diffusion and advection only:

$$\frac{\partial(\rho q)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{A_x}{\rho}\frac{\partial(\rho q)}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{A_y}{\rho}\frac{\partial(\rho q)}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{A_z}{\rho}\frac{\partial(\rho q)}{\partial z}\right) - \frac{\partial}{\partial x}(\rho q w_x) - \frac{\partial}{\partial y}(\rho q w_y) - \frac{\partial}{\partial z}(\rho q w_z)\,. \tag{9}$$

Here A_x/ρ, A_y/ρ, and A_z/ρ represent the diffusion coefficients (of dimensions $L^2\ T^{-1}$) which are supposed to be different in different directions, ρ is the density of the air, and w_x, w_y, and w_z are the velocity components. The definition implies that the following considerations are not valid if droplets are present in such numbers that condensation on or evaporation from these droplets cannot be neglected.

Equation (9) has been used in very simplified forms. If we assume stationary conditions ($\partial(\rho q)/\partial t = 0$), motion along the x-axis only ($w_y = w_z = 0$), and neglect horizontal diffusion ($A_x\,\partial(\rho q)/\partial x = A_y\,\partial(\rho q)/\partial y = 0$), equation (9) is reduced to

$$\frac{\partial}{\partial z}\left(\frac{A_z}{\rho}\frac{\partial(\rho q)}{\partial z}\right) = \frac{\partial(\rho q w_x)}{\partial x}\,. \tag{10}$$

In this form it has been applied by Sutton [22] to the problem of evaporation from a limited body of water. His theoretical conclusions are in good agreement with the results of experiments in the laboratory and conditions observed in the field. Other similar studies have been carried out and have led to the clarification of experimental results, but they are not applicable to the problem of evaporation from the ocean because the ocean surface must be considered as a water surface of infinite extension, directly above which the horizontal gradients of moisture content are negligible. Very near the sea surface the vertical velocity can be neglected in all circumstances and the density can be considered constant. Equation (9) is then reduced to

$$\frac{d}{dz}\left(A\frac{dq}{dz}\right) = 0$$

or

$$A\frac{dq}{dz} = \text{const}, \tag{11}$$

which simply expresses the fact that near the boundary surface the vertical flux of water vapor, expressed in units of mass per area and time, is independent of height. In the cgs system the flux is expressed in g cm^{-2} sec^{-1}.

If the vertical flux is directed upwards, dq/dz must be negative. In this case the flux must equal the evaporation from the water surface, and therefore

$$E = -A\frac{dq}{dz}\ (\text{g cm}^{-2}\ \text{sec}^{-1}). \tag{12}$$

Thus the problem of evaporation is reduced to the problem of finding the vertical flux of water vapor.

This may appear to be an easy matter, because in the air the eddy diffusivity A can be considered to be practically known from wind observations. However, difficulties arise because, very close to the surface, diffusivity and viscosity are *not* identical, and because different assumptions as to the character of the diffusivity close to the surface lead to widely different results.

The eddy viscosity increases with height and, furthermore, depends upon the stability of the stratification and the character of the sea surface, whether hydrodynamically smooth or rough. At indifferent stratification or slight instability, conditions which prevail over the sea, the eddy viscosity is a linear function of height.

In anticipation of material which follows, it can be stated that at some distance from the surface

$$A \approx \rho k_0 w_* z, \tag{13}$$

where k_0 and w_* are defined as below. Since $A\ dq/dz = \text{const}$, it follows that q must vary with the logarithm of height, and Montgomery [16] has therefore introduced an evaporation coefficient Γ, defined by

$$\Gamma \equiv -\frac{1}{q_s - q}\frac{dq}{d\ln z}\,. \tag{14}$$

Introducing (13) and (14) into (12) we obtain

$$E = k_0 \rho w_* \Gamma\,(q_s - q). \tag{15}$$

It is this expression which we shall attempt to evaluate for smooth and rough conditions; that is, we shall determine the evaporation coefficient Γ which is the unknown factor.

Over a *smooth* surface there exists, next to the surface, a thin boundary layer in which the flow is laminar and where ordinary viscosity acts. Above this laminar layer the turbulent layer begins. Placing the origin of the vertical axis at the top of the laminar layer, we have in

the turbulent layer

$$A = \rho(\nu + k_0 w_* z), \tag{16}$$

and in the laminar layer

$$A = \mu = \rho\nu. \tag{17}$$

Here μ is the viscosity and ν the kinematic viscosity of the air, k_0 is von Kármán's universal turbulence constant ($k_0 = 0.4$), and w_* is the "friction velocity" which is defined by the equation

$$w_* = \sqrt{\tau/\rho} \tag{18}$$

where τ is the shearing stress. This is defined as $\tau = A\, dw/dz$ and is supposed, near the sea surface, to be independent of height and equal to τ_0, the wind stress at the sea surface. The friction velocity can be expressed by the wind at any level w:

$$w_* = \gamma w, \tag{19}$$

where γ is the resistance coefficient which, according to von Kármán, can be obtained from the equation

$$\frac{1}{\gamma} + \frac{1}{k_0} \ln \frac{1}{\gamma} = 5.5 + \frac{1}{k_0} \ln \frac{w_* z}{\nu}. \tag{20}$$

The thickness of the laminar layer δ is supposed to depend on the viscosity and the stress as expressed by the friction velocity:

$$\delta = \lambda \frac{\nu}{w_*}, \tag{21}$$

where λ is a constant for which von Kármán found the value 11.5, whereas Montgomery [16] obtained 7.8. In the following applications we shall use von Kármán's value, $\lambda = 11.5$, but the choice is not of great importance.

The picture given above of the variation with height of the eddy viscosity appears fully applicable when turning to the eddy diffusivity. We must indeed expect that next to the sea surface there exists a layer through which the flux of water vapor takes place by molecular diffusion, and it seems reasonable that the thickness of this layer equals that of the laminar layer. In the turbulent layer the eddy diffusivity must then be

$$A' = \rho(\kappa + k_0 w_* z), \tag{22}$$

where κ is the coefficient of diffusion of water vapor through air. For values of $z >> \delta$ this eddy diffusivity differs only imperceptibly from the eddy viscosity.

The flux of water vapor through the two layers is then

$$E = -\rho\kappa \left(\frac{dq}{dz}\right)_l = -\rho\,(\kappa + k_0 w_* z) \left(\frac{dq}{dz}\right)_t, \tag{23}$$

where the indices l and t indicate that the differentials apply to the laminar and the turbulent layer, respectively. By integration, remembering that $z = 0$ at the top of the laminar layer where $q = q_\delta$:

$$q_s - q_\delta = \frac{E}{\rho\kappa}\delta = \frac{E}{\rho w_*}\frac{\lambda\nu}{\kappa}, \tag{24}$$

$$q_\delta - q = \frac{E}{k_0 \rho w_*} \ln \frac{\kappa + k_0 w_* z}{\kappa}. \tag{25}$$

Adding and considering that κ is small compared to $k_0 w_* z$ for $z >> \delta$, we obtain:

$$q_s - q = \frac{E}{k_0 \rho w_*}\left[\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{k_0 w_* z}{\kappa}\right], \tag{26}$$

or, introducing w,

$$E = k_0 \rho\gamma w \frac{q_s - q}{\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{k_0 w_* z}{\kappa}}. \tag{27}$$

Therefore:

$$\Gamma_a = \left[\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{k_0 w_* a}{\kappa}\right]^{-1}. \tag{28}$$

Here the ratio ν/κ is independent of temperature ($\nu/\kappa = 0.602$), but κ increases a little with temperature (at 0C $\kappa = 0.22$, at 20C $\kappa = 0.25$ cm^2 sec^{-1}). With $\kappa = 0.24$ cm^2 sec^{-1} and $a = 800$ cm, Γ_a is shown in Fig. 2 as a function of w.

According to Rossby [19], the sea surface has the character of a hydrodynamically *smooth* surface at wind velocities up to 6–7 m sec^{-1} as measured at a height of about 8 m. At wind velocities exceeding 7–8 m sec^{-1} the sea surface appears to be hydrodynamically rough.

Over a *rough* surface the eddy viscosity increases linearly with height and has, at the surface itself ($z = 0$), a value which is much larger than that of the ordinary viscosity:

$$A = k_0 \rho w_* (z + z_0). \tag{29}$$

Here z_0 is the "roughness length" of the surface. The resistance coefficient γ is determined by

$$\gamma = \frac{k_0}{\ln \frac{z + z_0}{z_0}}. \tag{30}$$

When this concept is applied to conditions over the ocean the question arises as to where to place the zero level. No wind profiles have been measured at high wind velocities when large waves have been present, and we have, therefore, no knowledge of the wind distribution very close to the sea surface.

Using other features—the angle between surface wind (*e.g.*, wind at about 8 m) and pressure gradient, the ratio between surface wind and gradient wind—Rossby arrived at the conclusion that over the ocean the roughness length is independent of the wind velocity and equal to about 0.6 cm. This result is substantiated by the facts that with $z_0 = 0.6$ cm and with 800 cm $< z <$ 1500 cm, equation (30) renders γ values in agreement with those found by entirely different methods, and that these γ values have been applied

successfully to such problems as the generation of wind waves and the maintenance of the equatorial currents [25].

We may therefore state that at wind velocities above 6–7 m sec^{-1} the eddy viscosity in the lower layer above the sea surface behaves *as if* the sea surface were a rough surface (z_0 = 0.6 cm) located about 8 m below the level at which ship observations are ordinarily made. The next question is then, How does the eddy diffusivity behave? If z is not small, the eddy diffusivity must be practically equal to the eddy viscosity, but what happens when z approaches zero? It appears quite probable that in this case, as well, we must introduce a layer next to the boundary surface through which the flux of water vapor takes place by ordinary diffusion. In 1937 the present writer argued as follows [23, p. 4]:

> One can conceive that for a short time a mass of air comes to rest in one of the hollows of the surface. This mass of air will immediately give off its momentum to the surface, but the vapor pressure will not immediately attain the value which is characteristic of the surface. As long as this mass remains at rest, its vapor contents will be increased by processes of ordinary diffusion, supposing that it originally was lower than the equilibrium pressure corresponding to the temperature and salinity of the surface. After a short time the air mass will be removed and replaced by another, and similar processes will be repeated. In order to describe these processes one can introduce a boundary layer, the thickness of which is a statistical quantity representative of the average conditions and within which the transport of water vapor takes place through ordinary diffusion.

A much more extreme view has been advanced by Millar [14] and Montgomery [16], who assume that even over a rough surface the laws pertaining to a smooth surface apply to a distance Z from the boundary surface, but that at that height an abrupt transition to "rough" conditions takes place.

So far, five different attempts have been made to discuss the evaporation from a rough surface, based on different assumptions as to the character of the diffusivity close to the surface. These assumptions can be described as follows:

1. Next to the surface there is a true diffusion layer followed by an intermediate layer of thickness Z in which the eddy diffusivity corresponds to that over a smooth surface, $A = \rho(\kappa + k_0 w_{*s} z)$, where w_{*s} is obtained from equation (20). At Z the eddy diffusivity suddenly increases to the value of the eddy viscosity over a rough surface, and for $z > Z$, $A = k_0 \rho w_{*r} (z + z_0)$ where w_{*r} is obtained from equation (30) [14, 16].

2. Next to the surface there is a true diffusion layer of thickness $\delta = \lambda\nu/w_{*r}$, where w_{*r} applies to a rough surface. From the top of this layer the eddy diffusivity increases at the same rate as the eddy viscosity above a rough surface. These assumptions agree with those of Bunker and others [4] if their notations are made to correspond to those that apply to conditions above a rough surface.

3. Next to the surface there is a true diffusion layer of thickness $\delta = \lambda\nu/w_{*r}$, where w_{*r} applies to a rough surface. At the top of this layer the diffusivity increases abruptly from the molecular value to the value over a rough surface: $A = k_0 \rho w_{*r} (\delta + z_0)$, and for $z > \delta$: $A = k_0 \rho w_{*r} (z + z_0)$ [23].

4. There exists no layer of true diffusion. The eddy diffusivity is at all levels identical with the eddy viscosity, $A = k_0 \rho w_{*r} (z + z_0)$ (Sverdrup [24]).

5. The character of the diffusivity corresponds to that assumed under (1), but at the top of the intermediate layer, at $z = Z$, the vertical flux of water vapor increases abruptly from E_0 to E where $E_0/E = (w_{*s}/w_{*r})^2$ [18].

On the basis of these assumptions it is possible to compute Γ_a. The results are:

$$1.\quad \Gamma_a = \left[\ln \frac{a}{Z} + \frac{w_{*r}}{w_{*s}} \left(\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{k_0 w_{*s} Z}{\kappa} \right) \right]^{-1}, \tag{31}$$

$$2.\quad \Gamma_a = \left(\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{k_0 w_{*r} a}{\kappa} \right)^{-1}, \tag{32}$$

$$3.\quad \Gamma_a = \left(\lambda k_0 \frac{\nu}{\kappa} + \ln \frac{a}{\delta + z_0} \right)^{-1}, \tag{33}$$

$$4.\quad \Gamma_a = \left(\ln \frac{a}{z_0} \right)^{-1}, \tag{34}$$

$$5.\quad \Gamma_a = \left(\ln \frac{a}{z_0} - 1.76 \frac{w_{*r}}{w_{*s}} \right)^{-1}. \tag{35}$$

In Fig. 2 five Γ_a curves, referred to a = 800 cm, are shown as functions of the wind velocity, corresponding to the five sets of assumptions. The curve marked (1) is plotted according to values given by Montgomery [16], taking into account that here we refer Γ_a to a height of 8 m, whereas he used 4 m. When computing Γ_a in cases (2) and (3), the question arises as to what value to assign to λ. It is not obvious that the value for a smooth surface (λ = 11.5) is applicable, but lacking means of determining a better value, we have used it. It may be observed that, on the basis of Montgomery's humidity measurements, Sverdrup [23] found λ = 27.5, but this value seems doubtful because it was derived

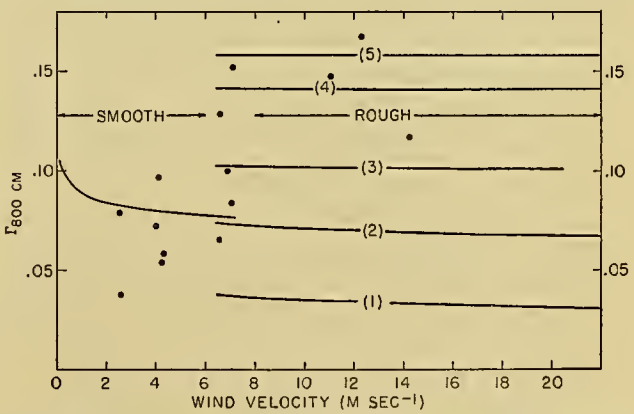


FIG. 2.—Theoretical values of the evaporation coefficient as function of the wind velocity at 800 cm, computed on the basis of different assumptions, and observed values shown by dots.

from observations that included many cases of flow over a smooth surface. With $\lambda = 27.5$, curve (3) would nearly coincide with curve (2).

The wide spread of the curves (2) to (5) clearly demonstrates the importance of the assumptions as to the diffusivity near the boundary surface. Assumptions (1), (4), and (5) appear too extreme, but none can be rejected offhand since we have no basis for describing what actually takes place, but are trying to introduce a model which gives results in consistent agreement with observed conditions. In order to advance a step further we have to compare the theoretical conclusions with the empirical results. Our empirical data are of two kinds. On the one hand we have available a few series of measurements of humidity at different heights above the sea surface from which Γ_a can be found, and on the other hand we have series of accurate ships' observations of meteorological conditions and sea surface temperatures from which we can compute evaporation, using different Γ_a values, and we can compare the results with those derived from energy considerations. In order to facilitate computations and comparisons we shall use vapor pressure instead of specific humidity. Approximately, $e = qp/0.621$ or with $p = 1000$ mb, $e = 161q$. We then write equation (15) in the form

$$E = K_a (e_s - e_a) w_a, \tag{36}$$

and shall call K_a the evaporation factor. We shall use units such that with pressure in millibars and wind velocity in meters per second we obtain the evaporation in millimeters per 24 hours. The curves for K_a ($a = 800$ cm) are shown in Fig. 3. The heavy curves

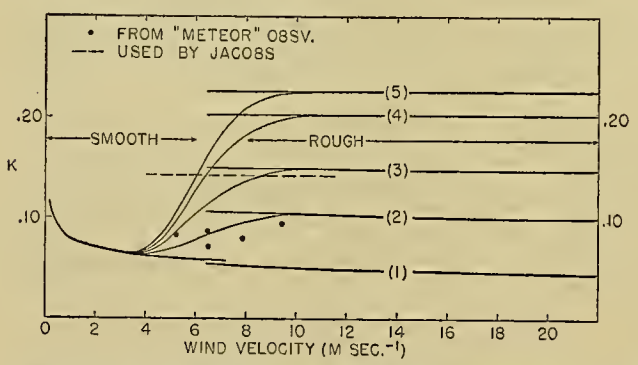


FIG. 3.—Theoretical values of the evaporation factor K as function of the wind velocity at 800 cm, computed on the basis of different assumptions, and empirical values shown by dots and a dashed line.

should be used for computing evaporation from single observations. If climatological averages are available, we must use some transition curves in an interval of velocities around 6–7 m sec^{-1}, the width of which depends upon the spread of values upon which the averages are based. In Fig. 2, arbitrary transition curves are indicated for the interval 2–11 m sec^{-1}.

Discussion. The theoretical considerations lead to such different evaporation factors that the theoretical approach is useless until there can be advanced adequate arguments for accepting one of the many models which at present can be used to describe the character of the eddy diffusion of water vapor close to the sea surface. The theory apparently leads to consistent results at wind velocities up to 6 m sec^{-1} for individual cases, or up to 4 m sec^{-1} for average conditions, but it should be emphasized that this consistency arises because it has been accepted that the sea surface is smooth at low wind velocities. The empirical evidence for this feature is, however, so meager that its reality is badly in need of confirmation. Should it not be confirmed, we must admit that the theoretical computation of the evaporation factor fails at all wind velocities. The only remaining conclusion is that such a factor can probably be established, either on a theoretical basis (if a new foundation for the development of a theory is found) or on an empirical basis.

In trying to do the latter we may follow one of three courses: (*a*) the evaporation coefficient Γ may be determined from measurements of humidity gradients over the sea; (*b*) where accurate meteorological observations are available, and where evaporation can be computed from energy considerations, the evaporation factor may be determined; or (*c*) where climatological charts are available, based on observations that may contain small systematic errors, such as temperature observations on shipboard, the procedure under (*b*) can be used. We shall examine these procedures more closely.

In Fig. 2 the values of Γ have been plotted which have been derived from observations of humidity gradients. Most of the observations were made at wind velocities below 6 m sec^{-1} and agree fairly well with the theoretical values for a smooth surface. The few observations at high wind velocities agree quite well with curve (4) for a rough surface, and this feature led the writer [24] to accept that approach although he previously [23] had adhered to different concepts. It will be shown here that the agreement between the empirical Γ values and the theoretical curve (4) must be accidental because the evaporation factor K, based on curve (4), gives far too high evaporation values.

The second method (*b*), can be used only if very accurate meteorological observations are available for a large area and a sufficiently long time. The *Meteor* observations in the Atlantic [12] satisfy this requirement, and from these the data in Table II have been obtained. The factor K, which has been plotted in Fig. 3, lies between 0.07 and 0.09, and agrees fairly well with values derived from curve (2). The result is in sharp contrast to that derived from examination of humidity gradients and indicates that of the different theoretical assumptions number (2) is the most nearly correct. It should be borne in mind that the computation of the evaporation factor on the above basis is somewhat uncertain, because the meteorological observations are extended over only from 1⅓ to 3 months in summer whereas the "energy values" apply to the whole year. The errors that may be introduced because of this circumstance should, however, not be serious, because in the tropics the annual variation in

evaporation is small and the same probably applies to the South Atlantic down to latitude 55°S.

The third approach has been used by Jacobs [8, 9] in computing the evaporation from the data contained in the *Atlas of Climatic Charts of the Oceans* [27].

TABLE II. EVAPORATION FACTOR K [$K = E/w\ (e_s - e_a)$] (Computed from the meteorological observations on board the *Meteor* and the evaporation rates determined by Wüst)

Region	Latitude range	Number of days	w (m sec^{-1})	$e_s - e_a$ (mb)	Evaporation (Wüst) (mm day^{-1})	K
NE—trade	10°–25°N	67	6.5*	6.4	3.58	0.086
Calm belt	10°N–0°	40	5.2	7.6	3.20	0.082
SE—trade	0°–25°S	107	6.5	7.6	3.48	0.071
Transition	25°–35°S	56	7.9	4.9	3.05	0.079
Westerlies	35°–55°S	86	9.4	2.0	1.75	0.093

* Reduced to the probable average value. The observed value was 8.3 m sec^{-1}, but observations were made in midwinter.

Data of this character may contain small systematic errors in sea-surface and air temperatures, and in humidities and wind velocities. It is, therefore, best to establish the evaporation factor on a strictly empirical basis, requiring that evaporation computed from meteorological data shall agree with that obtained from energy considerations. Jacobs followed this procedure. By undertaking a comparison between results from four limited areas he found

$$K = 0.142$$

and could show that by using this value the total evaporation from the North Atlantic and North Pacific Oceans agrees with that computed on the energy basis. It should be observed that the value 0.142 has no general significance, because it applies to the specific data used by Jacobs and may not apply if other climatological compilations are used. Also it should be mentioned that the factor K used by Jacobs may not be a constant, but may depend on the wind velocity. If so, some of the details of Jacobs' results are possibly in error, but the main conclusions can be expected to be correct.

Results. Using the equation

$$E = 0.142\ (e_s - e_a)w_a \qquad \text{(mm per 24 hr)}, \quad (37)$$

where e is in mb and w in m sec^{-1}, Jacobs has by means of the data in the *Atlas of Climatic Charts of the Oceans* [27] computed regional and seasonal values of the evaporation from the North Atlantic and North Pacific Oceans.

The average values between 0° and 60°N are entered in Table I. The average evaporation in this latitude range agrees with the corresponding average from the energy relations, indicating that a correct value of the factor K was obtained by means of results from four limited areas. Between 0° and 30°N they are *higher*. The observed evaporation rates (Table I, Wüst) display the same feature, which indicates that in low latitudes part of the radiation surplus is stored as heat in the water, is carried north by ocean currents, and is used for evaporation in middle and higher latitudes. The energy that is transported by the ocean currents across the parallel of 30°N amounts, according to Jacobs' values, to 1.4×10^{16} cal min^{-1}. The observed evaporation rates give a similar transport to the north, and at the same time they show no consistent transport in the Southern Hemisphere.

It is well known that the earth receives a radiation surplus in lower latitudes and a deficit in higher latitudes [26]. In order to maintain a steady average temperature distribution, heat must therefore be transported from the equator towards the poles across the parallels of latitude, and at 30°N this transport amounts to 3.6×10^{16} cal min^{-1}. It has been generally assumed that this transport takes place mainly by means of the air currents, but the results referred to above indicate that in the Northern Hemisphere about one-third of the transport takes place by means of ocean currents and two-thirds by means of air currents, whereas in the Southern Hemisphere the transport by ocean currents is negligible. This conclusion seems reasonable because currents that carry warm water into higher latitudes are far better developed in the Northern than in the Southern Hemisphere.

From the preceding discussion it is evident that the computation of evaporation from meteorological observations has so far not rendered an independent check on results obtained by other methods because energy considerations have been used to establish the evaporation factor to be applied. Still, the approach has greatly increased our knowledge as to the evaporation from the oceans because it has furnished a basis for a detailed examination of the energy exchange between the ocean and the atmosphere. This question is dealt with in an adjoining article in this Compendium, which also presents the available information as to the evaporation from different ocean areas in different seasons.

CONCLUSIONS[1]

The problem of evaporation from the oceans has been attacked by three different methods: (1) observations of evaporation from pans on board ship; (2) computations on the basis of the available energy; and (3) computations of vertical flux of water vapor.

1. Observations of evaporation from pans on board ship have rendered fairly consistent values, but the reduction of these values to true rates of evaporation from the sea surface is so uncertain that the two other methods of approach appear to be superior.

2. Different computations based on energy considerations have given similar results as to the average annual

1. When the present article was in press the author received a copy of the paper: Anderson, E. R., and others, *A Review of Evaporation Theory and Development of Instrumentation.* (Lake Mead Water Loss Investigations.) Interim Rep. Navy Electronics Laboratory, Rep. No. 159, pp. 1–70, February 1950. This report includes considerations applicable to a limited body of water and deals with effects of stability. Where the problems are similar, the conclusions agree with those in the present paper. The report contains a long list of references.

evaporation between parallels of latitude. Still, in order to place the results on a firmer basis, it is very desirable to obtain further direct measurements at sea of incoming radiation from sun and sky and of nocturnal radiation. Measurements of the reflectivity of the sea surface under different conditions are also needed, as well as further examination of the Bowen ratio (the ratio between heat losses from the sea surface by conduction and by evaporation). Even with improved knowledge of these conditions, the energy approach cannot render information as to evaporation rates under given meteorological conditions or seasonal and regional values of the evaporation. In order to compute such values the third approach has to be adopted.

3. For the computation of the vertical flux of water vapor from the sea surface several assumptions have been introduced as to the character of the eddy transfer processes near the sea surface. These assumptions lead to different formulas from which the evaporation rate can be obtained from observations of humidity and wind velocity at a standard level above the sea surface and observations of the sea surface temperature from which the vapor pressure at the sea surface can be derived.

These theoretical equations all presuppose indifferent stratification or slight instability. This restriction is not very serious, because stable conditions are not common over the open ocean and when they occur, evaporation is small.

The different approaches have been reviewed and it has been shown that they lead to widely diverging results. Furthermore, it has been shown that the available empirical evidence is far too scanty and too inconsistent to indicate what assumption may be acceptable.

In order to advance our knowledge, it is necessary to carry out a large number of detailed measurements of conditions directly above the sea surface. These should comprise measurements of wind profiles in order to examine the validity of the concept that at low wind velocities the sea surface is hydrodynamically smooth, measurements of humidity in order to examine the validity of the logarithmic law for the decrease of water vapor with height, and measurements of temperature in order to establish the stability under which the observations are made. It is possible that such extended observations may show that one of the suggested theoretical equations is applicable, but it is also possible that new concepts have to be introduced and more variables considered in order to arrive at a satisfactory basis for the computation of evaporation during short time intervals.

The problem is somewhat different if seasonal and regional values are to be computed from climatological data. In this case it may be assumed that the evaporation can be computed by means of an equation that is based on theoretical considerations, and the unknown factor in this equation may be established by means of evaporation values that are derived from energy considerations. Such an equation cannot be expected to give accurate results before the correctness of the character of the formula has been more firmly established, and before it has been shown that the unknown factor is a constant.

REFERENCES

1. Ångström, A., "Applications of Heat Radiation Measurements to the Problems of the Evaporation from Lakes and the Heat Convection at Their Surfaces." *Geogr. Ann., Stockh.*, 2: 237–252 (1920).
2. Bowen, I. S., "The Ratio of Heat Losses by Conduction and by Evaporation from any Water Surface." *Phys. Rev.*, 27: 779–787 (1926).
3. Brückner, E., "Die Bilanz des Kreislaufs des Wassers auf der Erde." *Geogr. Z.*, 11: 436–445 (1905).
4. Bunker, A. F., and others, "Vertical Distribution of Temperature and Humidity over the Caribbean Sea." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 11, No. 1, 82 pp. (1949).
5. Cherubim, R., "Über Verdunstungsmessung auf See." *Ann. Hydrogr., Berl.*, 59: 325–335 (1931).
6. Cummings, N. W., and Richardson, B., "Evaporation from Lakes." *Phys. Rev.*, 30: 527–534 (1927).
7. Hann, J. v., *Lehrbuch der Meteorologie*, 3. Aufl. Leipzig, Tauchnitz, 1915. (See pp. 213–219)
8. Jacobs, W. C., "On the Energy Exchange between Sea and Atmosphere." *J. mar. Res.*, 5: 37–66 (1942).
9. —— "Sources of Atmospheric Heat and Moisture over the North Pacific and North Atlantic Oceans." *Ann. N. Y. Acad. Sci.*, 44: 19–40 (1943).
10. —— and Clarke, Katherine B., "Meteorological Results of Cruise VII of the *Carnegie*, 1928–1929." *Carneg. Instn. Wash. Publ.* 544, 168 pp., Washington, D. C., (1943).
11. Krümmel, O., *Handbuch der Ozeanographie*, Band I. Stuttgart, J. Engelhorn, 1907. (See pp. 243–250)
12. Kuhlbrodt, E., und Reger, J., "Die meteorologischen Beobachtungen." *Wiss. Ergebn. dtsch. Alt. Exped. Forsch. Vermess.-schiff* Meteor, *1925–27*, Bd. 14, 2. Lief., Abschnitt B, SS. 215–392 (1938).
13. McEwen, G. F., "Some Energy Relations between the Sea Surface and the Atmosphere." *J. mar. Res.*, 1: 217–238 (1938).
14. Millar, F. G., "Evaporation from Free Water Surfaces." *Canad. meteor. Mem.*, 1: 41–65 (1937).
15. Mohn, H., *Meteorology* (The Norwegian North-Atlantic Expedition 1876–1878, II). (Translated into English by J. Hazeland). Christiania, Grøndahl & Sons, 1883.
16. Montgomery, R. B., "Observations of Vertical Humidity Distribution above the Ocean Surface and Their Relation to Evaporation." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 4, 30 pp. (1940).
17. Mosby, H., "Verdunstung und Strahlung auf dem Meere." *Ann. Hydrogr., Berl.*, 64: 281–286 (1936).
18. Norris, R., "Evaporation from Extensive Surfaces of Water Roughened by Waves." *Quart. J. R. meteor. Soc.*, 74: 1–12 (1948).
19. Rossby, C.-G., "On the Momentum Transfer at the Sea Surface, I." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 4, No. 3, 20 pp. (1936).
20. Schmidt, W., "Strahlung und Verdunstung an freien Wassenflächen." *Ann. Hydrogr., Berl.*, 43: 111–124, 169–178 (1915).
21. —— "Zur Frage der Verdunstung." *Ann. Hydrogr., Berl.*, 43: 136–145 (1916).
22. Sutton, O. G., "Wind Structure and Evaporation in a Turbulent Atmosphere." *Proc. roy. Soc.*, (A) 146: 701–722 (1934).

23. Sverdrup, H. U., "On the Evaporation from the Oceans." *J. mar. Res.*, 1: 3–14 (1937).
24. —— "The Humidity Gradient over the Sea Surface." *J. Meteor.*, 3: 1–8 (1946).
25. —— "The Wind and the Sea." *P. V. Ass. Ocean. Phys.*, No. 4, pp. 37–55 (1949).
26. —— Johnson, M. W., and Fleming, R. H., *The Oceans, Their Physics, Chemistry and General Biology*. New York, Prentice-Hall, Inc., 1942. (See pp. 99–100)
27. U. S. Weather Bureau, *Atlas of Climatic Charts of the Oceans*. W. B. Publ. No. 1247, 130 pp., Washington, D.C., 1938.
28. Wagner, A., "Zur Frage der Verdunstung." *Beitr. Geophys.*, 34: 85–101 (1931).
29. Wüst, G., "Die Verdunstung auf dem Meere." *Veröff. Inst. Meeresk. Univ. Berl.*, Neue Folge, Reihe A, Heft 6, 95 SS. (1920).
30. —— "Oberflächensalzgehalt, Verdunstung und Niederschlag auf dem Weltmeere." *Länderkundliche Forschung*, Festschrift Norbert Krebs, SS. 347–359 (1936).

FORECASTING OCEAN WAVES[1]

By W. H. MUNK and R. S. ARTHUR

University of California, Scripps Institution of Oceanography

Introduction

One result of the interaction of wind and ocean is the generation of surface waves. Prior to 1940 a few inadequate empirical rules formed the only basis for the prediction of wind-induced waves. A concentrated effort under the stimulus of wartime demands has resulted in the development of relationships which make possible the forecasting of ocean waves from synoptic meteorological data.

As an introductory example, reference is made to the meteorological situation north of the Hawaiian Islands in early January 1947. An intense low-pressure center remained virtually stationary 1000 miles to the northeast of the islands for a period of more than 72 hours (Fig. 1). On January 1 and 2, the winds in an area extending from 500 to 1200 miles north of the islands were force 8–10 and direction N-NNW for a period of 36 hours. This area comprised the *fetch* and computations based on wind velocity, duration, and the forecasting relationships give a wave height of 38 ft and a period of 11.3 sec at the southern boundary [1]. The characteristics of the wind waves which are in the process of generation in such a fetch are referred to as *state of sea* or simply *sea*. After the wind waves left this particular area of direct influence of intense winds, they continued onward as *swell* through a region of relatively weak winds, and reached Hawaii in 28 hours after having traveled a distance of 600 nautical miles. Over this *decay distance* the height of the significant waves decreased and the period and length increased. Computations yield a swell of height 22 ft and period 14 sec, and available observations compare favorably with these calculations. Further computations show the swell reaching Palmyra Atoll, which is 900 miles south of Hawaii, with a height of 12.5 ft and a period 17 sec after an additional travel time of 29 hours.

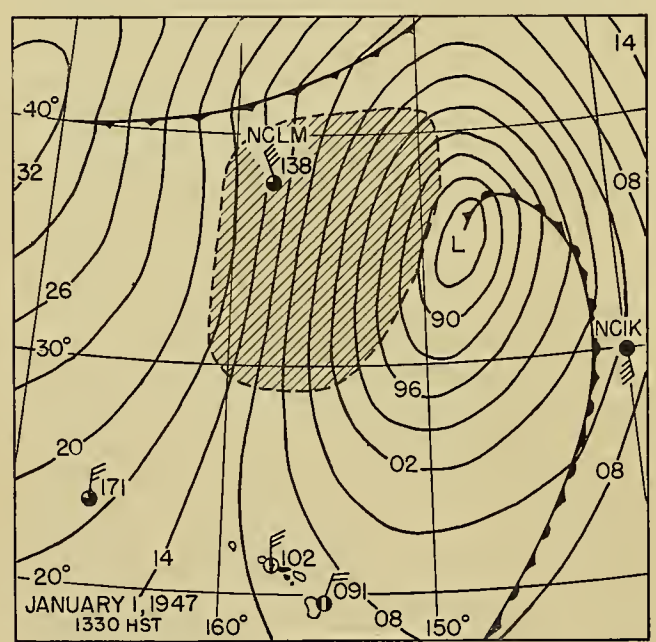


FIG. 1.—Weather map for 1330 Hawaiian Standard Time, January 1, 1947. The shaded area represents the fetch.

The damage in Hawaii was estimated at between one and two million dollars. Three waves are reported to have inundated low-lying Cooper Island in the Palmyra group. Although these wave conditions are extreme for Hawaii and Palmyra, wave damage in general is by no means rare. Damaging waves may even move from one hemisphere to another. In April 1930, more than 20,000 tons of superstructure stone were displaced from the Long Beach, California, breakwater by the action of swell which in all probability had its origin in the "Roaring Forties" of the South Pacific Ocean. The "rollers" of St. Helena and Ascension, which cross the equator after moving from the North Atlantic, have been famous since at least 1846, when thirteen vessels were driven from their moorings and wrecked on the shores of St. Helena. Reports of the disaster indicate that waves broke seaward of the vessels at a depth of around 20 m which denotes that breaker height was approximately 50 ft.

The primary motivation for the development of a wave-forecasting method has been the fact that waves are a nuisance. This problem came into particular prominence during the planning in 1942 of the invasion of North Africa. In that year, Instructor Commander C. T. Suthons, R. N., began an investigation of wave-forecasting techniques in England, and the same problem was studied in the Oceanographic Section, Directorate of Weather, Headquarters, Army Air Forces. The study was continued at Scripps Institution of Oceanography under Air Force and, later, Navy support, and in 1943 the method which was used throughout the war had in essence been developed [8]. Although oceanographers played an important role in the development of forecasting methods, the application is best carried out by meteorologists.

Since waves represent the effect of the wind integrated over time and distance, forecasts of waves can be made from a knowledge of only the large-scale features of the wind distribution. Such forecasts are likely

1. Contribution from the Scripps Institution of Oceanography, New Series No. 511. Many of the results discussed here are based on research supported by the Office of Naval Research, Navy Department, under contract with the University of California.

to be made more easily and with greater certainty than most types of meteorological forecasts.

The present status of wave forecasting is that there are available methods which permit useful predictions of certain wave characteristics from synoptic meteorological observations. The theoretical basis of the methods does not adequately explain the physical mechanism of wave generation and decay. We shall consider wave forecasting, the deficiencies of the methods, and some of the unsolved problems.

Sea, Swell, and Surf

The categories *sea*, *swell*, and *surf* form a natural division of the topic of wave forecasting. The latter, or more generally, the transformation of waves in shallow water, is dealt with by methods which are somewhat different from those used in the study of sea and swell. The essential features of the surf problem have been accounted for physically, whereas this is not the case for sea and swell, as previously mentioned. Sea is determined by the nature of the winds, and the forecasting of sea depends therefore upon the current and predicted synoptic meteorological situation. Swell depends to a lesser extent upon the synoptic situation, but its characteristics can still be altered by a following or an opposing wind during the course of decay. Surf is determined by factors which are largely nonmeteorological, and we confine our consideration of surf to this section, where we note some of the important physical factors in the problem.

As waves move into shallow water the wave length and wave velocity decrease. Waves moving at an angle to the bottom contours are subject to refraction. The wave crests tend to orient themselves parallel to the bottom contours. Changes in wave-crest orientation are associated with convergence and divergence of energy along the crest [11]. As a result, waves over a submarine canyon are relatively low compared to waves on either side of the canyon; the pattern is reversed for a submarine ridge. These changes can all be quantitatively accounted for by the linear-wave theory, assuming conservation of energy and constancy of wave period [22]. An exception occurs where the waves travel over great stretches of shallow water such as exist along the coast of the Gulf of Mexico and the shelf east of Long Branch, New Jersey. In such cases the effects of bottom friction [14] and percolation [13] should be taken into account.

As the waves approach the breaking point the height increases rather suddenly, and the wave breaks at a depth approximately equal to $\frac{4}{3}$ the wave height. Adequate predictions can be based on the application of the solitary-wave theory [12]. Waves breaking at an angle to the shore set up longshore currents whose average velocity can be roughly estimated when bottom contours are straight and parallel [15]. In the case of nearly normal wave approach, longshore currents are variable in direction and form an integral part of a system of currents, called *rip currents* [19], which include narrow zones of fast flow away from shore. Active research is being conducted on the near-shore circulation [19], including rip currents, on orbital velocities, and on the important role played by waves, particularly breakers, in beach erosion and in the design of engineering structures.

Although a number of characteristics of surf are predictable if the deep-water wave characteristics and the bottom topography are known [9], it must be noted that the basic relationships apply only approximately, and even large deviations are to be expected. More research on surf is indicated, particularly on the problem of the mechanism of breaking.

Forecasting Significant Waves

The notations given below will be used in the discussion which follows.

F—length of wind fetch (subscript F denotes value of variable at end of fetch);
D—decay distance (subscript D denotes value of variable at end of decay distance);
H—wave height;
C—phase velocity of wave;
T—wave period;
L—wave length;
δ—H/L (wave steepness);
E—mean energy of wave per unit area;
x—horizontal distance coordinate;
t—time;
t_d—duration time of wind;
U—wind velocity at about 8 m above surface;
β—C/U (wave age);
τ—tangential stress of wind on sea surface;
u_0'—mean surface mass transport velocity;
R_N—mean rate of energy transfer to wave as a result of normal wind pressure;
R_T—mean rate of energy transfer to wave as a result of tangential wind stress;
g—acceleration of gravity;
s—sheltering coefficient;
ρ—density of air;
γ^2—resistance coefficient applicable to wind.

The outstanding characteristic of sea and swell is its irregularity, and a complete description of the sea surface requires the introduction of statistical terms. Nevertheless, a practical purpose can be served by dealing only with the most prominent, or "significant," waves. A definition of the term "significant" is given in the following section. One of the major achievements of wartime wave forecasting was the development of operationally useful methods, one with a theoretical-empirical basis [24], and the other primarily empirical [20]. We next consider the results and basis of the first method, and make a brief comparison with the second.

Observers have noted that as the wind blows over a limited fetch, the wave height and period over the upwind part of the fetch reach a steady state after a limited interval of time. As time passes the steady-state region expands over the whole fetch. The wave height and period at the end of the fetch are determined from the wind velocity, fetch, and duration. The relationships given below, in terms of nondimensional parameters, follow from the forecasting theory.

Steady state:

$$gH_F/U^2 = f_1(gF/U^2), \tag{1}$$

$$C_F/U = f_2(gF/U^2); \tag{2}$$

Transient state:

$$gH_F/U^2 = f_3(gt_d/U), \tag{3}$$

$$C_F/U = f_4(gt_d/U). \tag{4}$$

Height and period of swell at the end of the decay distance and travel time are determined from the decay distance and the period at the end of the fetch according to:

$$H_D/H_F = f_5(D/gT_F^2), \tag{5}$$

$$T_D/T_F = f_6(D/gT_F^2), \tag{6}$$

$$t_D/T_F = f_7(D/gT_F^2). \tag{7}$$

Relationships (1)–(7) are translated into practical graphs for use in forecasting[2] (examples are presented in Figs. 2 and 3).

The basis of the theory is the development of equations which express the energy budget between wind and waves. Use is made of results from two hydrodynamical theories of irrotational waves: (1) the Airy theory of waves of small amplitude, and (2) Stokes' theory of waves of finite amplitude. Certain constants which occur in the solution are evaluated empirically.

2. R. S. Arthur, "Revised Wave Forecasting Graphs and Procedures." *Scripps Institution of Oceanography Wave Report No. 73*, March 1948 (unpublished). These forecasting relationships, as revised according to [24], are to be included in a forthcoming manual on wave forecasting to be issued by the Hydrographic Office.

The assumption is made that energy is transferred from wind to waves by normal pressure and tangential stress. Following Jeffreys, the expression used for the mean rate of energy transfer per unit area as a result of normal wind pressure is

$$R_N = \pm \frac{\pi^2}{2} s\rho(U - C)^2 \delta^2 C, \tag{8}$$

where the sign is positive for $U > C$ and negative for $U < C$. Thus, a transfer expression which depends upon variations in pressure along the surface is introduced, although use is made of hydrodynamical theories which assume constant pressure along the surface.

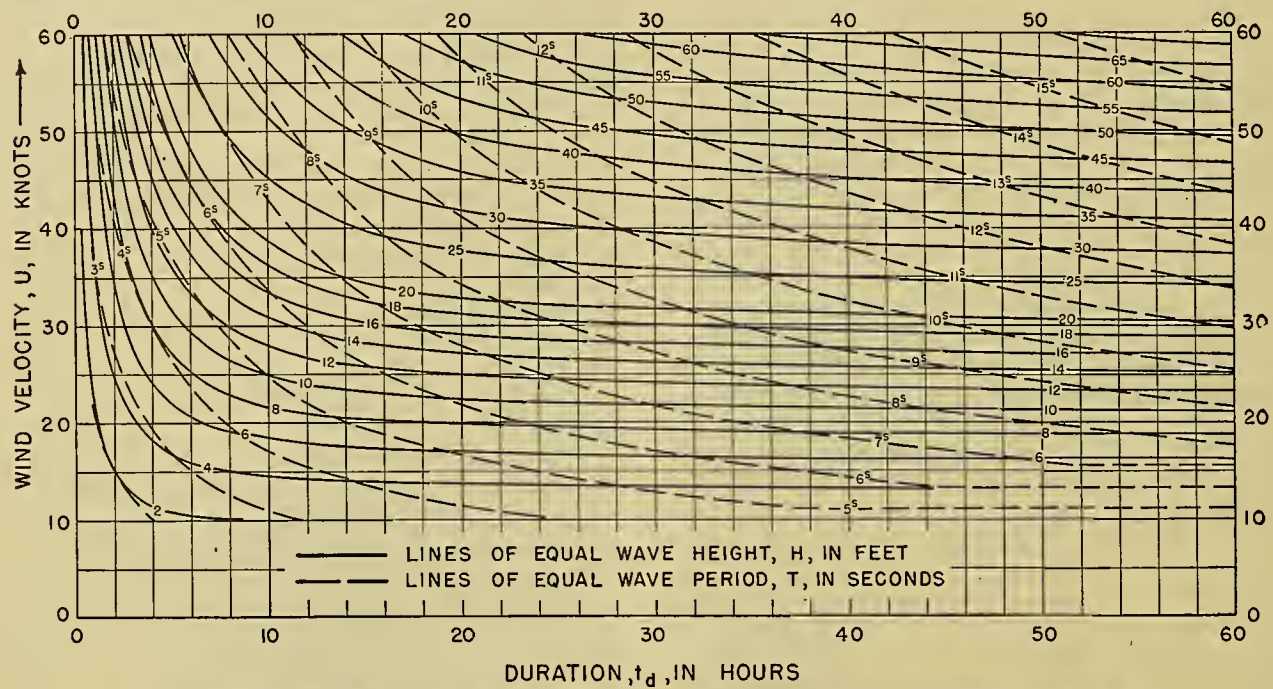


Fig 2—Transient state wave-generation relationships in practical form for use in forecasting.

For the tangential stress, the form

$$\tau = \gamma^2 \rho U^2, \tag{9}$$

where $\gamma^2 = 2.6 \times 10^{-3}$, is introduced with the understanding that wind velocities are great enough so that the sea surface is hydrodynamically rough. The expression for R_N is derived using the horizontal component of particle velocity at the surface from the Airy theory, but on this basis no energy is transferred by tangential stress. However, if the average surface mass transport velocity, $u_0' = \pi^2\delta^2 C$, from the Stokes theory is utilized, the mean rate of energy transfer per unit area by tangential stress is

$$R_T = \frac{1}{L}\int_0^L \tau u_0' \, dx = \pi^2 \rho \gamma^2 \delta^2 C U^2, \tag{10}$$

and on the basis of this expression a transfer of energy from wind to waves is possible even if the waves move faster than the wind. Further investigation of this point is indicated because if the difference between the wind velocity and the horizontal component of particle velocity is introduced in the expression for τ, seemingly

a more accurate formulation of the transfer expression, no energy is added on the basis of either the Airy or the Stokes theory if the waves move faster than the wind. The validity of Stokes' expression for mass transport in the case of waves on a rotating globe has also been questioned.[3]

If the waves are conservative, that is if they maintain their identity, the knowledge from theory of the way in which energy advances with the wave form leads to the following differential equations:

Steady state:

$$\frac{C}{2}\frac{dE}{dx} + \frac{E}{2}\frac{dc}{dx} = R_T \pm R_N; \qquad (11)$$

Transient state:

$$\frac{dE}{dt} + \frac{E}{C}\frac{dC}{dt} = R_T \pm R_N \qquad (12)$$

where the positive sign is used for $C \leqq U$ and the negative sign for $C > U$. Actually the waves comprising sea cannot be conservative and the equations apply at best only approximately.

Since the wave energy is related to the wave height, the two unknowns in the equations above may be considered as wave height and wave velocity and the knowns as distance and time. An additional equation is obtained by assuming that the energy transferred from the wind is divided in a certain fixed manner between the energies required to increase wave height and wave period. The solution of the resulting system of differential equations leads to a relationship between wave steepness and wave age,

$$\delta = f(\beta) \qquad (13)$$

and, finally, to the relationships (1)–(4) above.

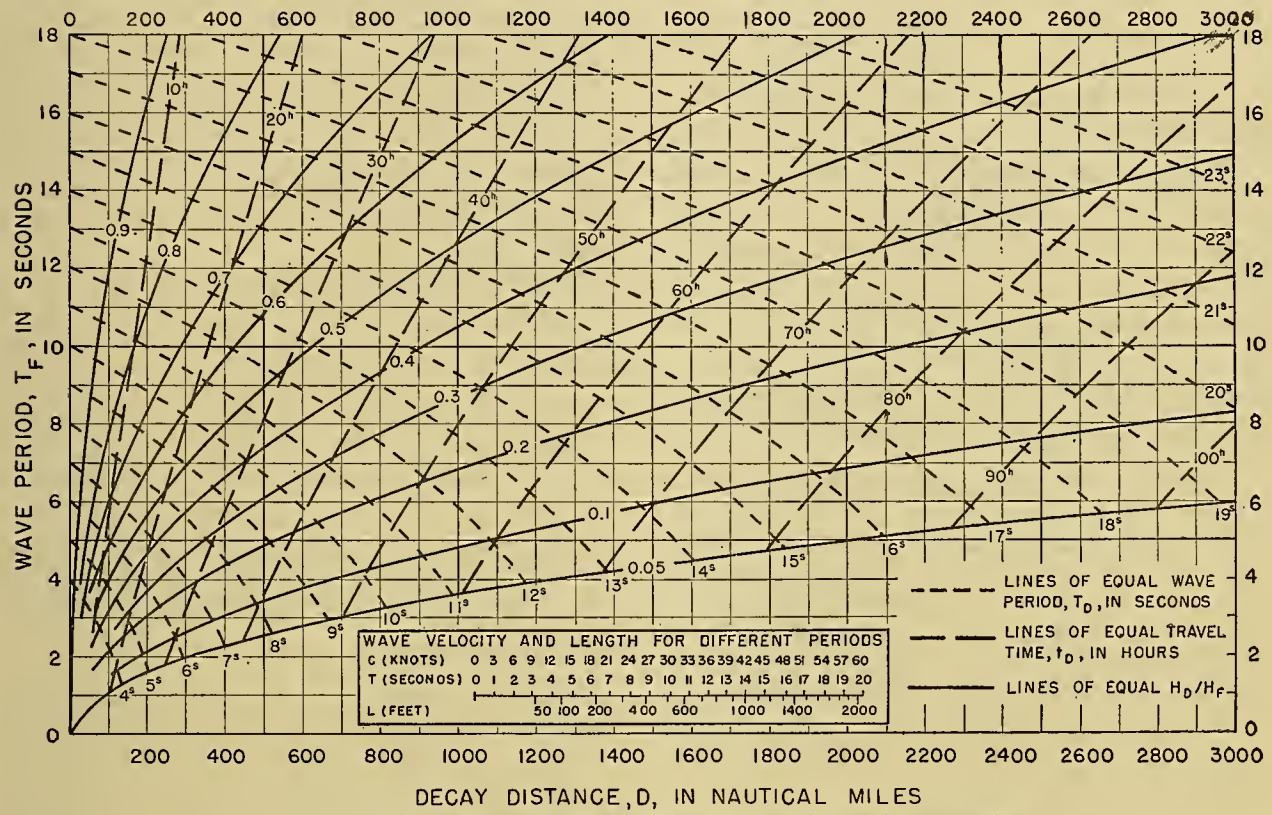


FIG. 3.—Wave-decay relationships in practical form for use in forecasting.

The relationship between δ and β is particularly useful in comparison with wave observations. Many of the observations available for the first checking of the forecasting relationships were inadequate in that either duration time or fetch or both were missing. Most observations are sufficiently complete for the computation of wave steepness and age. Certain constant parameters appearing in the equation are evaluated by fitting the analytical expression (13) to the available observations. The resulting relationship (solid line curve, Fig. 4) gives a satisfactory fit.

Waves leaving the fetch and advancing as swell through a region of calms lose energy by the normal pressure effect of air resistance, but no effect of tangential stress need be considered. The solution of the equation expressing the energy budget of swell yields an increase of wave period during decay and a decrease of wave height, and these facts are in agreement with observations.

Sverdrup [21] compares the decay of swell to the advancement under the influence of air resistance of a train of impulsively generated waves and shows that the period increase is explicable as a result of selective

3. F. Ursell, "The Theoretical Form of Ocean Swell on a Rotating Earth." Admiralty Research Laboratory, A.R.L./R.6/1.03.41/W, February 1947 (unpublished).

attenuation. The comparison suggests that travel time be computed on the basis of the decay distance and the group velocity of the swell at the end of the decay distance. The effect of a following wind during decay is to give lower swell arriving earlier, and the effect of an opposing wind is the opposite. The results for travel

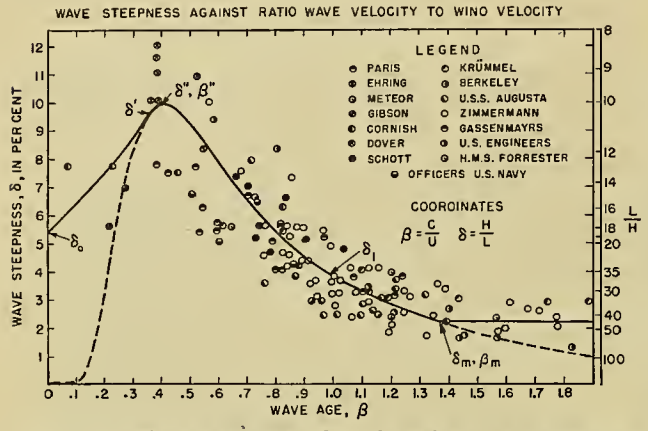


FIG. 4.—Wave steepness plotted against wave age. The dashed lines represent the part of the theoretical relationship which must be modified to conform to empirical evidence.

time, following wind, and opposing wind are incorporated into the forecasting method.

Groen and Dorrestein [7] attribute the decay and the period increase of swell to a selective damping by eddy viscosity. The eddy-viscosity coefficient used depends on the size of the systems under consideration, and the coefficient is larger for long waves than for short ones. Groen and Dorrestein point out that it is improbable that the values of the sheltering coefficient would be the same for swell and wind waves as Sverdrup and Munk [24] have assumed.

TABLE I. PREDICTED AND OBSERVED WAVE CHARACTERISTICS

	Per cent of cases
Predicted height within 1 ft of observed height	53
Predicted height within 2 ft of observed height	80
Predicted period within 2 sec of observed period	63
Predicted arrival time within 6 hr of observed time	69

Isaacs and Saville [10] have compared forecasts made according to the method in its present form with records from wave meters at Point Sur, California, and Heceta Head, Oregon. The results of the comparison for a total of more than 200 forecasts for each station during the period April to December, 1947, are given in Table I. The predicted heights are appropriately corrected for the transformation from deep water to the position of the wave meter. They do not show any tendency to be higher or lower than the observed heights. The predicted arrival times, however, show a tendency to be too early and the periods to be too low.

According to Isaacs and Saville, "97 per cent of the recorded significant *increases* in wave height were forecast, but 23 per cent of the forecast wave trains failed to arrive. The rather large proportion of non-arrivals apparently resulted from the erroneous selection of fetches. . . ." They conclude that the forecasting technique "results in a high degree of reliability for forecasting the arrival of significant increases in wave height, and prognosticating the wave heights," but that "the forecast of wave period and arrival time of the peaks of the wave trains does not display the same degree of reliability. . . ." Bates [3] reviews the forecasting of sea, swell, and surf for landing operations in World War II and concludes that the forecasts were basically correct.

Suthons [20] presents graphs for the forecasting of sea and swell, but the empirical data on which they are based and the exact methods of preparation are not indicated. Predicted wave periods for sea are longer from these graphs than from the first method, and wave heights lower. If a δ,β-relationship is derived from Suthons' graphs, the agreement with the observational data mentioned above (Fig. 4) is not good over a wide range of β values. Suthons attributes the increase in wave period to the downward transport of wave energy by molecular and eddy viscosity, but no quantitative discussion of the mechanism is presented.

The Wave Spectrum

The great complexity of the waves in the sea obscures the exact meaning of earlier wave observations which form the empirical basis for the forecasting methods discussed in the preceding section. Only recently has it been possible, by analyzing records from pressure-type wave meters, to clarify the relation of the predicted wave height and period to the irregular pattern actually recorded.

An individual wave record shows a wide range of heights and periods. For many practical applications, such as the landing of craft, wave action on structures, and sand movement, only the heights and periods of the higher waves in the record are significant. The terms *significant wave height* and *significant wave period* have accordingly been defined as the average height and period, respectively, of the highest one-third of the waves, after ripples and waves of height less than one foot are eliminated from consideration [23].

Experience shows that observers who attempt to determine visually the characteristics of waves tend to record heights and periods which are approximately those of the significant waves rather than of the average waves. Since the forecasting method is based on such visual observations, one might expect that the predicted wave heights and periods are to be considered significant heights and periods. This expectation has subsequently been confirmed [10].

The characterization of a wave record in terms of the

significant waves gives a useful description of the record.[4] Wiegel [25] finds that the daily average height of the highest ten per cent of the waves bears a relatively constant relationship to the significant height at the locations on the Pacific Coast of the United States. This fact is of practical importance because less time is required to compute the average for the highest ten per cent. A good correlation between the maximum waves recorded each day and the significant waves is also noted, and Seiwell [17] finds at Cuttyhunk and Bermuda that the ratio of the daily average height of all waves to the highest one-third is relatively constant. These results (Table II) taken collectively indicate that in general a similar *distribution* of wave heights prevails in the waves generated in different storms although the absolute magnitude varies.

Table II. Ratios of Mean Wave Height and of Average of Highest 10 Per Cent to Significant Wave Height

	Observation Point					
	Scripps[4] (swell)	Point Sur [25]	Heceta Head [25]	Point Arguello [25]	Cutty-hunk [17]	Ber-muda [17]
Interval over which averages formed	46 waves	three 20-min intervals per day			2-min intervals	
Ratio of mean to significant	0.67				0.64	0.64
Ratio of highest 10% to significant		1.27	1.30	1.30		

A more detailed interpretation is obtained by subjecting the records to harmonic analysis. Barber and Ursell [2] have pioneered in the development of instruments for obtaining the Fourier amplitude spectrum. The spectra are taken to indicate that a storm produces a mixture of trains of waves of all lengths up to a maximum which depends on the greatest wind strength. The component trains act independently during decay, and each travels at the group velocity appropriate to its period. In an individual spectrum at a very distant station the swell covers a narrow range of periods, but at a closer station the wave band is wider and broadens rapidly toward shorter periods. Seiwell and Wadsworth [18], on the other hand, take issue with such an interpretation of the component periods in the Fourier amplitude spectrum. From a computation of the autocorrelation function they reach the following rather surprising conclusions which are quoted from their summary:

(1) Periodogram analyses performed on oceanic wave records do not appear to give correct geophysical information. The numerous wave periods, and bands of periods, indicated by this type of analysis do not necessarily possess physical significance.

(2) Application of the hypothesis of generalized harmonic analyses to western North Atlantic wave records indicates that ocean wave patterns are not complex interference patterns resulting from combinations of many wave frequencies, but frequently consist of a single sinusoidal wave on which is superimposed an oscillatory component. . . .

This question of the interpretation of records is fundamental in the consideration of wave generation and propagation, and it merits further study.

Problems in Forecasting

Three main stages in the study of forecasting ocean waves from storms can be recognized. In the earliest stage the forecasting was based on empirical "engineering rules," mutually inconsistent and often incomplete. The well-known Stevenson's law, for example, relates wave height to fetch without regard to wind speed. The second stage was the development during World War II of consistent, dimensionally correct relationships. These relationships have made it possible to forecast the significant waves to an accuracy sufficient for many practical applications. Most of the foregoing discussion deals with this stage of the development. The relationships are strengthened by the application of hydrodynamic theory, although the theoretical basis is not at all secure. The third stage, and one which we are now entering, centers about the British studies of the wave spectrum [2, 5], and some recent attempts to forecast the entire spectrum [4].

According to Deacon [4],

> The width of the spectrum when waves are being generated by a rising wind suggests that the energy begins to be distributed over a range of wave lengths as soon as the waves are formed, some being communicated to the longer wave lengths before the shorter ones are fully energized. Each spectrum has an *optimum band* in which the waves are highest. Waves shorter than this optimum period are lower, presumably because they have a smaller capacity for absorbing energy without becoming unstable; and longer waves are lower, probably because their speed, which is nearly equal to that of the wind, allows them less opportunity of absorbing energy. . . . The period of the highest waves is approximately 25 per cent less than the period of the longest waves.

From this point of view the forecasting of significant waves is closely related to the forecasting of the optimum band. An increase in significant wave height with fetch, or with wind speed, corresponds to an increase in the maximum energy contained in this band, and an increase in the period of the significant waves represents a shift of the optimum band towards longer periods. Since the characteristic of this optimum band depends closely on the character of the component wave trains longer and shorter than those of the band, it is not possible to develop a satisfactory physical theory for forecasting the optimum band without regard to the remaining portion of the spectrum. Nor is it desirable to do so, for there are many practical problems for which the character of the optimum band is only incidental. In the study of the acoustic and optical properties of the sea surface the short-period portion of

4. W. H. Munk, "Proposed Uniform Procedure for Observing Waves and Interpreting Instrument Records." *Scripps Institution of Oceanography Wave Report No. 26*, December 12, 1944 (unpublished).

the spectrum is the important one. In the problem of storm tracking we are most interested in the longest periods present, as those travel most rapidly.[5] It is therefore desirable to divorce from the problem of wave forecasting the selection of that particular portion of the spectrum which happens to be important to a definite application.

The derivation of the physical laws governing the growth of the component wave trains under wind action remains the outstanding theoretical problem in the development of a method for forecasting the wave spectrum. The problem promises to be a difficult one for at least two reasons: (1) the limitations on the growth of the shorter waves imposed by stability considerations introduce a nonlinearity into the problem; (2) to the extent to which the lower, shorter waves are sheltered from the wind by the large waves, the energy budgets of the component wave trains are not mutually independent.

The problem of forecasting the transformation of the swell traveling through the region of decay also resolves itself into the problem of forecasting the change in the spectrum of the swell. The observed increase in the period of the significant swell is the result of the relatively rapid attenuation of the short-period components. One possible mechanism is provided by the viscous effects of the water, according to which the dissipation of energy is inversely proportional to the fourth power of the wave period. It can, however, easily be demonstrated that the effect of viscosity is several orders of magnitude too small to explain the observed transformation. Sverdrup [21] has, however, demonstrated that the selective attenuation resulting from the air resistance on waves traveling through regions of calm (equation (8) with $U = 0$) provides a possible mechanism for selective attenuation. Groen and Dorrestein suggest that selective damping by eddy viscosity is the mechanism. Further study is indicated.

It should also be noted that laws governing the generation of waves at low wind speed can be expected to differ from those governing the generation at moderate and high speeds. The reason lies in the character of the sea surface, which appears to be hydrodynamically smooth at wind speeds less than 6 m sec^{-1}, and hydrodynamically rough at wind speeds exceeding 7 m sec^{-1} [16]. Equation (9) and all subsequent development is based on the assumption of a rough sea surface.

In addition to the more fundamental problems mentioned above, there remain many problems of practical applications. Among the more important ones is the prediction of waves and swell from a tropical storm. The rules of thumb which have been established[6] do not in all cases lead to adequate forecasts.

Most of the actual testing of wave forecasting methods has been carried out for west-coast situations. Forecasting for an east coast offers some special problems, and these are being investigated by the Department of Meteorology, New York University, under the sponsorship of the Beach Erosion Board, and by the Woods Hole Oceanographic Institution [6].

5. Consult "Ocean Waves as a Meteorological Tool" by W. H. Munk, pp. 1090–1100 in this Compendium.

6. W. J. Francis, Jr., Commander, USN, "Waves and Swell from a Tropical Storm." *Scripps Institution of Oceanography Wave Report No. 29*, Nov. 1944 (unpublished).

REFERENCES

1. Arthur, R. S., "Forecasting Hawaiian Swell from January 2 to 5, 1947." *Bull. Amer. meteor. Soc.*, 29: 395–400 (1948).
2. Barber, N. F., and Ursell, F., "The Generation and Propagation of Ocean Waves and Swell. I. Wave Periods and Velocities." *Phil. Trans. roy. Soc. London*, (A) 240: 527–560 (1948).
3. Bates, C. C., "Utilization of Wave Forecasting in the Invasions of Normandy, Burma, and Japan." *Ann. N. Y. Acad. Sci.*, 51: 545–569 (1949).
4. Deacon, G. E. R., "Recent Studies of Waves and Swell." *Ann. N. Y. Acad. Sci.*, 51: 475–482 (1949).
5. —— "Waves and Swell." *Quart. J. R. meteor. Soc.*, 75: 227–238 (1949).
6. Donn, W. L., "Studies of Waves and Swell in the Western North Atlantic." *Trans. Amer. geophys. Un.*, 30: 507–516 (1949).
7. Groen, P., and Dorrestein, R., "Ocean Swell: Its Decay and Period Increase." *Nature*, 165: 445–447 (1950).
8. Hydrographic Office, United States Navy, *Wind Waves and Swell—Principles in Forecasting*. H. O. Misc. 11,275, 61 pp., 1943.
9. Hydrographic Office, United States Navy, *Breakers and Surf—Principles in Forecasting*. H. O. No. 234, 55 pp., 1944.
10. Isaacs, J. D., and Saville, T., Jr., "A Comparison Between Recorded and Forecast Waves on the Pacific Coast." *Ann. N. Y. Acad. Sci.*, 51: 502–510 (1949).
11. Johnson, J. W., O'Brien, M. P., and Isaacs, J. D., "Graphical Construction of Wave Refraction Diagrams." *U. S. Hydrogr. Off. Tech. Rep.* No. 2, H. O. Publ. No. 605, 45 pp. (1948).
12. Munk, W. H., "The Solitary Wave Theory and Its Application to Surf Problems." *Ann. N. Y. Acad. Sci.*, 51: 376–424 (1949).
13. Putnam, J. A., "Loss of Wave Energy Due to Percolation in a Permeable Sea Bottom." *Trans. Amer. geophys. Un.*, 30: 349–356 (1949).
14. —— and Johnson, J. W., "The Dissipation of Wave Energy by Bottom Friction." *Trans. Amer. geophys. Un.*, 30: 67–74 (1949).
15. Putnam, J. A., Munk, W. H., and Traylor, M. A., "The Prediction of Longshore Currents." *Trans. Amer. geophys. Un.*, 30: 337–345 (1949).
16. Rossby, C.-G., and Montgomery, R. B., "The Layer of Frictional Influence in Wind and Ocean Currents." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 3, No. 3, 101 pp. (1935).
17. Seiwell, H. R., "Sea Surface Roughness Measurements in Theory and Practice." *Ann. N. Y. Acad. Sci.*, 51: 483–500 (1949).
18. —— and Wadsworth, G. P., "A New Development in Ocean Wave Research." *Science*, 109: 271–274 (1949).
19. Shepard, F. P., and Inman, D. L., "Nearshore Water Circulation Related to Bottom Topography and Wave Refraction." *Trans. Amer. geophys. Un.*, 31: 196–212 (1950).
20. Suthons, C. T., *The Forecasting of Sea and Swell Waves.*

Naval Meteorological Branch Memo, No. 135/45, Oct. 1945.

21. Sverdrup, H. U., "Period Increase of Ocean Swell." *Trans. Amer. geophys. Un.*, 28: 407–417 (1947).
22. —— and Munk, W. H., "Theoretical and Empirical Relations in Forecasting Breakers and Surf." *Trans. Amer. geophys. Un.*, 27: 828–836 (1946).
23. —— "Empirical and Theoretical Relations between Wind, Sea, and Swell." *Trans. Amer. geophys. Un.*, 27: 823–827 (1946).
24. —— "Wind, Sea, and Swell: Theory of Relations for Forecasting." *U. S. Hydrogr. Off. Tech. Rep.*, No. 1, H. O. Publ. No. 601, 44 pp. (March 1947).
25. Wiegel, R. L., "An Analysis of Data from Wave Recorders on the Pacific Coast of the United States." *Trans. Amer. geophys. Un.*, 30: 700–704 (1949).

OCEAN WAVES AS A METEOROLOGICAL TOOL[1]

By W. H. MUNK

University of California, Scripps Institution of Oceanography and Institute of Geophysics

Introduction

Since ancient times seafaring men have recognized the significance of ocean waves as a sea-borne storm warning. However, their interpretation was intuitive, based upon the experience gained in a lifetime at sea.

The first attempt to apply this ancient art as an aid in weather forecasting appears to have been made in New Zealand around the turn of the century.[2] Yet no *quantitative* methods were developed until World War II. We shall describe the three general methods which present themselves at this time. Each of these has its advantages and disadvantages, but it is too early to state whether these methods will find widespread application.

The Height-Period Method Applied to Visible Swell

This method is based on the relationships established for forecasting sea and swell from weather maps [15, 16].[3] It consists of computing the wind speed U in a storm area, the distance D from the storm area, and the travel time t_D of a swell, making use of measurements of the swell height H_D and period T_D. A discussion of these quantities is found in another paper.[3] Wave height and period refer to deep water, and to the *significant* waves, that is, the average of the highest one-third of all waves present.[4] Observations from shipboard can therefore be taken without further modification. In the case of land-based observations the wave *period* can be determined at any depth since it does not change as waves enter shallow water. Along steep coastlines or exposed beaches with simple bottom topography the deep water wave *height* can, for practical purposes, be taken as the wave height one or two wave lengths outside the breaker zone. Otherwise it is necessary to take into account, by means of specially constructed "refraction diagrams" [7, 12], the effect of bottom topography and of the configuration of the coastline.

The nondimensional relationships given below follow from the forecasting theory [16]:[5]

$$\frac{U^2}{gH_D} = \frac{1}{2\pi}\frac{1}{\delta_F\beta_F^2}\left(\frac{\delta_F}{\delta_D}\right)^{.57}, \tag{1}$$

$$\frac{D}{gT_D^2} = 1.72 \times 10^3\left[1 - \left(\frac{\delta_D}{\delta_F}\right)^{.429}\right], \tag{2}$$

$$\frac{t_D}{T_D} = 2.16 \times 10^4\left[1 - \left(\frac{\delta_D}{\delta_F}\right)^{.429}\right], \tag{3}$$

$$\delta_D = \frac{2\pi}{g}\frac{H_D}{T_D^2}, \tag{4}$$

$$\delta_F = \delta_F(\beta_F), \tag{5}$$

$$\beta_F = \beta_F(gt/U), \tag{6}$$

where g is gravity, and t the storm duration. The relationships (5) and (6) are very complicated analytically, and are here presented in graphical form (Fig. 1). Equations (1) to (6) contain seven unknowns, and the quantities U, D, and t_D are therefore not uniquely determined from measurements of swell height and period. To obtain a complete solution a relation

$$t = t\,(U) \tag{7}$$

will be assumed between the duration of the storm and its intensity, based on the common experience that high winds are usually of short duration. For two specific relationships, a storm of unusual duration (line A in Fig. 2 inset), and a short-lived storm (line B), equations (1) to (7) have been solved and the results shown graphically (Fig. 2), using practical units. The examples in Table I illustrate that D and t_D do not depend very critically upon the nature of the relationship $t(U)$ and can be determined with greater certainty than U. To obtain greater accuracy the forecaster may attempt an interpolation between Cases A and B, based on his experience with the duration of storms over particular regions. It should be noted that the results ob-

1. Contribution from the Scripps Institution of Oceanography, New Series No. 512. Many of the results discussed here are based on research carried out for the Hydrographic Office, the Office of Naval Research, and the Bureau of Ships of the Navy Department, under contract with the University of California.

2. Prof. C. E. Palmer, in a letter to the author, writes, "In the early days of weather forecasting in New Zealand (towards the end of the last century) no information from Australia was available, consequently the only meteorological warning of storms coming from the west was the local fall in the barometer. In an endeavor to get more advanced warnings, Captain Edwin, who was then in charge of a very small weather forecasting organization in the Marine Department, used the observations of sea and swells from light houses to supplement the data on the synoptic maps. These old maps are still in the archives of the New Zealand Meteorological Service and I have studied them with some interest because the reconstruction of the storm positions and tracks by means of the ocean observations seems remarkably good."

3. Consult "Forecasting Ocean Waves" by W. H. Munk and R. S. Arthur, pp. 1082–1089 in this Compendium.

4. In estimating wave heights from visual observations there is a tendency to arrive at values well in excess of the mean heights for all waves present. The definition of significant waves given above has been found in accord with the estimates of most observers.

5. Equation (3) has been revised according to equation (1) in R. S. Arthur, "Revised Wave Forecasting Graphs and Procedure." *Scripps Institution Wave Report 73*, March 1948 (unpublished).

tained become increasingly uncertain as the distance between storm and observer becomes greater.

For a discussion of the assumptions underlying the foregoing equations the reader is referred to the original paper [16]. The advantage of this method is that for

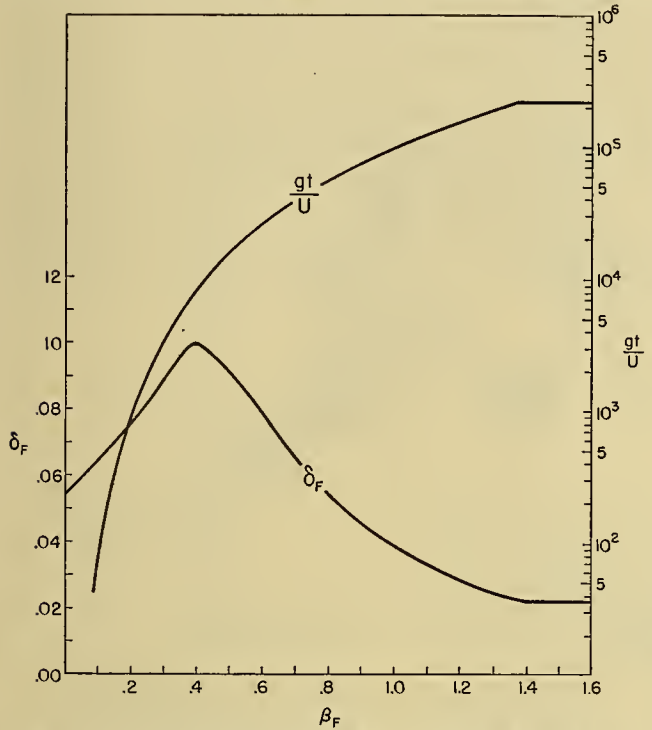


FIG. 1.—Plot of equations (5) and (6) relating wave steepness δ (height/length), wave age β (wave velocity/wind velocity), wind speed U, and wind duration t. Subscript "F" refers to values at end of feteh. (*From Sverdrup and Munk* [16], *Figs. 5 and 7.*)

shipboard observations, or for exposed and steep coasts, it permits a rapid determination of the storm distance and a rough estimate of the storm intensity without the need of special wave-measuring equipment. In this sense the method was first used during World War II by Capt. D. F. Leipper, AAF, as an aid in forecasting the arrival of intense storms in the Aleutian area. In many instances the arrival of heavy swell preceded all other indications of the storm. From a climatological point of view this method can be applied by "mapping" on an H_D,T_D-diagram typical meteorological situations associated with wave conditions in certain areas (Fig. 3).

The principal disadvantage of this method is the slow velocity of the significant waves. As a result the information concerning storm location and intensity may be several days old before the waves reach shore.

The Frequency-Spectrum Method Applied to Forerunners of Swell

The limitations outlined above can be overcome in part by dealing with the low, long, imperceptible waves which precede the visible swell. Since these waves are only a few inches high it is not possible to apply simple, visual methods of observation, but the advantages involved often appear to justify the use of special instrumental techniques.

In the first place, the forerunners travel considerably faster than the visible swell. On the basis of instrumental observations now available [1] the period of the forerunners may reach 25 sec, compared to 12–13 sec

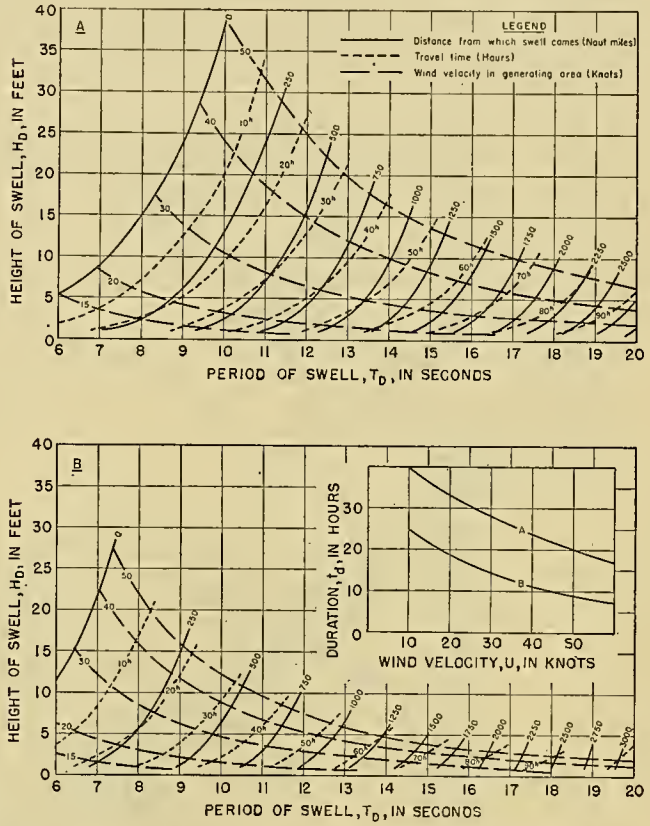


FIG. 2.—Distance from which swell comes, travel time, and wind velocity in generating area as functions of observed height and period of swell. The upper figure is drawn for a long-lived storm, the lower figure for a short-lived storm. The assumed relation between wind speed and duration for these two cases is shown in the inset. (*From Sverdrup and Munk* [16].)

as a typical maximum period for the swell. The corresponding group velocities with which the disturbance created by a storm is effectively propagated by surface wave motion are of the order of 900 nautical miles per day for the forerunners, in constrast with 400 miles per day for the swell.

TABLE I. SAMPLE CALCULATIONS FOR A STORM OF UNUSUAL DURATION (CASE A) AND A SHORT-LIVED STORM (CASE B)

H_D (ft)	T_D (sec)	Storm duration (see inset, Fig. 2)	D (naut. mi)	t_D (hr)	U (knots)
6	12	Case A	700	38	30
		Case B	825	45	45

A second advantage of the "frequency-spectrum method" over the "height-period method" is that travel time and storm distance are *uniquely* determined from a knowledge of wave periods. The need for measuring wave height is eliminated.

Instrumentation. Shore-based instruments for recording ocean waves have been in existence for many years.

Most of the older instruments recorded the elevation of the sea surface against time by some suitable arrangements of floats. During World War II very effective methods were developed which derive from the underwater pressure fluctuations induced by the surface waves. It is curious that this mode of attack should have been initiated independently in Great Britain and the United States. An early model developed by the

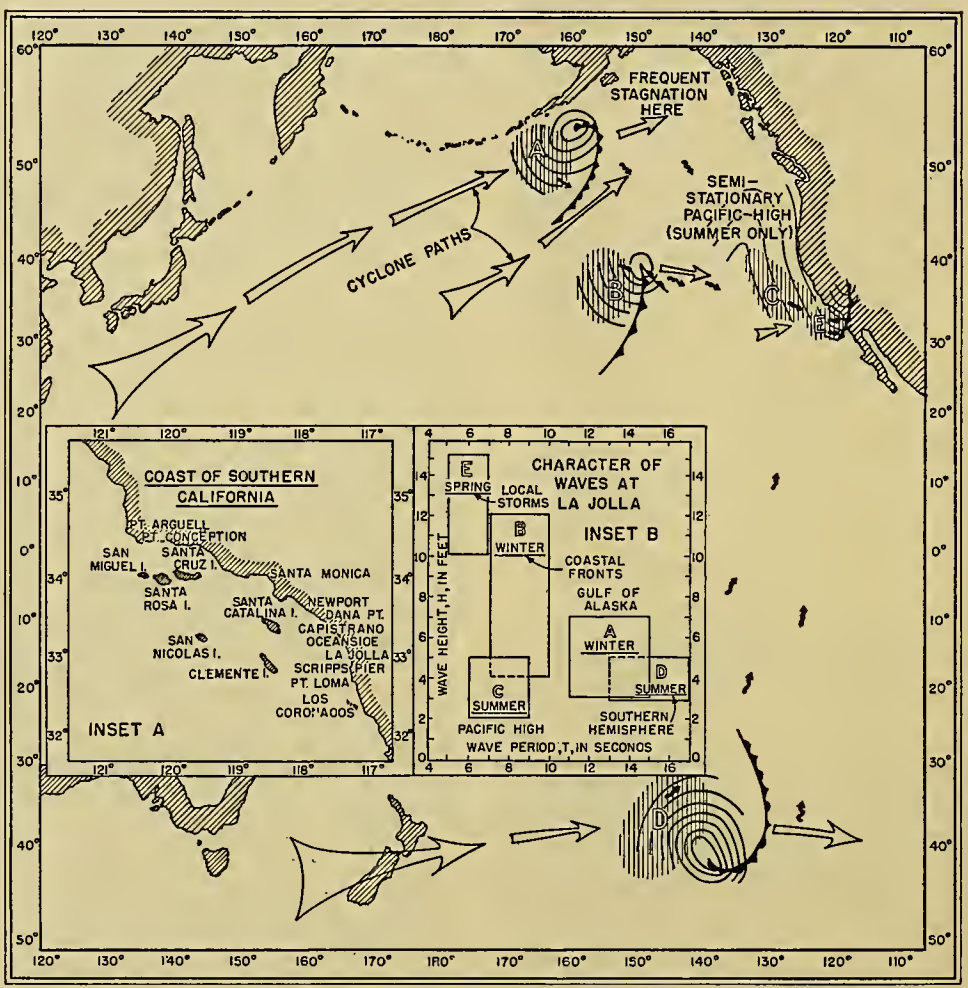


Fig. 3.—Origin of waves reaching La Jolla, California. The capital letters on the large chart represent typical meteorological situations, which give rise at La Jolla to waves of heights and periods shown in inset B. Inset A is a chart of the region near La Jolla on an enlarged scale. The shaded areas on the large chart represent typical fetches, that is, areas where the waves are generated. The isobaric pattern and the nature of the meteorological fronts are also indicated. The large open arrows represent typical paths of the storm systems. The wavy arrows indicate the direction of the waves leaving the storm systems. The figure cannot represent the nature of the meteorological situations with any degree of accuracy, but it is supposed to illustrate how different meteorological situations can be identified by the height and period of the waves with which they are associated. (*From Munk and Traylor* [12].)

Mine Design Department, British Admiralty, has been in operation at Pendeen, near Lands End, England, since 1944. In the United States the development of wave instruments has been carried out chiefly at the Woods Hole Oceanographic Institution and at the Department of Engineering of the University of California. In an early model designed by M. Ewing the recording apparatus was contained directly in the underwater unit. Later, units recording on shore were installed near Cuttyhunk, Massachusetts, and on the island of Bermuda [8]. A number of shore recording models have been developed at the College of Engineering, University of California [6], and these have been installed at Quillayute River, Washington; Heceta Head, Oregon; Point Cabrillo, Point Sur, Point Arguello, Oceanside, and La Jolla, California; and on the island of Guam.

Although the various instruments differ in many important aspects, the following principles are common to most wave instruments of the underwater pressure type and should be understood for a proper interpretation of the records. Surface waves induce pressure fluctuations in the entire column of water between the surface and the sea bottom (Fig. 4). For any given wave height and depth of water, the amplitude of these fluctuations depends on the wave period in such a manner that waves of very short period are virtually eliminated. The underwater unit measures the deviation of the fluctuating pressure at the sea bottom from the mean hydrostatic pressure. A "slow leak" in the underwater unit elimi-

nates the effect of tides and other very long waves. The pressure fluctuations are converted into electrical modulations, which are transmitted through a cable to a shore recorder. Finally the records are subjected to a harmonic analysis by a special instrument [2]. The natural wave spectrum is therefore modified in three stages: (1) waves of short period are removed by hydrodynamic filtering; (2) waves of very long period are removed by the slow leak in the underwater unit; (3) the remaining record is broken down further by harmonic analysis.

Hydrodynamic filtering is particularly adapted for extracting the low forerunners from the complex pattern of the sea surface. The response characteristics of this filter can be expressed by the ratio $\Delta P/\Delta P_0$ of the amplitude of the pressure fluctuation at the bottom to

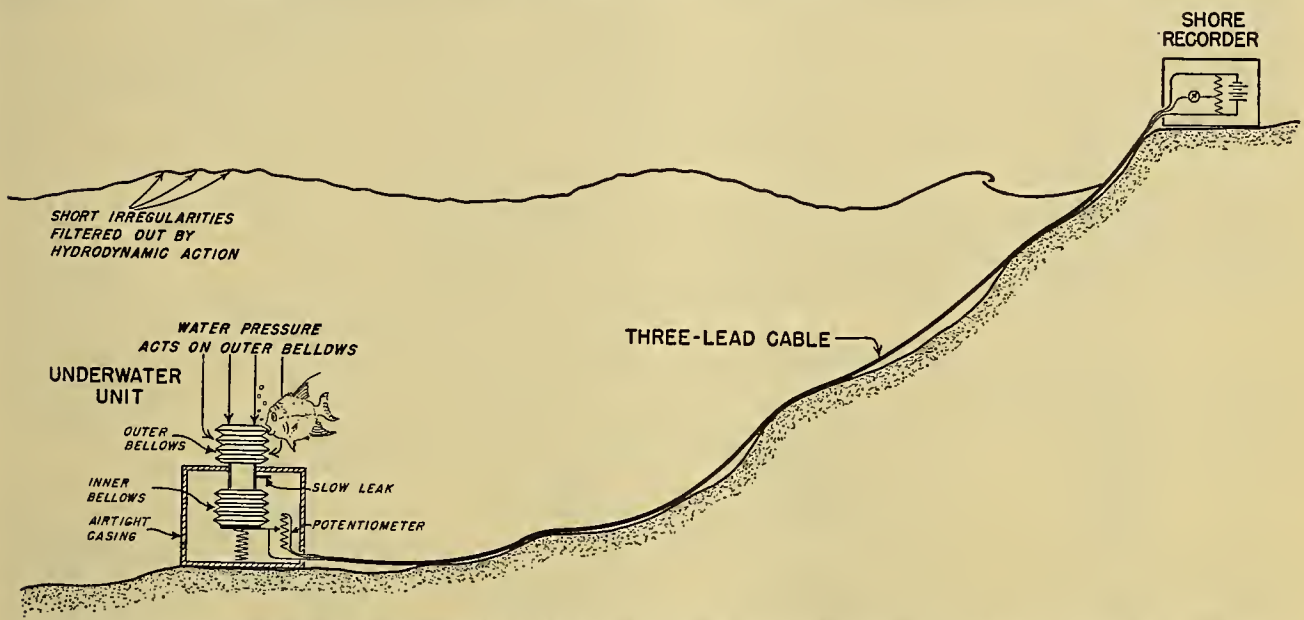


Fig. 4.—Recording mechanism. The pressure is transmitted through an outer bellows into a second bellows inside the instrument casing, so that the pressure of the air inside the two bellows always equals the pressure in the water outside. The air inside the bellows can pass through a slow leak into the instrument casing, and the *average* pressure inside the casing equals the *average* pressure in the water, that is, the hydrostatic pressure. The leak is so slow that the pressure inside the instrument does not change appreciably during one wave period, yet it permits pressure equalization related to tides. The total displacement of the inner bellows depends on the *difference* in pressure between the inside and the outside of the bellows and therefore measures the deviations of pressure (amplitude ΔP) from the hydrostatic mean.

that just beneath the surface. Let h designate the water depth, L the wave length, T the wave period, and g the acceleration due to gravity. The length L can be eliminated between the equations

$$\frac{\Delta P}{\Delta P_0} = \frac{1}{\cosh\,(2\pi h/L)}, \tag{8}$$

$$\left(\frac{2\pi}{T}\right)^2 = \frac{2\pi g}{L}\tanh\frac{2\pi h}{L}, \tag{9}$$

so that the hydrodynamic filtering effect $\Delta P/\Delta P_0$ can be computed as a function of depth and wave period. For instrument depths (on the bottom) of 40, 150, and 600 ft, the pressure fluctuations for wave periods of 4, 8, and 16 sec, respectively, are reduced to 10 per cent of their surface value ($\Delta P/\Delta P_0 = 0.10$).

Measurements by Seiwell [13] would indicate that equations (8) and (9) are in error by approximately 20 per cent.[6] Ewing and Press [4] have suggested that this discrepancy might be related to the nonrigidity of the sea bottom. Folsom [5] obtains about half this discrepancy in laboratory experiments employing metal wave tanks, thus indicating that the discrepancy might also be related to other factors.

Two methods for determining *wave direction* are being investigated. At the Department of Engineering, University of California, Berkeley, J. D. Isaacs has measured wave direction by means of a Rayleigh disk attached to the underwater unit. The orientation of this disk is recorded ashore by means of a Selsyn motor. In the case of a simple wave train the disk will align itself normal to the plane of orbital motion. In the case of an interference pattern the records are difficult to interpret. At the Scripps Institution, R. S. Arthur is investigating the determination of wave direction from a comparison of the phase relationships of records from two underwater instruments placed roughly parallel to shore. The development of a reliable method for measuring wave direction is an essential requirement in making use of wave records for tracking storms.

Interpretation of Records. In order to overcome the almost prohibitive labor involved in carrying out harmonic analysis by numerical computations, a very effective instrument for frequency analysis has been developed at the Admiralty Research Laboratory in Teddington, England [2]. Figure 5 shows a set of period

6. Seiwell uses the deep-water formula $(2\pi/T)^2 = 2\pi g/L$ rather than equation (9) (see discussion by Ewing and Press [4]).

analyses obtained by the methods discussed above.[7] The first indications of the storm on the wave record were some "low-frequency noises" at 1400, 14 March 1945. At that time visual observations at Pendeen revealed only the presence of 10-sec and 12-sec waves from another source; the long forerunners of the swell were too low to be visible.

The band of gradually diminishing periods was the result of an intense storm which formed on 10 March off the coast of Cuba, 3000 miles from Pendeen, and

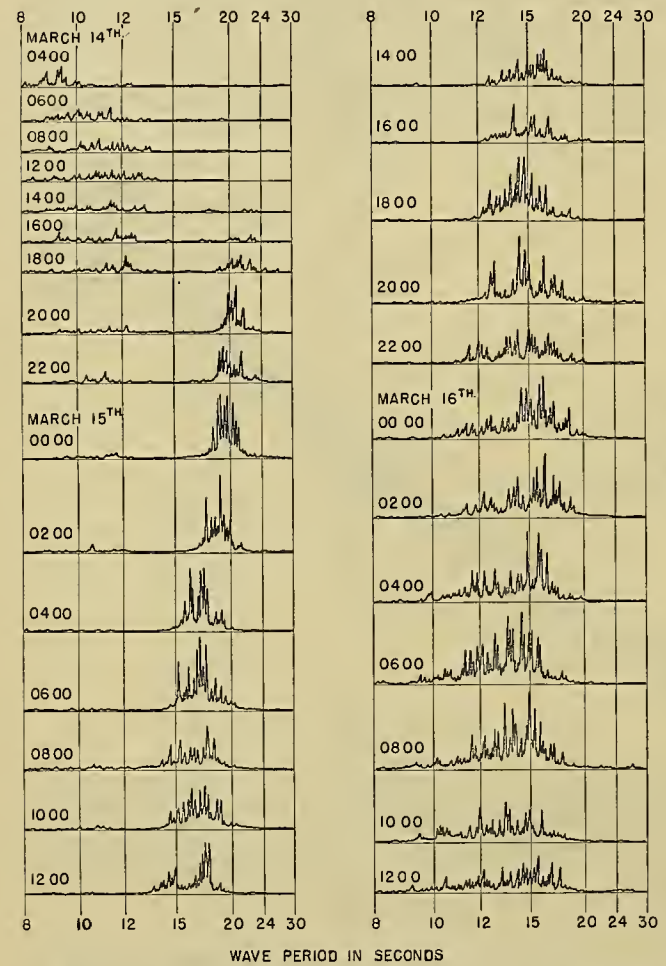


Fig. 5.—Spectrograms of waves recorded during 14–16 March 1945 at Pendeen, England [1]. Each spectrogram gives the distribution of energy among wave periods for a 20-minute record. The records were taken at 2-hourly intervals.

traveled in a general northeast direction, passing west of the British Isles. Four representative weather maps are shown in Figs. 6–9. The positions indicated for the "fetches" were determined according to the rules developed for wave forecasting.[3] It should be noted that the fetch lengthened during the time interval between 1830, 11 March, and 1830, 12 March.

The spectrogram in Fig. 5 can be interpreted in terms of the classical wave theory, according to which component wave trains travel independently with the group velocity appropriate to their period:

$$V = \frac{gT}{4\pi} = \frac{x_w - x_s}{t_w - t_s}. \tag{10}$$

Here x_w and x_s represent the location of a point disturbance and the wave station, respectively; t_w and t_s the time of the disturbance and the time at which waves of period T arrived at the wave station. Equation (10) has a simple geometric interpretation on a time-distance

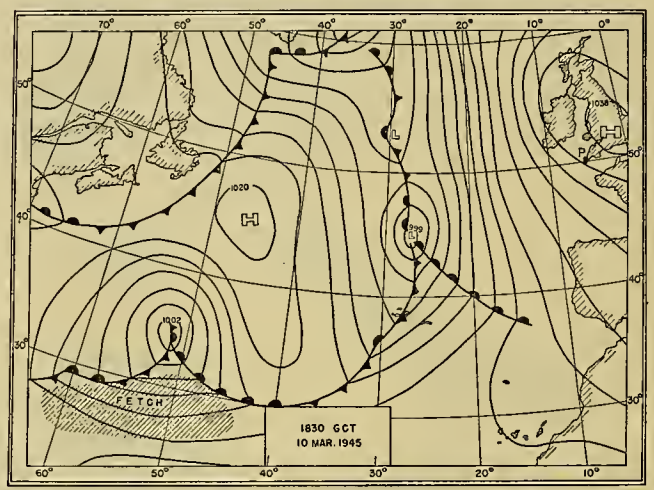


Fig. 6.—Surface weather map for 10 March 1945. Fronts are represented in the conventional manner, and isobars are labeled in millibars. Point P denotes the wave station at Pendeen, England. Figs. 6–9 show the passage across the Atlantic Ocean of the storm for which the wave spectrograms are shown in Fig. 5. (*From Munk* [9].)

diagram (Fig. 10): Rays through the point source represent the paths of wave periods, the value of each period being determined by the slope of the ray. The measurement of two periods, T_1 and T_2, arriving at times $t_{w,1}$ and $t_{w,2}$, suffices to determine distance and time of the source.

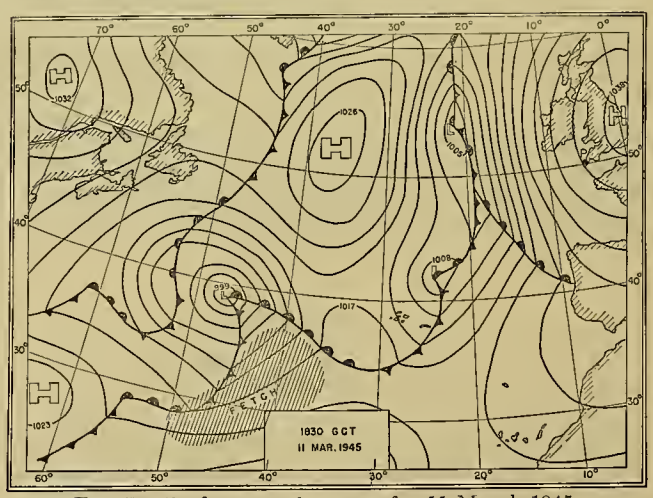


Fig. 7.—Surface weather map for 11 March 1945.

Figure 11 shows the *propagation diagram* used by Barber and Ursell [1]. For each spectrogram (Fig. 5) the maximum and minimum periods were determined, and

7. A different type of analysis based on the autocorrelation function has recently been applied to ocean waves by Seiwell and Wadsworth [14].

lines of appropriate slope drawn. It is apparent from the figure that the foregoing interpretation of the wave observations is consistent with the meteorological observations. The author has found it convenient to use a graph [9, Fig. 5] with a special coordinate system on

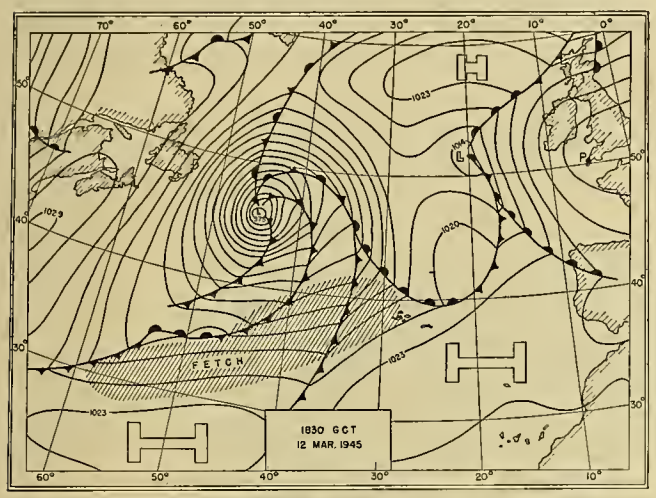


FIG. 8.—Surface weather map for 12 March 1945.

which periods from a single disturbance fall along a straight line. In this manner certain outstanding features on the spectrogram (such as the upper and lower limits) were used to determine the distance and time of the source region with which they are associated.

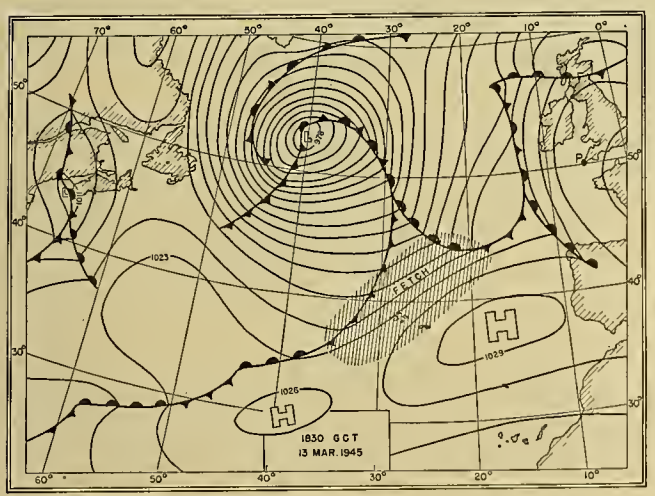


FIG. 9.—Surface weather map for 13 March 1945.

The information taken from the weather maps and the periods recorded at the wave station have been replotted in Fig. 12 to emphasize the similarity, with regard to both shape and width, between the storm path and the period band. The computed positions, designated by A, B, . . . F, agree fairly well with the information on the weather maps, especially if it is remembered that these "foci" were obtained from theory without the introduction of any arbitrary constants. The foci A, B, and C were computed from the lower limit of the period band and seem to correspond to the forward edge of the fetch; D, E, and F were computed from conspicuous features near the upper limit of the period band and give the approximate position of the rear edge of the fetch. It can be seen, however, that focus A in the upper part of Fig. 12 is located 500 miles

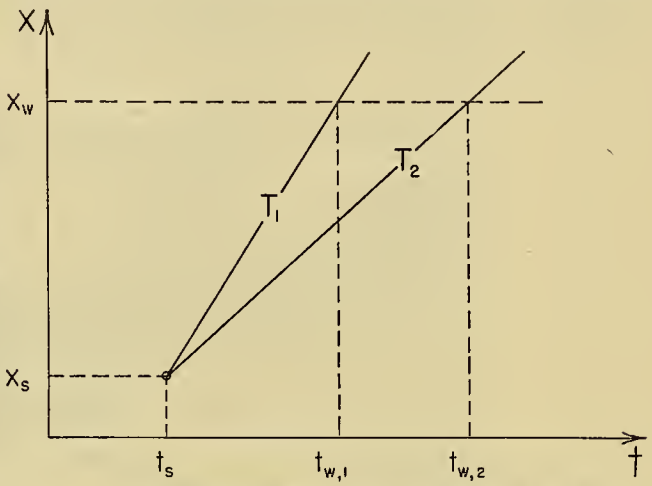


FIG. 10.—Schematic presentation of propagation of wave periods on a time-distance diagram. x_s and t_s represent the position and time of a point disturbance; x_w is the position of a wave-recording station, showing the arrival of waves of periods T_1 and T_2 at times $t_{w,1}$ and $t_{w,2}$, respectively.

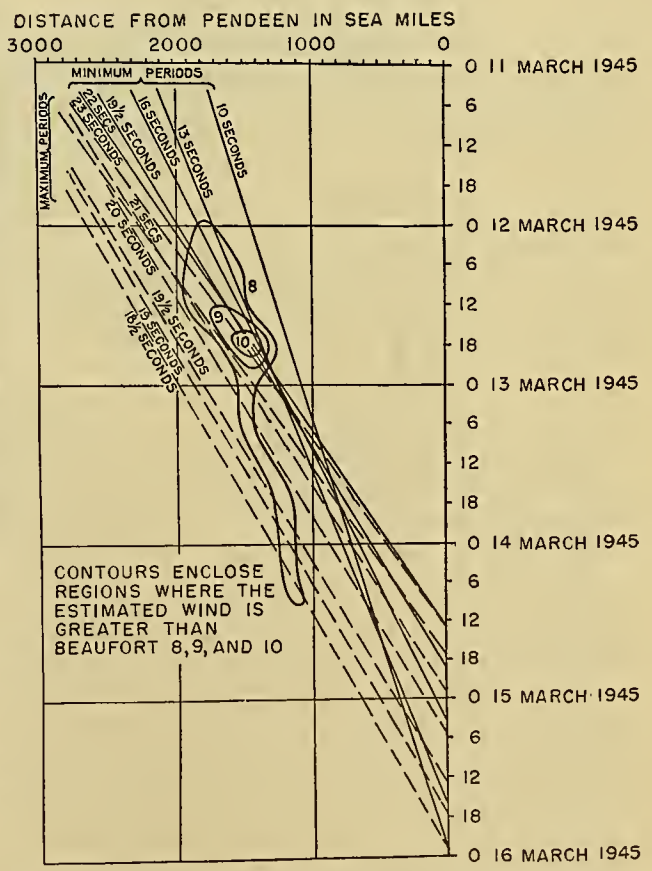


FIG. 11.—Propagation diagram for 11–16 March 1945, using winds estimated from the pressure distribution, flowing within 30° of a line of the recording station. (*From Barber and Ursell* [1].)

behind the storm front, and the other two foci about 200 miles behind the front. These discrepancies must be expected, since waves generated at the front itself could not gather sufficient energy to be perceptible at the wave station; furthermore, the distance between the point of wave origin and the storm front must be larger for the farthest disturbance, which is almost 3000 miles from Pendeen.

Iceland, Pendeen came under the influence of the fetch in the cold sector, which remained at about the same distance from Pendeen. In the meantime the fetch in the warm sector of the second storm moved slowly eastward, and on 18 February this and the fetch remaining from the first storm appeared as a single generation area. By 19 February the first storm had undergone complete frontolysis. The only remaining fetch was that

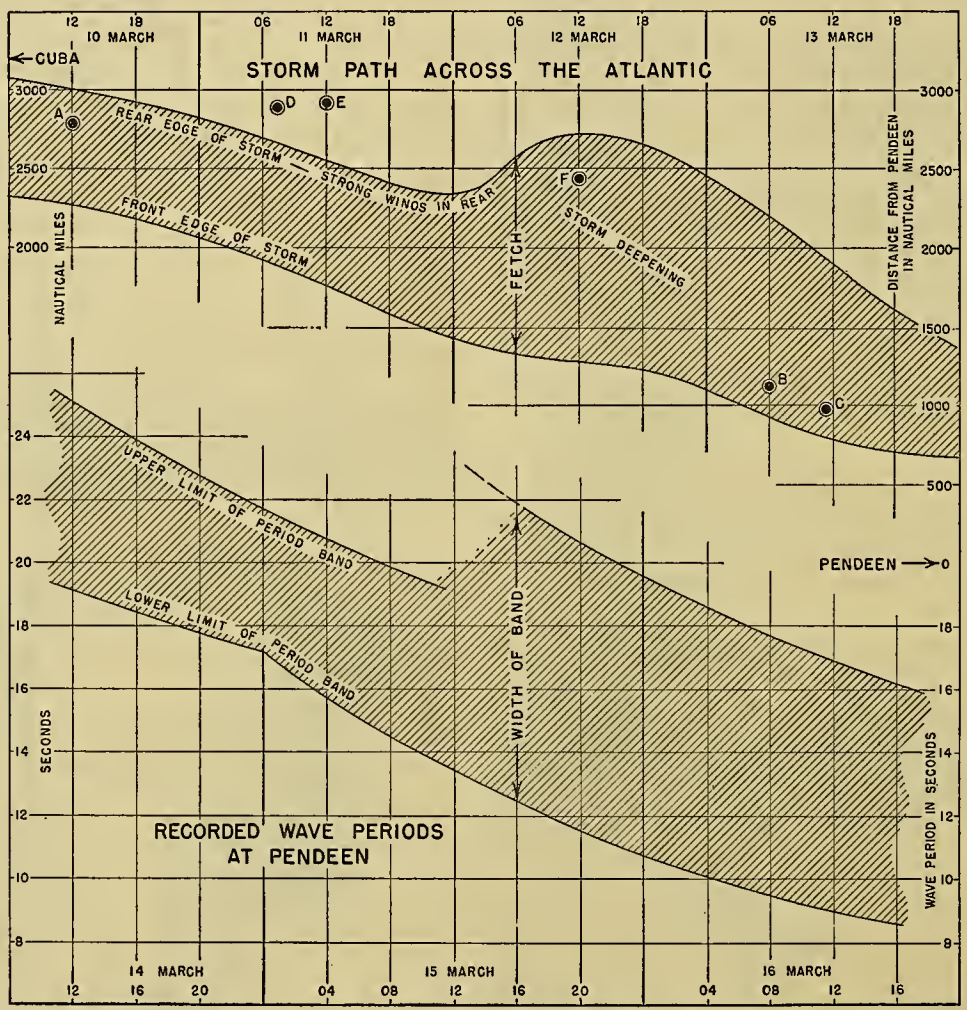


FIG. 12.—Comparison of storm path with period band. The shaded band in the upper part shows the movement of the storm system toward the wave station. The forward and rear edges of the storm were determined by 6-hourly weather maps, of which every fourth map is shown in Figs. 6–9. The computed storm positions (foci) are marked *A*, *B*, . . . *F*. The limits of the period band in the lower part were determined from the wave spectrograms shown in Fig. 5. (*From Munk* [9].)

Figure 13 illustrates the application of the method to a more complex meteorological situation. Two storm systems which existed simultaneously could be identified and traced separately. On 15 February 1945 the development of a cyclone started off the east coast of the United States. On 17 February the center of the cyclone had reached mid-Atlantic, and an unusually well-defined 1000-mile fetch in the warm sector was pointing directly toward the wave station. At the same time a new low-pressure area had just moved off the east coast of the United States and formed a second fetch immediately north of Bermuda (Fig. 13, upper right). As the first storm stagnated and veered northward toward in the cold sector of the second storm, which was moving northward toward Iceland.

The spectrograms shown in Fig. 13 are more complicated than those in the preceding examples, and the interpretation is much more difficult. The first three foci fall in the first storm as it moved across the Atlantic. Foci *D* to *H* are associated with wave trains that were formed as the first storm veered northward and its fetch was joined with the fetch of the second storm. Only by that time had the main-period band moved sufficiently into the lower periods to permit identification of a wave train, *F*, from the distant second storm. Foci *J* to *N* originated from the joint fetch existing from

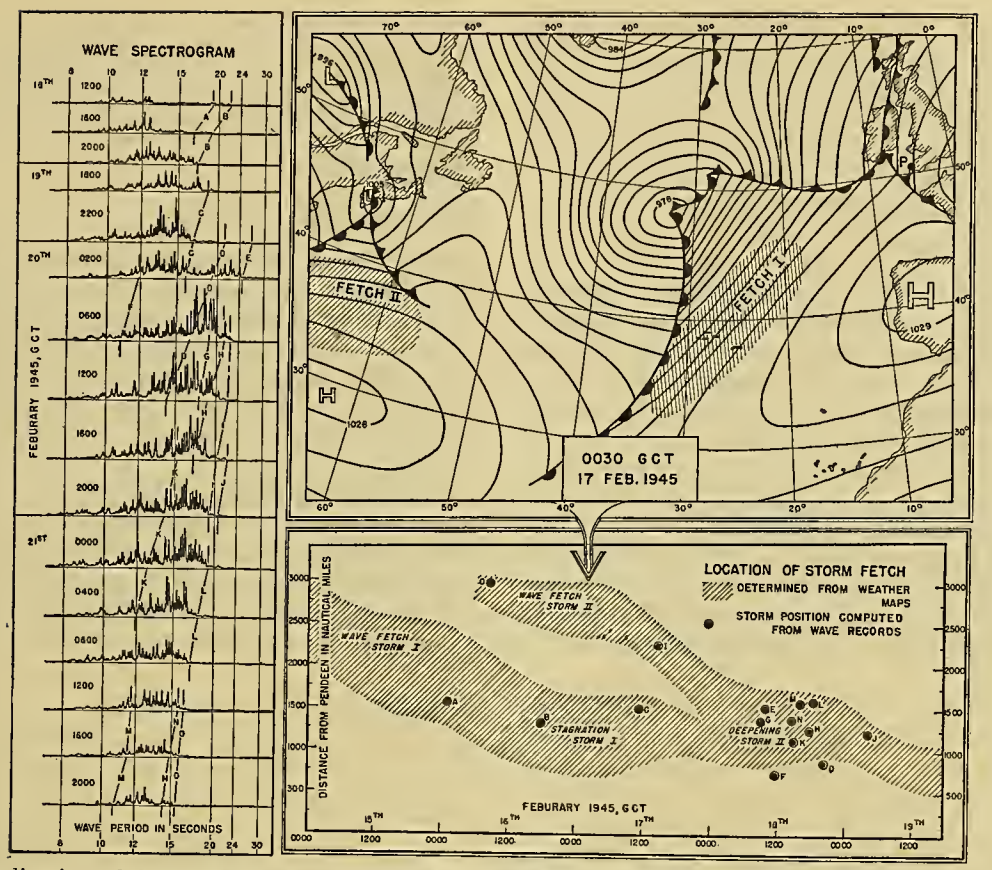


FIG. 13.—Application of storm-tracking method to wave records of 18–21 February 1945. The period spectrograms recorded at Pendeen are shown to the left. Significant features on adjoining records are connected by dashed lines, marked *A, B, . . . O*. The corresponding foci are shown by the circles in the diagram at the lower right, where the shaded band gives the path of the storm determined from weather maps. One of these weather maps is shown at the upper right. (*From Munk* [9].)

18 to 19 February, but focus *O* had its origin in the second storm system when it was still 3000 miles from the wave station.

Discussion. A considerable number of meteorological sequences have been studied by means of the Pendeen wave records, and in all instances the agreement between computations based on the wave records and the information on the weather maps was encouraging. In one instance, 26 June 1945, waves from a hurricane off the coast of Florida were faintly recognizable above the disturbance caused by a moderate local storm. To the author's knowledge, attempts based on wave records taken along the coasts of the United States have not been equally successful. This can be attributed at least partly to instrumental difficulties.

At one time the author had hoped that by shifting the range of maximum response of the instruments toward longer periods (by placing recording units of greater sensitivity in deeper water) it should be possible to record even longer forerunners, with periods, say, up to 60 sec, and correspondingly high velocities. Experience in England [1] does not support this expectation. Figure 14, based on thirteen well-substantiated storms, shows a correlation between maximum wind force of *distant* storms and the maximum recorded swell period. A possible explanation is contained in the forecasting theory [16] according to which waves gain energy from the wind only if the wave age (wave velocity/wind velocity) is less than 1.37. The dashed line in Fig. 14 corresponds to values of the wave age ranging from 1.26 to 2.26. The preliminary conclusion

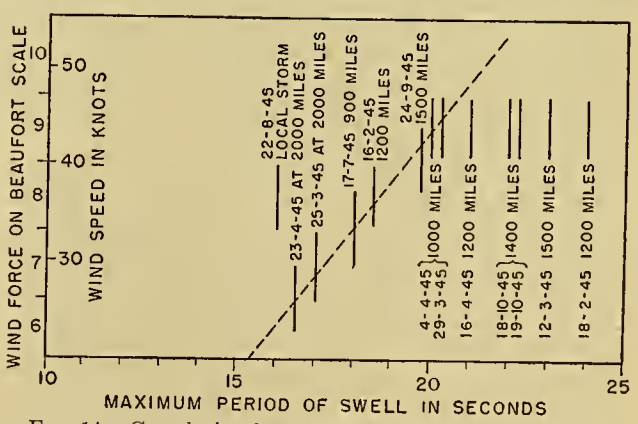


FIG. 14.—Correlation between the maximum wind strength and the maximum period of swell generated for thirteen storms. (*From Barber and Ursell* [1].)

is that fore-runners with periods longer than 25 sec, or perhaps 30 sec, are so low compared to other types of wave motion in that period range [10] that they would be associated with an unfavorable signal-to-noise ratio. Accordingly, the highest group velocities which one may

hope for are of the order of 1000 nautical miles per day. Not until many more situations have been analyzed will an evaluation of the usefulness of the method be possible. The results so far obtained demonstrate the feasibility of locating and tracking storms at distances exceeding 3000 miles and of making rough estimates regarding the size and character of these storms.

The Storm Surge Method

The preceding discussion has indicated that the swell spectrum, visible or invisible, is limited by the maximum velocities of the winds in a storm area. This conclusion is in accord with the hypothesis that swell was originally caused by the normal and tangential stresses applied by the wind to the sea surface, but it does not preclude the existence of other types of waves from storm areas of much longer period than those of the swell. To explore this possibility an instrument [11] has been built which is sensitive to waves whose periods lie between those of the swell and tides. It is essentially a modified tide gauge, and the filtering of the swell and tides is accomplished by means of capillaries and air volumes. The instrument does not require the expensive and troublesome underwater cable and is therefore simpler and much less expensive than the underwater pressure recorder described in the preceding section. One such unit has been installed at the end of the Scripps Institution pier at La Jolla, California; a second unit has recently been installed in Hawaii.

During the spring of 1948 five severe storms passed off the coast of California. In each of the five cases, waves of periods of approximately 15 min (storm surges) were received from 24 to 36 hours prior to the arrival of the heaviest surf and the strongest winds.

Some of the data are summarized in Table II, giving the height H and period T of the storm surges, and also of the swell as it was recorded on the underwater recorder at a depth of 40 ft. The mean height of the storm surges was only 1 in. and it is quite clear that these waves could be detected only by special instruments. According to Table II the earliest arrival of the storm surges preceded the heavy swell by one and one-half days on the average.

It must be emphasized that the study of these storm surges is in a preliminary stage, and that the manner in which they are generated is not at all understood. Should it be possible to use storm surges in locating storms at sea, they would have the advantage of relatively high speed, $\sqrt{gh}$ in water of depth h, or 10,000 nautical miles per day in mid-ocean. It is hoped, therefore, that they may eventually provide a tool to the synoptic meteorologist.

Discussion and Conclusions

The relative merits of the three methods involving the propagation of ocean surface waves are summarized below. We shall also point out some of the disadvantages and advantages of these methods as compared to methods involving the propagation of sound waves through the sea bottom (microseisms) and electromagnetic waves through the atmosphere (radar and spherics).

Instrumentation. The height-period (HT) method requires no special instrumentation. The frequency-spectrum analysis (FA) method requires an underwater recording unit connected by a cable to a shore recorder. Along rocky coasts and coral reefs the chafing of the cable may create quite a serious problem and the use of expensive and unwieldy armored cable may be required. A frequency analyzer must be available at each location if the method is to be used synoptically. The storm surge (SS) method involves a much simpler installation, provided a pier or breakwater is available for securing the instrument.

Storm Location. The bearing and distance of a storm can be located by the HT and FA methods from a single station. In the case of the HT method a visual estimate of wave direction should suffice in view of the uncertainties regarding the determination of distance. The theory underlying the method is insecure, and it is also necessary to make assumptions regarding the duration of the storm. Furthermore, an estimate of deep-water wave height from shore observations may be subject to considerable error. The HT method has the advantage that it permits some estimate of the storm location from a single set of simple observations.

The FA method, on the other hand, depends upon the determination of a *shift* in the frequency spectrum and requires therefore (intermittent) observations extending over a minimum period of perhaps six hours. The determination of distance is based on sound physical theory, and furthermore does not involve the deter-

Table II. Data for Five Severe Storms off the California Coast

Date (1948)	Earliest storm surges			Maximum storm surges			Maximum swell			
	Date and time	H (ft)	T (min)	Date and time	H (ft)	T (min)	Date and time	H (ft)	T (sec)	Warning period (hr)
Feb. 9–10	9:0500	.04	14	9:2230	.12	15.5	10:1500	6	8	34
Feb. 21–23	21:1800	.06	15	22:1100	.15	15.5	23:0400	8	9	34
Feb. 27–29	27:1400	.06	12	28:0600	.12	14	29:1200	4.5	11	46
Mar. 30–31	30:1000	.10	20	30:1600	.14	12	31:1200	7.5	8	26
Apr. 22–24	22:0800	.08	18	22:2200	.09	13	24:0700	9.5	8.5	47
								Mean:		37

mination of wave height. In the determination of wave height the response characteristics of the instrument enter as an important factor, and it may be necessary to take into account, by means of specially constructed "refraction diagrams" [7, 12], the effect of bottom topography and of the configuration of the coastlines. Wave periods, on the other hand, are relatively easy to ascertain from wave records and are not subject to the complex changes in wave pattern that occur as waves come into shallow water. Determination of wave direction by means of a Rayleigh disk or a two-unit wave station are being investigated.

It is not likely that the SS method will make it possible to compute the storm distance from records at a single station. The distance determination for the FA method depends upon the high dispersiveness of ocean swell, and storm surges are almost nondispersive. In view of the great length of storm surges they will also be subject to considerable refraction. The principal hope lies in the determination of the arrival time of identical phases at two or three stations, and the location of storms from special charts based on the propagation at $\sqrt{gh}$ velocity.

Range. It is estimated that the HT method can be used to distances up to 3000 nautical miles, the FA method to distances exceeding 3000 miles. The range of the SS method is not known. The values given above are in excess of those that have been obtained by seismic and electromagnetic methods.

Group Velocity. The group velocities applicable for the HT, FA, and SS methods are of the order of 400, 900, and 10,000 nautical miles per day respectively. The relatively long interval between the time the waves are generated and the time they are recorded at the wave station is a disadvantage in the application of the HT and FA methods. Their practical use will, therefore, be largely to verify or modify the interpretation already made on the basis of meteorological information. In the case of compact meteorological disturbances, such as young hurricanes, these waves may well provide the first clue as to the existence of the storms. The methods should be particularly useful over the southern oceans [3] and for other regions where the network of observing stations is widely spaced. The group velocity of the storm surges, although slow compared to the seismic and electromagnetic waves, is sufficiently large to permit synoptic application.

Storm Intensity. The HT method gives an estimate of the storm intensity. In the case of the FA method an estimate of wind speeds from the maximum recorded period is suggested by Fig. 14. It is also hoped that studies dealing with the propagation of *energy* by the forerunners will lead to an interpretation of the spectrogram *ordinate,* which has received no attention so far but must be related to the storm intensity. The relationship between storm intensity and storm surges has not been studied.

Other Storm Characteristics. The application of the methods employing ocean surface waves is further enhanced by the possibility of determining not only the location and intensity of the storm, but also its size, type, acceleration, and other characteristics. In view of the close "coupling" between the wind pattern and the sea surface waves one might expect a better chance for success by the methods described in this report than by those based on seismic and electromagnetic waves.[8]

Desirable Future Studies. With regard to the HT method the research requirements coincide with those pertaining to the problem of forecasting ocean waves and are discussed in another paper.[3] Of particular importance is a better understanding of the decay of swell traveling through an area of calm or cross winds. In this connection it is planned to establish a wave recorder in the equatorial Pacific, possibly on the Marquesas Islands. These islands lie 2400 miles and about 6 days "up-swell" from California, and a comparison of wave records obtained there with those at Oceanside, California, would give information concerning the attenuation of swell over large distances. It is hoped that these studies may help to bring about, on a scientific basis, a revival of the ancient art of judging weather from the appearance of the sea surface.

The FA method appears to have reached a stage of development where it will be possible to work with records from a network of stations, rather than from a single station. We may expect marked progress as a consequence. Our present experience is almost entirely confined to stations in the eastern Atlantic *toward* which the storms were moving.[9] Development and standardization of suitable recording and analyzing equipment is urgently needed. The determination of wave direction is an essential requirement in making use of wave records for tracking storms.

The SS method is at such an early stage of development that further work along instrumental and theoretical lines is required before it is possible to make any definite recommendations. An improved instrument has been installed at the end of the Scripps Institution Pier, in La Jolla, California, and similar units will be installed during 1950 at Oceanside, California, 30 miles north of La Jolla (for direction determination) and in Hawaii. By means of this network of long-period-wave stations it is hoped that research on the generation and propagation of storm surges can be vigorously pushed.

In conclusion it must be emphasized that it would be most unfortunate if study were to be limited to the three methods with which we have been chiefly concerned here. A systematic exploration of the spectrum of ocean waves is likely to lead to surprising new developments. In this connection it should be remembered that the sea surface wave pattern provides a complete picture, though a distorted one, of the entire wind distribution over the ocean, and our interpretation of this

8. Recent evidence [3] indicates that some of the microseismic activity is caused by the interference pattern of the surface waves in the storm area. Apparently a second-order effect inherent in such a pattern is associated with pressure fluctuations which do not disappear at great depth.

9. Lt. R. L. Miller, USAF, has successfully applied the method to a storm over the Bering Sea, using wave records taken at Guam, located to the *rear* of the moving storm (*Scripps Institution Wave Report No. 90,* unpublished).

picture should improve with the advance of instruments, of theoretical studies, and of empirical experience.

REFERENCES

1. Barber, N. F., and Ursell, F., "The Generation and Propagation of Ocean Waves and Swell; I—Wave Periods and Velocities." *Phil. Trans. roy. Soc. London*, (A) 240: 527–560 (1948).
2. —— Darbyshire, J., and Tucker, M. J., "A Frequency Analyser Used in the Study of Ocean Waves." *Nature, Lond.*, 158: 329–332 (1946).
3. Deacon, G. E. R., "Storm Warnings from Waves and Microseisms." *Weather*, 4: 74–79 (1949).
4. Ewing, M., and Press, F., "Notes on Surface Waves." *Ann. N. Y. Acad. Sci.*, 51: 453–462 (1949).
5. Folsom, R. G., "Sub-surface Pressures Due to Oscillatory Waves." *Trans. Amer. geophys. Un.*, 28: 875–881 (1947).
6. —— "Measurement of Ocean Waves." *Trans. Amer. geophys. Un.*, 30: 691–699 (1949).
7. Johnson, J. W., O'Brien, M. P., and Isaacs, J. D., "Graphical Construction of Wave Refraction Diagrams." *U. S. Hydrogr. Off. Tech. Rep.* No. 2, H. O. Publ. 605, 45 pp. (Jan. 1948).
8. Klebba, A. A., "Details of Shore-Based Wave Recorder and Ocean Waves Analyzer." *Ann. N. Y. Acad. Sci.*, 51: 533–544 (1949).
9. Munk, W. H., "Tracking Storms by Forerunners of Swell." *J. Meteor.*, 4: 45–57 (1947).
10. —— "Surf Beats." *Trans. Amer. geophys. Un.*, 30: 849–854 (1949).
11. —— Iglesias, H. V., and Folsom, T. R., "An Instrument for Recording Ultra Low Frequency Ocean Waves." *Rev. sci. Instrum.*, 19: 654–658 (1948).
12. Munk, W. H., and Traylor, M. A., "Refraction of Ocean Waves: A Process Linking Underwater Topography to Beach Erosion." *J. Geol.*, 55: 1–26 (1947).
13. Seiwell, H. R., "Sea Surface Roughness Measurements in Theory and Practice." *Ann. N. Y. Acad. Sci.*, 51: 483–500 (1949).
14. —— and Wadsworth, G. P., "A New Development in Ocean Wave Research." *Science*, 109: 271–274 (1949).
15. Sverdrup, H. U., and Munk, W. H., "Empirical and Theoretical Relations between Wind, Sea and Swell." *Trans. Amer. geophys. Un.*, 27: 823–827 (1946).
16. —— "Wind, Sea and Swell: Theory of Relations for Forecasting." *U. S. Hydrogr. Off. Tech. Rep.* No. 1, H. O. Publ. 601, 44 pp. (March 1947).

BIOLOGICAL AND CHEMICAL METEOROLOGY

Aerobiology *by Woodrow C. Jacobs* 1103

Physical Aspects of Human Bioclimatology *by Konrad J. K. Buettner* 1112

Some Problems of Atmospheric Chemistry *by H. Cauer* 1126

AEROBIOLOGY

By WOODROW C. JACOBS

Headquarters, Air Weather Service, Washington, D. C.

INTRODUCTION

The field of extramural aerobiology is concerned with the distribution of living organisms by the exterior atmosphere and with some of the consequences of this distribution. It includes within its sphere of interest a study of the dispersion of insect populations, fungus spores, bacteria, viruses, molds, and pollens—in fact, all forms of life, both plant and animal, that are borne aloft and transported wholly or in large part by the atmosphere.

The present-day field of aerobiology had its origin in the pioneer experiments of Spallanzani in 1776, and in the work of Pasteur, Tyndall, and others who used the method of the aerobiologist in combating the theory of the spontaneous generation of life and in developing the germ theory of disease. It has been only within the last fifteen or twenty years, however, that aerobiology has emerged as a specialized field of investigation. An examination of the very extensive literature on the subject accumulated during these later years reveals that activity in the field has been largely confined to the biologist and has seldom included the meteorologist. Few of the biological data appear to have been analyzed from the standpoint of fluid mechanics which leaves the quantitative meteorological approach, based on theory and supported by meteorological-biological observations, as something that remains largely for the future.

The biologist has directed his attention primarily to the problems of entrapment, identification, and enumeration of organisms carried by air, but has more recently become interested in the atmosphere as the medium for the dispersal of the organisms. The meteorologist, on the other hand, has as yet shown comparatively little interest in the results of aerobiological investigation as a possible means for acquiring new knowledge on the nature of mixing processes in the atmosphere and on the movements of air masses. Nevertheless, the ever-increasing concern of the botanist, zoologist, geneticist, agriculturist, allergist, the general public, and, more recently, the militarist, in the results of the atmospheric dissemination of organisms, whether the latter be injurious insects, allergens, or pathogens, shows clearly that the subject is one of great importance and of wide general interest.

BIOLOGICAL FACTORS

Methods and Problems of Sampling and Counting. From the standpoint of the biologist the ultimate quantitative goal of atmospheric biological research is the determination and/or prediction of the type and number of viable organisms suspended in a given volume of air. For those microorganisms which may be readily identified by their shape or superficial surface characteristics and which cannot be grown on culture media, the adhesive-coated slide is widely used as a sample trap. A glass collecting slide is normally exposed horizontally to the air and, although particles are deposited by the combined action of turbulent air motion and by gravity, the technique is commonly referred to simply as the "gravity slide" method of collection. The method has several important disadvantages from the meteorological point of view. First, comparatively long exposures are required—24 hours being the time unit commonly used. The method, therefore, does not afford a means for determining the degree of atmospheric contamination at a given moment or through a given short time interval which, of course, limits the use of the data in the "synoptic" sense. The second, and principal, criticism of the gravity method, however, lies in the fact that the catch does not represent recovery from a definite volume of air. Attempts to convert gravity slide figures into volumetric terms have not been too successful. In spite of the obvious disadvantages of the method, a large proportion of our present statistics on pollen and fungus-spore distribution has been obtained from these gravity slide samples.

For the study of mold spores and bacteria which are too small for convenient direct microscopic examination, it is necessary to examine cultured specimens and count the colonies of each type of specimen. This type of sample is obtained by exposing standard Petri dishes containing suitable culture media to the air for periods ranging from two minutes to one-half hour. This is the so-called "plate method" of collection and suffers from the same disadvantage as the gravity slide technique in that the results cannot be interpreted in volumetric terms. The samples obtained by this method represent organisms released from the air during the brief period of sampling and no record whatsoever is obtained during the long gaps between exposures. The method is also time-consuming and expensive, and to secure reasonably complete data from only a few collecting locations requires the support of a large staff of laboratory technicians.

Because it is necessary to identify and count the microorganisms after they have been captured, it appears that some sort of impingement-slide method of collection offers the greatest promise of adaptability to aerobiological work. For surface collecting the equipment should be economical, simple to operate, and capable of sampling known volumes of air. Above all, the equipment, the techniques of its use, the method of counting, and the units of measure for reporting results

should be standardized among observers and technicians as far as is practicable.

Several types of volumetric sampling devices (such as the Owens dust counter) are available in some quantity, but most of them handle an insufficient volume of air to be adaptable to aerobiological investigation. While Durham [8] has reported the development of an apparatus which impinges the particles from a measured quantity of air on a slowly moving slide, thus affording both a record of organisms and of the volume of air sampled throughout the 24-hour period, the results of collections by this method have not been reported upon in detail. In a study of various types of air filters for trapping microorganisms, DallaValle and Hollaender [5] found that the relative efficiency of a viscous impinging surface tested was approximately 70 per cent. They report that the highly efficient Greenburg-Smith impinger, used as a dust-sampling apparatus by the U. S. Public Health Service, is not particularly adaptable to aerobiological work. It is the opinion of the writer, however, that the efficiency of the collector, itself, is not nearly so important as the condition that the relative efficiency be a known factor and approximately a constant.

For sampling the upper atmosphere the collecting equipment should be compact, light in weight (particularly if it is to be supported aloft by balloon, parachute, or special atmospheric sounding device), and fully automatic in operation; it should be designed to exclude the possibility of contamination[1] and to provide for the multiple sampling of measured volumes of air at known temperatures (and humidities) between specified small pressure intervals (or at known altitudes).

Proctor [23] has perfected an apparatus for upper-air sampling by aircraft which, in addition to securing operation independent of the pilot, simultaneously records certain meteorological data. In addition, variation in air flow is recorded to permit comparative counts in known volumes of air. As in his previous models of collecting instruments, provision is made for photomicrography and subculture of the microorganisms with the exclusion of contamination or cross innoculation of the separate samples.

As would be expected, the problem of collecting and identifying the larger organisms and insects is simpler than the sampling and counting of air-borne microorganisms. The most common methods for collecting these larger forms at or near the surface involve the use of nets, flight or baited traps of some sort, or simply entail the visual counting of insects on the wing or as they alight on some object. None of these methods, however, serve to give the true population of organisms in the air at any given time, although it is possible, under certain conditions, to convert flight-trap collections into rough volumetric terms.

Kites and balloons were naturally the earliest devices used for sampling insect populations in the upper air, but few of the results of collections by either method appear to have been published. The most successful collections appear to have been obtained through the use of aircraft. Perhaps the most extensive study made of the insect population of the atmosphere was conducted at Tallulah, La., during a five-year period between 1926 and 1931 by the Bureau of Entomology and Quarantine of the U. S. Department of Agriculture. The results have been reported upon in considerable detail by Glick [9, 10]. In these collections some 28,739 specimens were taken, from altitudes as low as 20 ft above the ground to as high as 15,000 ft with specially designed traps fitted to the wings of several types of aircraft. The results are given in volumetric terms (average volume of air per insect) although desirable details concerning the testing and calibration of the collecting instrument are lacking. It is unfortunate that similar data are not more generally available for other regions.

Methods and Problems of Identifying Organisms and Locating Sources. Perhaps the greatest difficulty that confronts the aerobiologist in sampling the atmosphere arises from the fact that many species of organisms are so similar in appearance that important differences which serve to separate species defy description or pictorial representation. Even in the cases of some economically important plant pathogens, the spores cannot be identified by morphological characteristics alone. In the cases of other forms, such marked variations in size, shape, and septation exist within the species that identification is impossible unless many spores are examined. Because it is certain that many closely related species within the same genus differ greatly in cultural characteristics, allergenic or biochemical effects and pathogenicity, the solution to the problem of rapid and precise identification of organisms is an essential step toward the full development of an applied aerobiology.

The greatest difficulty that confronts the meteorologist in his investigations of the local and long-distance dissemination of organisms lies in his lack of knowledge concerning the exact (or even approximate) sources of the collected specimens. In the case of the more commonly collected spores and pollens, the parent plants or fungi may be so universally distributed that it is wasted effort even to attempt speculation as to the origin of the air-borne forms. In the case of bacteria in the atmosphere the problem is even more difficult, for here the entire earth's surface (including the oceans) must be considered as a potential source. It is unfortunate for the meteorologist that the organisms most frequently collected and identified are those whose sources are most cosmopolitan.

It is true that the surface of the earth itself can be considered the source of specimens collected in the upper atmosphere but this is far too broad a classification of origin to allow anything beyond gross analyses of the atmospheric factors involved in the vertical and horizontal transport of organisms. By the same line of reasoning, the shore line of a large body of water may be considered a *line source* for land forms

1. The element of contamination is an important consideration when sampling the upper atmosphere, the air over the oceans, or in other situations where population densities of microorganisms are low.

collected over water or for aquatic forms collected inland. The latter technique has been successfully employed by ZoBell and Mathews [40], Rittenberg [26], and Jacobs [14] in their analyses of data on the microbial populations of marine air.

In the investigation of the local dissemination of organisms the problem of identification of origin is usually less difficult. In such cases the collections can frequently be made near or above a source which is readily identified because of a local anomaly in surface cover or because the spore formers or pollinators can be isolated with a reasonable degree of fineness. Proctor and Parker [24] have suggested the artificial introduction into the atmosphere of an easily identified organism for studying dispersion but, because of the rapidity with which small organisms are dispersed throughout large volumes of air and the concurrent impracticability of introducing sufficiently large numbers of organisms to allow reasonable chance for a sample recovery, it appears that such a method would be limited to studies of purely local transport. The use of stained insects has been employed for the latter purpose with some success [30].

From the standpoint of the meteorologist, who is primarily concerned with the atmospheric processes involved in the dissemination of organisms and less interested in the results of such dissemination, it appears that the most pressing need in aerobiologic research is the establishment of a standard list of "biological indicators" or "markers" (as they are called by Stakman [31]). This list should contain properly described and representative spores, pollens, molds, bacteria, and, perhaps, insects that can be easily identified and whose local sources have been determined. The sources represented by the type specimens should be mutually exclusive as to character of surface and/or geographic location. In the case of fungi, considerable care must be exercised in selecting species that produce a vast number of spores in a relatively short time; they should not multiply on dead vegetation or survive in a given region from one year to another [3].

From the standpoint of agriculturists, medical researchers, and workers in related fields, who are more interested in the results of dissemination than in dissemination processes, the primary need in aerobiology is the perfection of techniques for collecting and identifying organisms of pathogenic importance and in the determination of circumstances governing their survival and multiplication after they cease to be airborne. This segregation of interest is meant as no reflection upon the attitude of the meteorologist, for it is taken without question that the physical processes governing the distribution of pathogenic and allergenic forms by the atmosphere are identical to those governing the distribution of the nonpathogenic and nonallergenic species.

The Atmosphere as an Environment for Organisms. From the pathological standpoint the viability of organisms transported by air currents is as important a consideration as is the distance to which they are carried. As stated by Stakman [31], "To show that propagative bodies of pathogens are disseminated by the wind is only the first step in finding out what needs to be known. Are they viable at the end of their trip; how long will they remain viable, even if on the host plant, if conditions are not favorable for germination and entrance into the host soon after their trip; do they belong to a virulent race; is the host plant in susceptible condition; are [meteorological] conditions favorable for rapid multiplication and spread if initial infection occurs? These are only a few of the most important questions that must be answered if the problem is to be studied in its broadest aspects."

It is, first of all, obvious to the meteorologist that the organisms must be able to withstand great extremes of temperature, pressure, humidity, and solar radiation if they are to survive transport over great distances. From the standpoint of temperature it may be argued that because solid particles (thus all organisms) in the air radiate and absorb very nearly as a black body, they must undergo extreme temperature changes which are not represented by temperature changes within the air mass. However, the heat capacity of the organism, or a particle with which it is associated, must be very small, thus it can be assumed that the temperature of the organism at any instant must be very nearly the temperature of the ambient atmosphere. It is therefore doubtful if such organisms are ever subjected to excessively high temperatures. Certain organisms probably succumb to the very low temperatures that exist at high altitudes [34] although there are some forms which survive temperatures approaching absolute zero [39].

Bacteriologists have shown that exposure to ultraviolet radiations for short periods is lethal to most bacteria. In this connection, however, it is of particular interest to note that bacteria which are killed when suspensions (or agar plates streaked with suspensions) are exposed directly to sunlight are the same organisms which have been recovered from air by exposing agar plates for different periods of time. It appears that at least the air-borne parent cells are relatively resistant to sunlight. Whether this resistivity is due to extrinsic or intrinsic conditions is not clear. Some results [29] indicate that certain individual bacteria within a species survive dosages of ultraviolet radiation which are ordinarily lethal to the species, but that the resistivity of the individual bacterium does not persist in the subculture progeny. On the other hand, the writer will show later that many bacteria suspended in the atmosphere must be associated with a dust particle of some sort (such as surface debris or sea salt) or a water droplet, and this particle may afford sufficient protection against the lethal type of radiation. Nevertheless, the bacterial counts should show a diurnal variation due to this cause.

There is the possibility that desiccation of the organism in dry air may render it nonsusceptible to ultraviolet radiations because of its reduced water content. There appears to exist, however, a marked difference of opinion among investigators concerning the effect of humidity. Rentschler [25] concludes from

laboratory experiments that high relative humidity of the air does not increase the resistivity of air-borne bacteria. Wells [35] and Whisler [36], on the other hand, maintain that ultraviolet radiations are ten to twenty times more germicidal in dry air than in humid air. These antithetical results strongly suggest that the differences in germicidal effects were in some way related to differences in the number or character of condensation nuclei present in the several experiments.

In this connection it should be stressed that soil or sea-salt particles play a possible dual role in the maintenance of viability in the organism. In addition to containing aggregates of microorganisms they may, under certain circumstances, act as condensation nuclei (if hygroscopic) for water vapor in the atmosphere and in so doing provide favorable conditions of moisture for the persistence of organisms in the viable state. If this reasoning is valid, then the consideration of variations of atmospheric moisture should be important.

Wellington [34] has reported upon a series of laboratory experiments designed to determine the effects of substantial decreases in pressure, temperature, and humidity upon certain species of insects common in Canada. Attempts were made to simulate those combinations of pressure, temperature, and humidity that an organism might experience upon being lifted from the surface to a height of around 20 km. He concludes that the decrease in atmospheric pressure may be safely neglected as either a limiting or lethal factor among the elementary environmental changes experienced by insects distributed at higher levels; he does show, however, that most of the species tested become insensible at pressures below 120 mb. He found, also, that the flight activity of the insects varied somewhat with pressure, remaining constant for most insects up to simulated altitudes of about 7.5 km, then decreasing to zero at about 16 km. In the cases of *Coleoptera* and *Diptera*, however, there was a distinct increase in activity down to a pressure equivalent to 1.5 km where it again approached normality. This latter finding is significant from the standpoint of the actual vertical distribution of these orders. That the responses are barotactic and not due to distress occasioned by the lowered oxygen pressure was also proved by the experiment. The effects of reduced pressures on microorganisms appear not to have been similarly investigated but, because of their small size and simple internal structure, it is assumed that the effect can be neglected in the cases of these groups also.

There appear to be few, if any, actual quantitative data concerning the fraction of organisms that are able to survive significant transport. Proctor [21], in his investigation of the number of bacteria at high levels, also collected and counted the number of dust particles. He found the ratio of microorganisms (bacteria and molds) to dust particles to be 1 to 108 for all levels and 1 to 118 above 9000 ft. As the writer has previously pointed out [14], this would show about eight per cent fewer bacteria per particle for the higher levels, which might indicate that this proportion was killed, since the physical factors governing their removal from the air would be the same as for the dust particles. Proctor did not give average values for the layer below 9000 ft; it can therefore be assumed that the proportion that did not survive was somewhat greater than eight per cent.

METEOROLOGICAL FACTORS

The Mechanics of Exchange of Organisms Between Earth and Atmosphere. It is assumed that nearly all surfaces, whatever their nature or composition, will contribute organisms to the atmosphere; but that some surfaces, because of greater populations of organisms, more extensive vegetative cover, greater mobility of surface materials, or more intensive atmospheric turbulence, will contribute far more than others. In the case of bacteria, it is a statistically remote possibility that any single organisms in the soil will be lifted by themselves into the air; in all probability they are most frequently associated with debris of some kind, such as bits of organic matter, dust motes, or, in the case of marine forms, with salt particles or droplets of concentrated sea water. These organisms must therefore exist most frequently as colonies rather than as individuals. The laboratory method of plating and counting the bacteria and molds permits of no distinction between individuals and groups. Because the number of bacteria in a cubic centimeter of soil may be numbered in the millions or hundreds of millions, the potential supply of these organisms must be considered almost unlimited.

By this reasoning, the processes governing the exchange of soil bacteria between the earth and atmosphere are the same as for the exchange of other terrigenous materials. The quantity of material exchanged will be greatest in those areas where ample supplies of loose particles exist on the surface, surface winds are strong enough to stir up these materials, and steep lapse rates exist in the atmosphere to favor the vertical transport of the particles. The supply of microorganisms should be greatest where the surface is dry due to deficient rainfall, where there is an abundance of organic matter in the soil, where intensive cultivation is practiced, and in and near wooded areas. A spring maximum of dust in the atmosphere is noted in the solar radiation measurements in the United States. This is to be expected since the soil is driest at this season, cultivation is most extensive, surface winds are strongest, and the vertical temperature lapse rates are steepest. It can be assumed that the population of soil microorganisms in the atmosphere will be greatest during this season for the same reasons.

It has previously been pointed out that Proctor [21], during his determinations of the number of bacteria at high levels, also collected and counted the number of dust particles. Although he made no volumetric calculations, rough computations by the writer indicate that he collected approximately 510 visible and collectable dust particles per cubic meter of air averaged for all levels. Above 9000 ft there were approximately 25 per cent fewer particles than the general average. His ratio of 1 microorganism to 108 dust

particles as an average for all levels gives a figure of only 4.7 collectable bacteria (or molds) per cubic meter of air. If one considers the great number of bacteria that must exist in the lowest layers of air, these results show that either a very small percentage of the organisms survive or that the original number is dispersed rapidly throughout exceedingly large volumes of air. It is, perhaps, also possible that the processes governing the removal of organisms from the atmosphere are more effective than has previously been assumed.

The processes governing the exchange of spores and pollens between the earth and atmosphere are not wholly analogous to those governing the exchange of bacteria and molds between soil and air. The pollens and spores of fungi present a wide range of types with respect to their morphology, physiology, and mode of production and liberation [15]. Consequently they vary greatly in adaptation to aerial dissemination. In the case of certain fungi, very elaborate mechanisms exist for the forcible ejection of the fungus spores from their sporophores or spore mother-cells into the air. The force with which the spores are ejected varies greatly with different species, the most common ejection distances being of the order of one or two millimeters to several centimeters. Although many fungi produce an almost incomprehensibly large number of spores,[2] none of the latter appear to be provided with special adaptations for flotation as is the case with some seeds and insects.

Of the various classes of organisms so far discussed, only the insects are capable of aerial locomotion. Their initial presence in the lower layers of the atmosphere and, to some extent, their vertical distribution and local dissemination are due to the flight activity of the insects themselves. Even here, however, the activity of the insect is dependent in some degree upon such meteorological factors as temperature, wind, and humidity [10, 20, 33, 37]. Some nonflying insects are specially adapted to aerial dissemination, good examples being several species of spiders whose young spin webs from some elevated object to be later carried with the wind. Glick reports that at times great masses of such webs are found floating in the air, even at altitudes of 7000 to 10,000 ft.

A large number of seeds are particularly well adapted to aerial dissemination through very effective structures that increase their frictional resistance to air. Because of the size of the seeds and their very common occurrence, knowledge regarding them is more general than is true of smaller organisms.

Viruses which are produced within the tissues of plants or animals do not appear particularly well adapted to aerial dissemination except as they may be carried by insect vectors or birds. There is the possibility, however, that such virus material may occasionally be associated with bits of organic matter or microorganisms in the atmosphere. Additional research on the possible (free) aerial dissemination of virus substances is needed.

In the case of marine bacteria, it is only when the sea surface is stirred sufficiently to produce spray that a mechanism exists for the introduction of the sea-surface organism into the atmosphere. The spray droplets will, of course, include any bacteria present in the sea water and subsequent evaporation of the droplet will leave the organism associated with a salt particle or droplet of concentrated sea water. Since such particles are extremely effective nuclei for the condensation of water vapor in the atmosphere, they are more readily removed than are the nonhygroscopic dust particles which may contain soil bacteria. Since the number of marine bacteria in a cubic centimeter of sea water seldom exceeds 500, it can be computed (on the basis of data concerning evaporation from spray)[3] that the populations of marine bacteria in the air must be sparse everywhere except, perhaps, at times of rough sea or in the vicinity of a coastal "breaker zone."

An understanding of the physical processes which serve to remove organisms from the atmosphere once they have become air-borne is as important to the aerobiologist as is a knowledge of the mechanism governing the lifting of the organisms into the atmosphere in the first place. Most investigators in the field of aerobiology have taken great pains to point out the usual small size and mass of air-borne organisms, and data concerning the computed or observed free rates of fall under the influence of gravity (in still air) appear in almost every analytical paper.[4] The present author has no desire to minimize the importance of small size and mass in furthering the dissemination of microorganisms for those are their primary adaptations to aerial transport. It should be pointed out, however, that of all the forces tending to move the single microorganism vertically, the gravitational factor is merely another component of force additive to others usually of greater magnitude. The possibility that a single pollen grain, spore, or bacterium carried through convection to an altitude of several kilometers will ever be allowed to settle back to earth through the effects of gravity alone is exceedingly remote. Some organisms of near colloidal dimensions could be considered to remain almost permanently in the atmosphere were they not brought down to the surface again through turbulence or through capture by condensation or precipitation products.

Many investigators have made note of the fact that rainfall is an extremely effective mechanism for clearing

2. Christensen [3] cites as an example of the remarkable power of reproduction of many plant pathogens, that *Ustilago zeae* (corn smut) in two weeks may produce a gall of 20–40 cubic inches, each cubic inch of which contains about six billion spores. One acre of corn with 10 per cent infection would produce 5×10^{13} spores during the same period.

3. On the basis of determinations of the sea-salt content of marine air [13], the author has previously computed that the number of marine bacteria near the sea-surface source averages about five per cubic meter of air [14].

4. The free rates of fall of the largest fraction of bacteria, spores, and pollens (at sea-level air densities) are within the range of 0.1 to 20 mm sec^{-1}. In some cases the velocities approach those predicted by Stokes' law and in other cases substantially reduced velocities are indicated.

the atmosphere of organisms. The comparatively large number of microorganisms found in rain water by various investigators is further evidence that at least a sizeable fraction of the air-borne organisms have been captured and removed from the atmosphere by falling raindrops. While it has not been definitely proved that microorganisms ever act directly as condensation nuclei in the atmosphere, at least one investigator has noted that the rate of fall of spores in moist air is several times greater than in dry air, presumably because the organism has increased its mass through the absorption of water vapor.

It would thus appear that the condensation and precipitation of water vapor are the important mechanisms for removing microorganisms from the atmosphere. The transport to the surface through the effects of turbulence with subsequent settling or impingement appears to be the secondary process. This reasoning, of course, does not apply to the larger organisms whose free rates of fall are appreciable and which are supported aloft only within strong convective currents or through their own flight activity.

Vertical and Horizontal Transport of Organisms. Turbulence in the atmosphere tends to create a uniform distribution of properties throughout the air mass. This action, however, is of far greater magnitude than the mere diffusion of properties between layers; apparently the mixing takes place through the motions of eddies or discrete parcels of air which are pushed from their original surroundings at irregular intervals. Thus, it would be expected that air-borne microorganisms, whose source is at the surface of the earth, would be distributed throughout the entire depth of at least the troposphere,[5] but that the density of organisms would decrease the greater the distance from the source. Biological work to date shows this to be true. Proctor [21] notes a general decrease in both dust particles and bacteria with elevation and an increase in the bacterial count in the vicinity of wooded areas, while Rittenberg [26] has presented data which show a general decrease in the mold count over the oceans with increasing distance from shore. It would appear that the horizontal distance over which the individual microorganism may be transported is almost limitless and is largely determined by its ability to survive the atmospheric environment.[6] The meteorologist, of course, is interested in the total distribution of microorganisms; the biologist, on the other hand, concerns himself only with what he terms the *effective* distribution. The effective distribution is a function of the virulence and concentration of organisms and of the conditions bearing upon their multiplication and spread at the *place* and *time* of deposition.

As an example of the sometimes complex meteorological circumstances that must precede the effective long-distance dissemination of important plant pathogens, Stakman and Christensen [32] cite the spread of stem rust on wheat through the North Central States during 1942 and 1943. In both years, barberry bushes (the secondary host) in the Virginias became heavily infected about May 15 because of favorable weather circumstances during the preceding period of development. Throughout the first half of June, southeast winds, accompanied by rains (favoring multiplication), carried the rust spores west-northwestward, causing heavy infection through much of Ohio and Indiana and as far as southern Michigan and northeastern Illinois. While the south-to-north air movement was not unusual, the destructive epidemics resulted only when all meteorological factors favorable for rust development operated in conjunction.

Supporting evidence on the long-distance dissemination of plant pathogens is given by data collected in Canada which show that urediniospores of leaf and stem rust of wheat are regularly introduced each spring into Manitoba by spores from the United States. These rusts do not overwinter in Canada, but the urediniospores are invariably found in the air in advance of the infection of the wheat crop. Of even greater interest is the fact that the rusts do not survive the hot summers of the south and every fall must be reintroduced by wind-borne spores from the north. The meteorologist cannot help but note how the maintenance of such a biological cycle is favored by the normal seasonal change in the general atmospheric circulation over these areas.

Numerous similar examples of the long-distance dissemination of organisms appear in the very extensive literature on aerobiology. A large fraction of the papers contain aerobiological data which should be further analyzed by the meteorologist.

Although turbulence tends to create uniformity of properties throughout an air mass, at any given instant this air mass must contain variously large and small masses (eddies) with somewhat different properties from those of the surrounding air. This no doubt accounts for at least part of the extreme variability in bacterial counts within small areas, or over short intervals of time, which has been noted by various investigators. The more intense or prolonged the turbulent action, however, the more nearly will uniformity of properties be approached. At the same time, however, there must be a general decrease in the density of the property relative to the total air mass as the distance between the observation and the source area increases.

If this "uniformity of properties" were actually attained in a given region (which would be the case were the process one of pure diffusion), the spatial distribution of microorganisms would approximate a frequency distribution of the Poisson type, and the ratio of the variance (*i.e.* mean squared deviation) of the counts to the mean of the counts made in the region should approximately equal 1.00. A higher ratio indi-

5. Living spores of several common molds were caught in a spore trap released from the balloon *Explorer II* at 72,500 ft and set to close at 36,000 ft [27].

6. Living spores of various fungi have been caught from aeroplanes above the Caribbean Sea 600 miles from their nearest possible source [17], while allergens have been identified at least 1500 miles from their probable origins. In spite of the usual low concentrations at the source, marine bacteria have been collected 80 miles inland from the nearest seacoast [40].

cates the presence of eddy masses which still retain the properties derived from their sources. However, the longer the elapsed time since such an eddy mass left its source, the more completely will its population become equalized with that of its surroundings because of diffusion and small-scale turbulence; hence, for any class of microorganisms, the ratio referred to above should diminish with distance from the source.

The writer has previously analyzed data obtained by Rittenberg [26] on the distribution of molds[7] over the ocean from shore to a distance of 400 miles westward from the Southern California coast. The mold samples represent collections made on shipboard at oceanographic stations at approximately equal distances along courses normal to the coastline. The region off the coast of California presents unusually stable conditions throughout practically all months, therefore it is not expected that the mixing processes are as effective here as in some other areas. Nevertheless, the data accumulated by Rittenberg show the reasoning in the preceding paragraphs to be valid. His collections give an average mold count per hour of 155.6 for the region from 0–100 miles off the coast and a count of 29.0 for the region from 100–400 miles off the coast. If the mean of the differences in count between individual collecting stations is taken as a measure of the variability, a value of 250.1 molds per hour is found for the coastal region and 30.8 molds for the region beyond the 100-mile zone. These figures are interesting in that they give ratios between the mean differences in count and mean counts of 1.61 for the first region and 1.06 for the second. However, the actual mixing processes are probably best represented if the ratio of the variance to the mean of the counts is taken. In this case, the ratio is 402.0 in the first region and 32.7 in the second. As suggested, if the mixing were accomplished solely by pure diffusion, the ratios would approach 1.00 in both areas. Approximately the same number of counts were obtained in each of the areas under consideration.

One may speculate as to the origin of the molds. The prevailing winds in the Southern California area are from the sea, and thus the air masses usually present over the coastal region are essentially marine in character and may have passed over several thousand miles of sea surface without contact with land. Since the mold collections were made by Rittenberg at intervals of varying length and on several cruises, it may be assumed that a large fraction of the counts were made within truly marine air masses. If the molds are primarily of land origin, they must have originated either over the continental land masses to the east or over the far removed land areas bordering the North Pacific Ocean on the west or north. The fact that Rittenberg notes a decrease in the mold counts from shore seaward appears to rule out the latter possibility which indicates that these organisms actually originated from the land surface of the United States but were transported westward across the prevailing wind stream through lateral mixing. Whether the effect is the result of a more or less continuous mixing process or is brought about by sporadic incursions of continental air masses over the ocean is a matter for further speculation.

Considerations such as these suggest the use of bacterial counts in the identification of air masses and in the study of large-scale mixing processes in the atmosphere. The identification of the origins of species of organisms found in the atmosphere should give information concerning the origin of the air masses, while the study of the ratios (at the surface and aloft) between organisms representing two or more source areas should give additional information concerning the nature and effectiveness of lateral and vertical mixing. Few of the available aerobiological data, however, are complete enough to allow this type of analysis.

Many observers have reported great irregularities in concentrations of organisms collected at different levels in the atmosphere, often finding heavier concentrations at higher altitudes than at lower levels. In some cases the investigators have reported encountering clouds of spores or bacteria several thousand feet above the surface of the earth, seemingly borne along by the wind as more or less discrete, sharply bounded units. However, it should be pointed out that such collections have been obtained through the use of aircraft, and it is quite probable that in many cases the variations in samples merely represent the passage of the aircraft from and into convective (rising) columns of air which would be expected to have greater populations of organisms than the surrounding (and perhaps descending) portions of the atmosphere. Wellington [34] has suggested the use of the helicopter for the purpose of "selective sampling"—a technique which could be used to obtain more meaningful data on the vertical distribution of organisms in the atmosphere.

Several investigators have noted that increased counts of microorganisms are obtained in airplane collections upon entering clouds and have ascribed the condition to the fact that the cloud is a more favorable environment for the organisms than is clear air, although Durham [7] ascribes the increase to the loss of electrostatic charge and the consequent dislodging of particles which adhere to the plane. It does not seem to have occurred to the investigators that the vertical cloud boundary probably also represents the vertical boundary of a rising air column which should contain a heavier concentration of organisms than the surrounding clear (nonconvective) area.

The discussion so far is intended to indicate the importance of the horizontal and vertical mixing in distributing organisms throughout the atmosphere. The height to which the organisms may be transported depends upon the degree of stability (or instability) near the surface and upon the height of the turbulent or convective layer of air. The presence of a stable layer at the surface will prevent or retard the introduction of surface organisms into the upper atmosphere but will,

7. Rittenberg assumes that all the molds are of land origin regardless of whether they develop on sea-water media or tap-water media, although it is admitted that at least a portion of the molds collected may have had a temporary sojourn in the sea.

at the same time, maintain higher concentrations of organisms in the surface layers; the presence of a discontinuity surface in the upper air will limit vertical transport in either direction resulting in the concentration of organisms above or below such a surface. There are several recorded instances of such biologic stratification. For example, Durham [8] has noted that abrupt spore ceilings are often marked by a cloud layer or by a visible haze line. He noted one instance in which a detached cloud of fungus spores, almost 100 per cent *Alternaria*, was encountered about 1000 feet above a visible 4000-ft haze line that marked the ceiling for ragweed and the ground cloud of other spores. An analysis of the meteorological conditions that existed near New Orleans on September 12, 1940 (the date and place of collection), might prove interesting, particularly if the source of the "cloud" of *Alternaria* could be identified.

Nearly all investigators who have studied the vertical distribution of organisms in the atmosphere have noted the importance of thermal convection in increasing the populations at higher altitudes. Thermal convection over a region is a diurnal phenomenon and is usually at a maximum in the early afternoon; as a result, the maximum altitude populated by organisms should be highest at this time of day. Nearly all the data on insect populations verify this conclusion.

In general, the horizontal transport of organisms will be determined by the general stream flow in the atmosphere, but lateral mixing will transport the organisms laterally away from the trajectories indicated in the general wind observations. In temperate latitudes, the greater transport will be from west to east and, considering the greater carrying power of polar air masses, from north to south. As an example of the latter, Durham [8] calls attention to a remarkable mold-spore shower that occurred October 6–8, 1937, over the eastern half of the United States. During this period, extremely high slide counts of *Alternaria* and *Hormodendrum* were recorded at all collection points between southern Minnesota and Louisiana with somewhat increased counts in the tier of states between North Dakota and Texas. He presents a map of the area which shows, for each of the three days, the line along which the concentration of the organisms was a maximum. The writer has referred to the synoptic charts for those dates and finds it interesting to discover that Durham's lines of maximum spore concentration on each date coincide almost exactly with the daily positions of an extensive and rapidly-moving dry cold front.

A number of investigators have attempted to describe mathematically the vertical and horizontal distributions of organisms as functions of space or time but none appear to have been too successful. Wolfenbarger [38] has prepared regression formulas for the plots of a very large number of series of data on the spatial variations in organism counts; all are based on the form

$$E = a + b(\ln x), \text{ or}$$

$$E = a + b(\ln x) + c(1/x),$$

where E is the calculated number or concentration of organisms, a is the position on the graph and b and c are factors determining the slope of the curve as influenced by the logarithm or reciprocal of the distance x from the source. He presents the completed formulas for 251 separate plots of data. For example he finds, from one set of data, that the number of codling moths caught at a distance of x feet from an infested orchard can be given approximately by the expression:

$$E = 5.2904 (\log x) + 20114.0849 (1/x) - 18.0628.$$

Since the formulas nowhere take into account meteorological conditions or other factors governing distribution, the writer fails to appreciate the great amount of effort that must have gone into their preparation.

CONCLUDING REMARKS

No attempt has been made in these pages to cover in detail the list of specific possibilities for research in the field, either from the standpoint of the biologist or of the meteorologist. The principal gaps in the research programs conducted by aerobiologists have been pointed out, and suggestions for general lines of research have been offered in the several sections. The writer hopes that the discussion will serve as an added stimulus toward future investigations in the field by both groups and that it will, in some measure, give guidance to the biologist. It is the problem of the biologist to develop certain and rapid methods for identifying the organisms with respect to type and place of origin; it will be the problem of the meteorologist to apply the results of the biological investigations to a study of the atmosphere.

The author is indebted to Drs. B. E. Proctor, E. C. Stakman, J. J. Christensen, R. P. Wodehouse, G. W. Keitt, and P. A. Glick for their assistance in locating recent materials on the subject of aerobiology and to Mrs. Marguerite Gleeck of the Air Weather Service, for her careful reading of the manuscript. The writer also wishes to acknowledge his extensive use of the volume *Aerobiology*, published by the American Association for the Advancement of Science, as a source for biological information.

REFERENCES

No attempt is made here to cover the list of important papers concerned with aerobiological subjects. Care has been taken to include the more recent titles, those which, themselves, contain extensive bibliographies, and all titles which have been specifically referred to in the text. A valuable bibliography, which was published after this list of references was prepared, is that by Miss H. P. Kramer, "Cumulative Annotated Bibliography on Aerobiology and Its Applications to Meteorology." *Meteor. Abstr. & Bibliogr.*, 1: 119–135 (1950).

1. Berland, M. L., "L'exploration biologique de l'atmosphère en avion, et l'emploi possible de cette méthode en météorologie." *La Météor.*, Ser. 3, No. 1, pp. 28–35 (1936).
2. Chester, K. S., "Airplane Spore Traps for Studying the Annual Migration of Wheat Rust." *Proc. Okla. Acad. Sci.*, 19: 101–104 (1939).
3. Christensen, J. J., "Long Distance Dissemination of

Pathogens" in *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 78–87, 1942.

4. Committee on Apparatus in Aerobiology, Nat. Res. Comm., "Techniques for Appraising Air-Borne Populations of Microorganisms, Pollen and Insects." *Phytopathology*, 31: 201–225 (1941).
5. DallaValle, J. M., and Hollaender, A., "The Effectiveness of Certain Types of Commercial Air Filters against Bacteria." *Publ. Hlth. Rep., Wash.*, 54: 695–699 (1939).
6. Darling, C. A., and Siple, P. A., "Bacteria of Antarctica." *J. Bact.*, 42: 83–98 (1941).
7. Durham, O. C., "Methods in Aerobiology." *J. Aviat. Med.*, 12: 153–161 (1941).
8. —— "Air-Borne Fungus Spores as Allergens." *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 32–47, 1942.
9. Glick, P. A., "The Distribution of Insects, Spiders, and Mites in the Air." *U. S. Dept. Agric. Tech. Bull.* No. 673 (1939).
10. —— "Insect Population and Migration in the Air." *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 88–98, 1942.
11. Gregory, P. H., "The Dispersion of Air-Borne Spores." *Trans. Brit. mycol. Soc.*, 28: 26–72 (1945).
12. Humphrey, H. B., "Relation of Upper-Air-Mass Movement to Incidence of Stem Rust." (Abstract) *Phytopathology*, 28: 10 (1938).
13. Jacobs, W. C., "Preliminary Report on a Study of Atmospheric Chlorides." *Mon. Wea. Rev. Wash.*, 65: 147–151 (1937).
14. —— "A Discussion of Physical Factors Governing the Distribution of Microorganisms in the Atmosphere." *J. mar. Res.*, 2: 218–224 (1940).
15. Keitt, G. W., "Local Aerial Dissemination of Plant Pathogens." *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 69–77, 1942.
16. McCubbin, W. A., "Relation of Spore Dimensions to Their Rate of Fall." *Phytopathology*, 34: 230–234 (1944).
17. Meier, F. C., "Collecting Microorganisms from Winds above the Caribbean Sea." (Abstract) *Phytopathology*, 26: 102 (1936).
18. —— "Effects of Conditions in the Stratosphere on Spores of Fungi." *Nat. Geogr. Soc., Contrib. Tech. Papers, Stratosphere Ser.*, No. 2, pp. 152–153 (1936).
19. Pady, S. M., Kelly, C. D., and Polunin, N., "Arctic Aerobiology. II—Preliminary Report on Fungi and Bacteria Isolated from the Air in 1947." *Nature*, 162: 379–381 (1948).
20. Pratt, R. L., *Weather and Alaskan Insects*. Environmental Protect. Sec., Off. QMG, Lawrence, Mass., Rep. No. 156, 25 pp., mimeogr., August, 1949.
21. Proctor, B. E., "The Microbiology of the Upper Air. I." *Proc. Amer. Acad. Arts Sci.*, 69: 315–340 (1934).
22. —— "The Microbiology of the Upper Air. II." *J. Bact.*, 30: 363–375 (1935).
23. —— and Parker, B. W., "Microbiology of the Upper Air. III—An Improved Apparatus and Technique for Upper Air Investigations." *J. Bact.*, 36: 175–185 (1938).
24. —— "Microorganisms in the Upper Air." *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 48–54, 1942.
25. Rentschler, H. C., "Bactericidal Ultraviolet Radiation and Its Uses." *Trans. Illum. Engng. Soc., N. Y.*, 35: 960–975 (1940).
26. Rittenberg, S. C., "Investigations on the Microbiology of Marine Air." *J. mar. Res.*, 2: 208–217 (1940).
27. Rogers, L. A., and Meier, F. C., "The Collection of Microorganisms above 36,000 Feet." *Nat. Geogr. Soc., Contrib. Tech. Papers, Stratosphere Ser.*, No. 2, pp. 146–151 (1936).
28. Schrieber, E. A., "1946 Aerobiological Survey for Ann Arbor, Michigan." *Univ. Hosp. Bull. Ann Arbor*, 13: 35–37 (1947).
29. Sharp, D. G., "The Effects of Ultraviolet Light on Bacteria Suspended in Air." *J. Bact.*, 39: 535–547 (1940).
30. Smith, G. E., Watson, R. B., and Crowell, R. L., "Observations on the Flight Range of *Anopheles quadrimaculatus Say*." *Amer. J. Hyg.*, 32: 102–113 (1941).
31. Stakman, E. C., "The Field of Extramural Aerobiology." *Aerobiology* (publ. by Amer. Assoc. Adv. Sci.), No. 17, pp. 1–7, 1942.
32. —— and Christensen, C. M., "Aerobiology in Relation to Plant Disease." *Bot. Rev.*, 12: 205–253 (1946).
33. Van Overeem, M. A., "A Sampling Apparatus for Aeroplankton." *Proc. Acad. Sci. Amst.*, 39: 981–990 (1936).
34. Wellington, W. G., "Conditions Governing the Distribution of Insects in the Free Atmosphere." Parts I–III. *Canad. Ent.*, 77: 7–15; 21–28; 44–49 (1945).
35. Wells, W. F., "On Air-Borne Infection. II—Droplets and Droplet Nuclei." *Amer. J. Hyg.*, 20: 611–618 (1934).
36. Whisler, B. A., "The Efficacy of Ultra-Violet Light Sources in Killing Bacteria Suspended in Air." *Iowa St. Coll. J. Sci.*, 14: 215–231 (1940).
37. Wolf, F. T., "The Microbiology of the Upper Air." *Bull. Torrey bot. Cl.*, 70: 1–14 (1943).
38. Wolfenbarger, D. O., "Dispersion of Small Organisms." *Amer. Midl. Nat.*, 35: 1–152 (1946).
39. ZoBell, C. A., "Microbiological Activities at Low Temperatures with Particular Reference to Marine Bacteria." *Quart. Rev. Biol.*, 9: 460–466 (1934).
40. —— and Mathews, H. M., "A Qualitative Study of the Bacterial Flora of Sea and Land Breezes." *Proc. nat. Acad. Sci., Wash.*, 22: 567–572 (1936).

PHYSICAL ASPECTS OF HUMAN BIOCLIMATOLOGY*

By KONRAD J. K. BUETTNER

Randolph Field, Texas

According to the classical definition by Humboldt [50], the term *climate* designates all changes in the atmosphere that noticeably affect human physiology. With this definition, upon which the present article is based, the meteorological elements must be evaluated in a manner different from that used, for instance, in connection with an airway meteorological service or with plant geography. Thermal elements of the climate become the most decisive, specific radiations come next, whereas pressure, wind direction, and similar basic weather data play only an indirect role.

HEAT BALANCE

Heat Production. The body gains the energy necessary for maintaining its internal temperature by combustion of converted food. This energy is given off by conduction, convection, radiation, and evaporation from the skin, as well as by advection to, and evaporation from, the respiratory organs. A maximum of 20 per cent of the heat produced (less basal metabolism) may be used for external work. The excretion of stool, urine, etc., merely presents a decrease in the body's heat capacity. Unfortunately, metabolism, which is essential for the life of every cell, takes place in both hot and cold environments; however, it cannot be regulated beyond certain limits.

An adequate oxygen supply to the blood through diffusion in the lungs' surface requires a certain partial pressure of oxygen in the lungs. This partial pressure is determined almost exclusively by the atmospheric pressure because the fraction of oxygen in the air is very nearly constant. Pressure variations caused by weather, however, are too small to be significant in this relation. Consequently, only the well-known variation of pressure with altitude needs to be considered.

The combustion end products, CO_2 and H_2O, are removed by breathing and metabolism. Excessive CO_2 content of the inspired air is harmful; it occurs only near volcanoes and in artificial climates (crowded rooms, submarines, etc.), but in such cases the coexistent sultriness is often more important than the presence of CO_2.

The base value of heat production (at rest, "comfortable" climate) is about 40 Cal m^{-2} hr^{-1}. Exercise, fever, or cold increase this value up to a rate larger by more than one order of magnitude.

If, starting with "comfortable" conditions, the room temperature decreases by 10C, the heat production of the body increases by about one half. This heat production would have to double if the skin temperature were to remain constant. Actually, however, the latter decreases by about 7C because of a decrease in the advection of warm blood to the skin, particularly to that of the extremities whose temperature then drops almost to that of the air. The thermal conductance (taken between the interior of the body and the average skin) decreases from about 50 to 10 Cal m^{-2} hr^{-1} (deg C)$^{-1}$. On the other hand, it may rise to 120 Cal m^{-2} hr^{-1} (deg C)$^{-1}$ for portions of the skin heated separately and may drop almost to zero for cold extremities [7, 19, 84]. This complex of phenomena corresponds to variations in blood circulation which, in both extremes, constitute a severe load on the heart: since the temperature gradient between heart and skin is small in the case of warm surroundings, the heart must pump large quantities of blood to the periphery in order to rid the body of the heat it produces. In the case of cold surroundings, on the other hand, the necessary heat must be produced by arduous muscular activity which, again, strains the heart.

Rapid variations of the thermal environment produce temperature fluctuations first in the skin, then in larger portions of the body, until a new equilibrium is established. This involves a stock-piling of heat, using the body's heat capacity. The values of specific heat are:

Skin	0.77	Cal kg^{-1} (deg C)$^{-1}$
Fat	0.55	"
Muscle	0.91	"

Roughly, the entire body can lose j Cal m^{-2} without discomfort if its heat metabolism amounts to j Cal m^{-2} hr^{-1} [8]. The corresponding drop in the true body temperature is composed of 65 per cent core temperatures (rectum) and 35 per cent skin temperature [24]. This ratio may vary; nevertheless, the mean body temperature is usually not equal to the rectal temperature.

Heat Loss by Convection. In the simplest case, the temperature of a room may be used as a measure of its climate. However, thermal environments are rarely as simple as that; usually wind, radiation, and humidity act simultaneously. The total rate of heat loss can be determined only synthetically from these weather elements.

The known laws of heat transfer by convection have successfully been applied to the human body [13, 16, 19, 49]. The skin is surrounded by a laminar boundary layer several millimeters thick (in the case of calm air). At the inner portion of this layer, heat transfer takes place by pure conduction. A steady transition to turbulent austausch takes place towards the outside. The dependence of surface conductance h_a on wind velocity v, diameter d, and density of the air ρ is expressed by the following formula:

$$\frac{h_a d}{k} = CU^n, \qquad (1)$$

* Translated from the original German.

where k is the thermal conductivity of the air; $U = v\rho d/\mu$, Reynolds number; μ is the viscosity of the air, and d is the diameter of the cylinder or sphere, or the trunk width of the human body. The exponent n depends on the structure of the air flow. For the range which is of interest to us ($10^3 < U < 10^5$) the following values have been found:

	C	n	References
Cylinder, normal to axis	0.35	0.56	[72]
Sphere	0.70	0.52	[13, 16, 19]
Human body, supine, nude	1.0	0.54	[13, 16, 19]

(1a)

Under normal conditions, for the supine adult, we find approximately [13]:

$$h_a = 6.3\sqrt{v}\ \text{Cal m}^{-2}\ \text{hr}^{-1}\ (\text{deg C})^{-1}, \qquad (2)$$

where v is expressed in kilometers per hour and is greater than 2 km hr^{-1}. In these measurements [13], the exposed surface is taken as fifty per cent of the geometrical surface. More recent measurements made for sitting persons [84] and standing persons [63] showed values of $5.5\sqrt{v}$ and $3.9\sqrt{v}$, respectively. The scatter in the factors (6.3, 5.5, and 3.9) is caused by differences in the methods used for determining the area from which heat is lost.

Siple [27, 79] used the heat of fusion of water to determine the "wind chill" which, actually, is the cooling power at sub-zero temperatures. In his antarctic experiments he found the formula,

$$h = 10.9\sqrt{v} + 9.0 - v. \qquad (2a)$$

He used as a measuring device a cylindrical vessel which contained the water to be frozen. The plastic wall, one-eighth of an inch thick, has to be considered as a thermal insulator, or the whole instrument may be thought of as a clothed cooling device. Therefore, equation (15) (see below) should be applied with $S = 0$ and $h_r = 5$ Cal m^{-2} hr^{-1}. Best agreement with the measurements [27, 79] can be found by writing equation (15) in the form:

$$h = \frac{1}{(h_a + h_r)^{-1} + h_c^{-1}}, \qquad (2b)$$

where $h_r = 5$, $h_a = 9.0\sqrt{v}$, and $h_c = 80$, all in Cal m^{-2} hr^{-1} (deg C)$^{-1}$ (v in km hr^{-1}). The values of h_a and h_c thus extracted from the experiments fit the data to be expected for a cylinder of this size and a plastic wall of this thickness.

According to equation (1), the local value of h_a is large for highly curved surfaces such as the fingers. Moreover, h_a decreases with decreasing atmospheric pressure.

In the absence of forced flow, the higher surface temperature in itself produces convection. From the law of similarity and from experiments we obtain

$$\frac{h_{a0}\,d}{k} = c\sqrt[4]{G}, \qquad (3)$$

where $G = d^3 g\rho^2(\vartheta_s - \vartheta_a)\,\mu^{-2}\,T^{-1}$,
$T = 273 + (\vartheta_s + \vartheta_a)/2$,
g = acceleration of gravity,
ϑ_s = skin temperature,
ϑ_a = air temperature.

Values measured in calm air give
$c = 0.37$ for cylinders [72],
$c = 0.44$ for spheres [13].

Because of body motion and other sources of air flow, the human body is rarely surrounded by completely calm air, and hence equation (3) is rarely applicable. Actually, in rooms of normal size and in calm air, a value of $h_a = 3.3$ Cal m^{-2} hr^{-1} (deg C)$^{-1}$ was found for recumbent adults and $h_a = 3.9$ Cal m^{-2} hr^{-1} (deg C)$^{-1}$ was found for recumbent children [13, 16]. A value of $h_a = 2.0$ Cal m^{-2} hr^{-1} (deg C)$^{-1}$ was found for seated persons, on the basis of the geometric instead of the projected surface [84]. In narrow containers, such as calorimeters designed for human subjects, even smaller values have been observed [64].

Heat Transfer by Evaporation. Analogous laws are applicable to the evaporation of moisture from the skin to the air. The numerical values for boundary layer and austausch may be applied here. For the vapor transfer coefficient

$$h_w \equiv V/(e_s - e_a),$$

where V is the heat loss (in Cal m^{-2} hr^{-1}) by evaporation, and e_s and e_a are the vapor pressures of skin and air, respectively, theory and measurements [10, 11] yield the value:

$$h_w = 1.63\ h_a\ \text{Cal m}^{-2}\ \text{hr}^{-1}\ \text{mb}^{-1}. \qquad (4)$$

Experiments reveal that either the skin becomes wet, owing to active secretion of sweat glands, or water diffuses through a slightly permeable membrane at the surface (presumably the *stratum granulosum*). The second process, the so-called *perspiratio insensibilis*, is almost independent of wind, since the diffusion resistance is considerably greater in the skin than in the boundary layer of the air [19]. On the other hand, this portion of heat transfer by evaporation apparently increases with increasing E_s, the saturation vapor pressure at skin temperature. If f denotes the wet portion of the total surface, and c is a constant, then

$$V = [fh_w + c(1 - f)]\,(E_s - e_a). \qquad (5)$$

Equation (4) applies only at sea-level pressure. According to equations (1) and (1a), h_a is nearly proportional to $\sqrt{\rho}$ and thus decreases with altitude. However, since equation (4) contains the diffusion constant, which increases as $1/p$, we find $h_w \propto 1/\sqrt{\rho}$. We thus find an *increase* of the vapor transfer coefficient with height. In view of the fact that the *perspiratio insensibilis* is independent of h_w, the well-known desiccation at high altitudes must be attributed less to this factor than to the decrease in e_a.

Heat Transfer by Radiation. Radiation from the sun and sky, and solar radiation reflected from the surroundings (S) cover the spectral range between 0.3 μ and 3 μ; radiation of the skin and terrestrial objects (R), that between 3 μ and 50 μ. The optical constants for the skin differ widely for S and R: the mean energy albedos are 0.35 and 0.04, respectively [16, 42]. The mean absorption coefficients for incident solar and terrestrial radiation are 7 cm^{-1} and 100 cm^{-1}, respectively.

The reflection and absorption spectra in the solar region are complicated and depend upon pigmentation (race, suntan), congestion of blood vessels (erythema), etc. In the yellow and red bands, noticeable fractions of the radiation penetrate deep layers, whereas in the region from 0.5 μ to 0.9 μ, reflection is particularly strong. As far as its own radiation R is concerned, the skin acts nearly as a black body, its emission constant being $\epsilon = 0.96$ according to [16], or 0.99 according to [42]. We may write

$$R = 4.96\epsilon\left[\left(\frac{T_s}{100}\right)^4 - \left(\frac{T_r}{100}\right)^4\right] = h_r(\vartheta_s - \vartheta_r)\ \text{Cal m}^{-2}\ \text{hr}^{-1}, \quad (6)$$

where, approximately,

$$h_r = 19.84\epsilon\left(\frac{T_s + T_r}{2}\right)^3,$$

and ϑ_r is the radiation temperature of the surroundings, which in the case of a room is identical to the mean wall temperature (T_r and T_s are in degrees absolute). This identity applies also to specularly reflecting walls, as long as they are sufficiently far removed. In the open, the following approximation is usually adequate for a man or a sphere just above a plane surface:

$$R = \frac{h_r}{2}(\vartheta_s - \vartheta_r) + \frac{h_r}{2}(\vartheta_s - \vartheta_a) + R_h\ \text{Cal m}^{-2}\ \text{hr}^{-1}, \quad (6a)$$

where ϑ_r is the mean temperature of the solid surroundings, and R_h is the radiation to the sky by a sphere having an air temperature ϑ_a. Data on R_h are given after the calculations of Poschmann in [69]. The radiation R_h depends on cloud conditions, ϑ_a and e_a, and may rise to 100 Cal m^{-2} hr^{-1}. Its effect on human heat loss may be considerable, for example, during clear winter nights.

Total Heat Balance. The heat balance for the nude body is

$$Q - C\dot{\vartheta}_b + \epsilon_H S F_H = h_a F_a(\vartheta_s - \vartheta_a) + h_r F_r(\vartheta_s - \vartheta_r) + [h_w f + c(1 - f)]F_a(E_s - e_a). \quad (7)$$

In this expression the subscript H refers to the sun (Helios), Q is the heat production of the body (less loss in breathing), C is its heat capacity, and $\dot{\vartheta}_b$ is the change of the average body temperature with time.

At least three different surfaces have to be considered in the discussion of heat balance, namely, the projected surface for the solar radiation (F_H), the surface for the heat loss to the air (F_a), and the surface for the heat loss by outgoing radiation (F_r). All surfaces are smaller than the geometrical surface F which has been measured frequently because of its dependence on the height, weight, age, sex, and clinical factors of the subjects. For adults, F amounts to 1–2 m^2. The values of F_a and F_r are smaller for the recumbent person because of mutual contact between certain skin areas and contact with the insulating bed. For a standing person $F_a \approx 0.8F$, and for a person supine on an insulating bed it may be as low as $F_a \approx 0.5F$. For the average of all possible positions, $F_H \approx F_a/4$. Sky and reflected solar radiation might require special consideration. Usually, $F_a = F_r$.

Sultriness and Heat. In an ordinary room, $S = 0$ and $\vartheta_a = \vartheta_r$. In calm air at sea level, the value of h_a has been found to be close to that of h_r. With this simplification in mind (which should not be generalized), we write $h_{a0} = h_r = h$. In order to maintain the body temperature at a steady state ($\dot{\vartheta}_b = 0$), equation (7) becomes

$$\frac{Q}{2F_a h} = \left[\vartheta_s + \frac{h_{w0} f + c(1 - f)}{2h} E_s\right] - \left[\vartheta_a + \frac{h_{w0} f + c(1 - f)}{2h} e_a\right]. \quad (8)$$

According to experiment [17], $c \approx 0.2h_{w0}$. If equation (4) is used, the physiological quantities ϑ_s and f, skin temperature and wetness factor, are thus related to the environmental values ϑ_a and e_a.

We may now introduce certain standard data such as $Q/F_a = 120$ Cal m^{-2} hr^{-1}, $2h_{a0} = 8.4$ Cal m^{-2} hr^{-1} (deg C)$^{-1}$, and $\vartheta_s = 34.5$C. We find from equation (8) that $f = 1$ when

$$\vartheta_a + 0.815e_a = 63,$$

in which ϑ_a is expressed in degrees centigrade and e_a in millibars. When the wetness factor f equals unity, thorough wetness and consequently unbearable heat conditions are indicated. This has been proved by different observations. The majority of deaths during military maneuvers in the United States [78] or caused by heat waves in the United States[1] have occurred with values greater than those approximated by the formula: $\vartheta_a + 0.75e_a = 61$. In general, similar climatic strain in the range of sultriness (*i.e.*, high temperature and high humidity combined) is given by equations of the form

$$\vartheta_a + ae_a = b, \qquad (a \leqq 0.815). \quad (8a)$$

The more comfortable the surroundings, the smaller the values of ϑ_s and f, and the smaller the values of a and b according to equation (8). Such systems of physiologically equivalent combinations of (ϑ_s, e_a) correspond to the "effective temperatures" [49] of the American Society of Heating and Ventilating Engineers. From equation (8) we can readily interpret them as lines of constant (ϑ_s, f).

Evaporative cooling devices for homes and automobiles leave the equivalent temperature of the cooled and moistened air,

$$\vartheta + 1.5e = \vartheta' + 1.5E', \quad (8b)$$

approximately constant (psychrometric equation). Comparison of equations (8*a*) and (8*b*) shows that relief from heat may be expected only in the case of low relative humidities.

Overheating of the body ($\dot{\vartheta}_b > 0$) occurs for corresponding values of S, e_a, ϑ_a, etc., particularly when $\vartheta_a > \vartheta_s$ and $\vartheta_r > \vartheta_s$, so that only evaporation has a cooling effect. This becomes inadequate either if h_w is too small and e_a too large (jungle climate), or if the

1. T. Düll and B. Düll, personal communication.

body is not capable of supplying sufficient water to make $f = 1$. This is known to occur in hot deserts, that is, wind no longer produces cooling. The average person is rarely capable of filtering more than 1 kg hr^{-1} of water through the skin; in certain morbid conditions, for instance prior to heat prostration or after prolonged thirst, the rate of filtering is even less. The subsequent overheating is associated with desiccation of the skin and the formation of a salt crust.

Since f as well as ϑ_s represents a physiological variable, the total effect of ϑ_a, ϑ_r, e_a, and v can be determined only by calculations of the type referred to above, but hardly by means of instruments which seek to simulate actual conditions. The problem of defining climatic strain in one unit becomes even more difficult in the case of sunshine ($S > 0$). A solution of this problem seems more promising in the region between sultriness and cold.

Cooling. Perspiration ($f > 0$) is an emergency function of the body; ϑ_s values should stay between certain limits. We feel comfortable when $\vartheta_s \approx 33$C and $f \approx 0$, that is, when $a = 0.2$ and $b = 26.4$ (equation (8a)). This corresponds to an effective temperature of approximately 22C, which is the "summer comfort" standard of the American Society of Heating and Ventilating Engineers for lightly clothed persons. For $a = 0.2$, the effect of atmospheric humidity on skin temperature is already negligible; however, values of relative humidity above 70 per cent and below 20 per cent are undesirable for secondary reasons (effect on clothing and microorganisms, or desiccation of the skin, respectively). If the last term (evaporation) is omitted and $\vartheta_b = 0$, equation (7) can be used to derive a numerical criterion describing the total effect of several climatic elements in one index:

1. *Température résultante* [58] and *operative temperature* [38]. These quasi-temperatures comprise the radiation and air temperature if $S = 0$ and $v = 0$:

$$\text{op. temp.} = \frac{h_r \vartheta_r + h_{a0} \vartheta_a}{h_r + h_{a0}}. \qquad (9)$$

2. *Cooling power.* The problem here is to determine the (artificial) heat supply Q of a device required to keep its temperature ϑ_s constant (usually at 34C). The quantities ϵ_h, ϵ, F_H/F_a, and the constants of equations (1) to (3a) are supposed to have the values corresponding to those of the human body. Instruments of this type are the katathermometer of L. Hill and the frigorimeter of C. Dorno (see [13, 16, 69, 74, 80]).

3. *Standard operative temperature.* Calculation [38] permits the comparison of two rooms ($v \geqq 0$, $S = 0$), considering the skin temperature ϑ_s constant:

$$\text{std. op. temp.} = \frac{h_r \vartheta_r + h_a \vartheta_a - \vartheta_s (h_a - h_0)}{h_r + h_a}. \qquad (10)$$

4. *Cooling temperature* (ϑ_s'). This is the fictitious skin temperature of a sphere, heated with $Q = \text{const}$, whose constants correspond to those of a blond human. In this case,

$$\vartheta_s' = \frac{h_r \vartheta_r + h_a \vartheta_a + Q/F_a + \epsilon_H S/4}{h_r + h_a}. \qquad (11)$$

The value ($\vartheta_s' - 12$C) is the "temperature of equal effectivity," that is, the temperature of a "normal room" ($v = 0$, $S = 0$, $\vartheta_a = \vartheta_r$), which is physiologically equivalent. (For existing instruments by Pfleiderer and Buettner [13, 16, 69, 80] the constants are $Q/F = 96$ Cal m^{-2} hr^{-1}, $\epsilon_H = 0.60$, $\epsilon > 0.90$, and $d = 15$ cm (sphere).) For an additional, smaller model, the "Frierkoerper": $Q/F = 120$ Cal m^{-2} hr^{-1}, $d = 7$ cm [19].

Numerous other forms may be reduced to those given above. We give cooling temperature preference over cooling power and standard operative temperature, since in the mean cooling range the human body tends to achieve a constant Q with variable ϑ_s, rather than to keep ϑ_s constant.

The purposes of establishing and observing these combination elements are as follows:

1. Establishment of a comparative physiological climatology for nonhumid regions. The thermal load on a human body may be calculated from the elements v, ϑ_r, ϑ_a, and S, if these are known at the location of the body; unfortunately, this is rarely the case. On the other hand, series of cooling measurements are available only sporadically.

2. Evaluation of artificial climate and clothing (see below).

3. Determination of the dosage for exposure to the open air in sanatoriums in moderate and cool climates. The tonic effect of the change from indoors to a veranda increases with the deviation from the comfort climate and with the duration of outdoor exposure. The ϑ_s value recorded by the frigorigraph serves as a measure of comfort or load; values are to be measured at the patient's bed or calculated by means of equation (11); $\vartheta_s' = 37$C is comfortable. A formula found successful in a sanatorium on the North Sea coast (calculated from data in [68]) is

$$(|\vartheta_s' - 37|)t = BD, \qquad (12)$$

where t is the exposure time in minutes, $B = 8$ for $\vartheta_s' < 35$C, $B = 20$ for $\vartheta_s' > 40$C, and $t = 4D$ for $35\text{C} < \vartheta_s' < 40$C. The dose D increases from 5 for hypersensitive persons to 60 for healthy persons. Because of possible damage from ultraviolet radiation, the exposure time must be limited on the basis of special measurements or tables (see Table II).

Clothing. The main purpose of clothing is to produce a bearable skin temperature through thermal insulation. Let us assume that the incident radiation S is absorbed at the surface of clothing having a temperature ϑ_x; outside, $\vartheta_a = \vartheta_r$; underneath the clothing, which has the thickness d, the skin temperature ϑ_s prevails. The quantity h_c is the "thermal conductance" of the fabric and $1/h_c$ is its "insulation value" (as a measure of this, one also uses 1 Clo = 1.715 deg C m^2 hr Cal^{-1}). The values of h_a and h_r are practically the same for clothed and for bare sections except perhaps in the case of thickly clothed fingers. The increase of the surface area has to be taken into account. In the

case of clothing such as fur, the true surface is difficult to determine.

We then have, for the surface concerned,

$$Q + \epsilon_H S F_H = F_a(h_a + h_r)(\vartheta_x - \vartheta_a), \qquad (13)$$

and

$$Q = F_a h_c(\vartheta_s - \vartheta_x); \qquad (14)$$

hence

$$\frac{Q}{F_a h_c} + \frac{Q + \epsilon_H S F_H}{F_a(h_a + h_r)} = \vartheta_s - \vartheta_a . \qquad (15)$$

The following conclusions can be reached from equation (15):

1. Above certain wind speeds (large h_a), solar radiation S becomes ineffective.

2. For $h_c \ll (h_a + h_r)$, that is, for good insulation, the effects of wind (assuming h_c to be invariant with respect to v) and of solar radiation vanish.

3. Protection against solar radiation by reflection (small ϵ_H) is particularly effective in the case of thin fabrics (large h_c), deep penetration of the radiation being undesirable, of course. Solar rays are best reflected by white surfaces; those originating from sources below 1500C (fire), by shiny metals [20].

4. The insulation value $1/h_c$ necessary to maintain the Q and ϑ_s of the body constant can be determined experimentally or by calculations on the basis of the frigorigraph values ϑ'. Let Q, ϵ_H, and ϵ be the same for the body and for the frigorigraph. For the body, $F_a = 4F_H$. Finally, let $\vartheta_a = \vartheta_r$. It then follows from equations (11) and (15) that

$$\frac{1}{h_c} = \frac{F_a(\vartheta_s - \vartheta_s')}{Q} \text{ deg C m}^2 \text{ hr Cal}^{-1}. \qquad (16)$$

For clothing comprising several layers ($i = 1, 2 \ldots n$),

$$h_c = \frac{1}{\sum d_i/k_i + \sum \dfrac{1}{k_a/d_a + h_r}} + f(v), \qquad (17)$$

where d_i and k_i denote thickness and thermal conductivity of the clothing layers, respectively; d_a and k_a are the corresponding values for the intermediate layers of air. (Since convection would occur in thick air layers, equation (17) is valid only for $d_a < 0.01$ m.) The factor d_a is particularly sensitive to pressure and body motion. The function $f(v)$ represents symbolically the total effect of wind pressure on insulation (wind penetration through porous fabrics and openings; flapping, etc.).

The conductivity of cloth (k_i) is determined by three elements: (1) conduction by the interstitial air, (2) conduction along the component of the fabric fiber oriented in the direction of heat flow (the same component is responsible for resistance against compression) which becomes undesirably high for dense, compressed, and wet fabrics, and (3) radiation from fiber to fiber, a process which has hitherto not been considered, but which becomes noticeable in the case of large fiber separations. With increasing temperatures, the two conduction terms rise only slightly, but the radiation term increases considerably [21]. Table I lists several values to illustrate the foregoing considerations.

Table I. Heat Conductivity and Density of Different Materials

(*According to Buettner* [14, 18])

Material	k (Cal m^{-1} hr^{-1} (deg C)$^{-1}$)	ρ (kg m^{-3})
Air	0.023	1.3
Water	0.52	1000
Wool cloth, dry	0.030	100
Wool cloth, wet	0.092	250
Wool cloth, dry, under pressure	0.035	170
Asbestos	0.079	860
Glass fabric, 20C	0.027	210
Glass fabric, 350C	0.081	210
Human skin, upper layer, excised	0.18	—
Human skin (0–2 mm deep), *in situ**	0.32	—
Human skin (2 mm deep), *in situ*, cool*	0.47	—
Human skin (2 mm deep), *in situ*, warm*	3.60	—

* After [14]. They include the effect of blood advection.

Rain or perspiration increases k_i; water in the cloth may even freeze, thus causing even higher k_i. These processes might make the cloth windproof; consideration must be given to the question when and where this water or ice can be eliminated.

The question of how the extremities should be protected in extreme cold is difficult to answer. If the body as a whole produces enough heat, even poorly protected extremities are safe [40]. On the other hand, for geometrical reasons, clothing for the hands and feet would have to be prohibitively thick.

In the case of hot desert winds, loose clothing affords protection against overheating and dehydration resulting from the inadequate water supply to the skin [1, 19].

Breathing. In breathing, a heat exchange A takes place which depends on the transferred volume M (m^3 hr^{-1}), the heat capacity and density of the air, and the difference in the equivalent temperatures of the inhaled and exhaled air.

The mechanism by which the exhaled air is heated and humidified may be conceived of as follows [70]: Incoming air having the equivalent temperature J_i cools the mucous membranes of the upper respiratory organs (nose, mouth, throat) by convection. During this process the air itself is raised to blood temperature and saturated with moisture. The air, returning after a brief pause in the lungs, is cooled by the still relatively cool, mucous membranes. As long as the mucous membranes are moist, the air remains saturated and emerges from the mouth or nose with the equivalent temperature,

$$J_e = \vartheta_e + 1.5E_e.$$

The process may be treated schematically by assuming a constant mean membrane temperature. We introduce as fictitious quantities the equivalent temperature of saturation at blood temperature J_b and at the average temperature of the mucous membrane J_s, as well as the thermal conductance of the membrane h_s, the sur-

face conductance of the passing air h_a, and the effective surface area of the membranes F. Then the heat transferred to the inhaled air during its passage into the lungs is given by

$$h_a F(J_s - J_i) = M\rho c_p(J_b - J_i), \qquad (18)$$

and the total heat loss of the body by respiration,

$$A = M\rho c_p(J_e - J_i) = h_s F q(J_b - J_s), \qquad (19)$$

in which q is a relative time factor depending on the

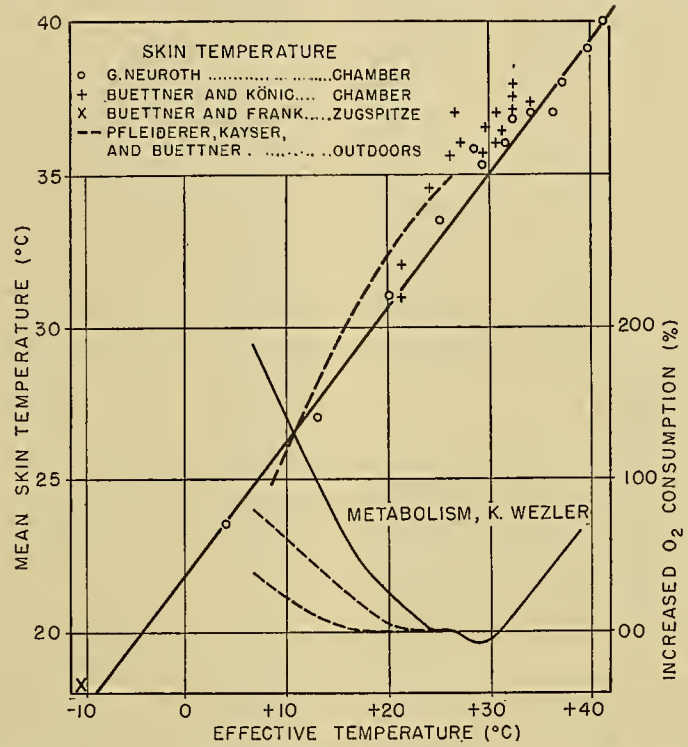


FIG. 1.—Mean skin temperature of supine, nude persons at rest, in terms of "effective temperature." Data are partly from a climatic chamber and partly from various climates in the open air (North Sea coast, Alps, central Africa). The abscissa has been converted to "effective temperature" on the basis of original values (temperature and humidity) according to the method of the A.S.H.V.E. for unclothed persons. In the case of cool and windy conditions, which prevailed for the most part at "effective temperatures" below 20C, the "temperature of equal effectivity" [18, 19] was used instead of the ordinary air temperature. Only in the central range were climatic conditions such that they could be withstood for several hours. The duration of experiments in the two extremes was about ½ hr. The metabolism curves (*after Wezler* [83]) in the lower right show the heat production measured by oxygen consumption of supine, nude persons as a function of "effective temperature." Data were obtained from four persons in a climatic chamber and show noticeable differences at the lower temperatures.

rhythm and relative length of the breath pause; hence

$$\frac{J_e - J_i}{J_b - J_i} = P(h_s, q, M, h_a, F). \qquad (20)$$

This calculation may be considered as a first attempt to find a connection between the observed temperatures without knowing J_s.

A comprehensive study of all available measurements for the case of normal mouth or nose breathing (not yet published; see [11, 17, 19, 70]) reveals a drop of J_e from 125C (mouth) and 125C (nose) down to 82C (mouth) and 65C (nose), respectively, if J_i changes from 100C to −40C. The average $M = 0.6$ m^3 hr^{-1} in these experiments. The variation of the function P with J_i is remarkably small: it drops from 0.84 to 0.75 (mouth) and from 0.84 to 0.62 (nose) if J_i changes from 100C to −40C. The slight dependency of P on M may be explained by a compensatory effect of M and F (cooling of deeper tracts with a deep breath). When the inhaled air is very warm and dry, the exhaled air is no longer saturated [11]. In any case, J_e is by no means constant but decreases with J_i.

At higher altitudes, A becomes an important part of the total heat loss: If M increases to compensate for the decreased partial pressure of O_2 ($M\rho = M'\rho'$), then the "dry" loss $M'\rho'(\vartheta_e - \vartheta_i)$ remains the same (observations in a pressure chamber have yielded the same values of temperature and saturation as at sea level). However, the "wet" portion $M'C(E_e - c_i)$ increases with M'/M. (C includes the heat of vaporization and the conversion from e to absolute humidity.) Exercise, fever, etc., have little influence on the relative amount of heat loss by breathing because increased internal combustion requires a higher M.

In general, the heat transfer due to breathing may be considered as small compared to the transfer through the skin.

The General Effects of Climate. Let us consider on the basis of our formulas how the body stabilizes its temperature, that is, how it balances its gain and loss of heat. In the "comfort" range, Q is at a minimum, ϑ_s is about 33–34C, and $f = 0$. If ϑ_a or c_a rises, Q at

first remains constant; increasing ϑ_s and f makes the removal of heat possible. In addition to the "dry" climatic factors, e_a becomes ever more important. With even higher values of ϑ_a and e_a, f finally reaches unity, and $(\vartheta_b - \vartheta_s)$ becomes too small to produce a sufficient "radiator effect." To remove a constant Q through the skin, a peripheral blood flow proportional to $1/(\vartheta_b - \vartheta_s)$ is necessary. This value becomes unbearably high. In addition, the increased blood temperature and work done in internal circulation raise Q [83] (see Fig. 1).

Cold restricts the peripheral blood circulation and thereby decreases ϑ_s and thus also E_s. By physiological regulation the decrease in skin temperature finally causes an increase in Q, varying considerably from individual to individual, as a consequence of shivering, conscious muscular activity, or thermally excited, purely physiological control (Fig. 1).

In spite of these effects, the core temperature remains constant in tropical as well as in arctic climates (except for extremes) after a brief adaptation period. The most important climatic elements in the tropics are humidity and temperature; in the Arctic, temperature and wind. In each case they are modified favorably or unfavorably by solar radiation. Exercise increases Q and thus makes warm climates less bearable and cold climates more bearable.

In a cool and moderate climate, variation in peripheral blood circulation and the stimulation to increased heat and food metabolism represent beneficial stimuli for healthy people, but a considerable strain for those with circulatory diseases. Metabolic and cardiovascular diseases as well as mental exhaustion or instability, chronic respiratory diseases and rheumatic afflictions occur frequently in the area of maximal stimulation, for instance, west of the Great Lakes [51, 56]. In these same areas we also find the earliest onset of growth and sexual maturity in adolescents.

In the warm, humid Tropics, a small amount of human activity accompanies a low Q. This condition is further augmented by the absence of diurnal, interdiurnal, and annual fluctuations of weather elements. The uniformly high values of temperature and humidity of the skin decrease its resistance to diseases. This decreased resistance, together with the abundance of microorganisms and a low social standard, increases the number of skin diseases; there are in addition innumerable internal tropical diseases. These illnesses might have influenced previous tropical cultures more than the climate proper could have done.

The dry, hot climates and the tropical highlands assume a position between the two extremes noted above, since they show a high variability of the weather, although at times the heat stress is great. The number of microorganisms is small. The dryness of the air is favorable to the cure of tuberculosis of the lungs. Cultures in these areas (Egypt, Near East, Inca) are outstanding and old.

Survival and culture in the Arctic seem to be a matter of providing only food, shelter, clothing, etc.

Medical geography and climatology [51, 56, 57] may have to be rewritten in the next few decades because a large number of the indirect effects of climate, especially the infectious diseases (malaria, sleeping sickness, food poisoning) can be eliminated. New means of transportation and storage will eliminate unbalanced diets; air conditioning can alleviate peaks of heat stress. The sociological situation can thus be changed. The Pygmy in the primeval forests of the Congo, the Bedouin in the Sahara, and the Eskimo in the Arctic are influenced not only by climate, but also by such factors as seclusion and the poverty of the country.

RADIATION

The caloric effect of solar radiation was discussed above in connection with the over-all heat metabolism of the human body. The heat radiation from the clear sky is slight. The differences in spectral composition between direct solar radiation and sky radiation, each considered for various solar elevations, cloudiness, and turbidity, are of little importance for the caloric effect.

On the other hand, a number of photochemical processes are connected with shorter wave lengths and are thus highly variable with regard to weather. The most important of these processes are vision by means of the spectrally selective, sensitive retina and the formation of erythema, pigment, and vitamin D in the skin (for a summary see [12, 16, 69]).

For measurements in the biologically important range below 0.32 μ one uses, in addition to spectrographs, special instruments such as the cadmium cell of which C. Dorno made excellent use, the special photocell of Coblentz [26], and the UV-Dosimeter of IG-Farben [37] (for a summary see [60, 69]). Measurements of the ultraviolet sky radiation are indispensable in all biological investigations since more than half the total incoming ultraviolet radiation is scattered by the sky.

Erythema solare, erroneously referred to as sunburn, is produced by the direct action of ultraviolet radiation below 0.32 μ upon the cells of the skin, perhaps directly on the nucleic acid of the cell nuclei (Fig. 2). In this process the photochemical sensitivity curve is modified by absorption in the uppermost layers of the skin. The primary cell damage is followed after about one-half hour by reddening of the skin, subsequent swelling, and in the case of excessive doses, by the formation of blisters and epidermic defects after several days. In the case of skin with a certain predisposition, the same cell damage causes the formation of "delayed pigment." This occurs after two to three days and consists of a migration of melanin particles to the surface. By a different photochemical process, the "immediate pigment" is formed even more rapidly than the erythema. This process is caused by long-wave ultraviolet radiation (Fig. 2). It consists of a darkening of already present melanins [12, 43, 44, 46].

In the case of erythema and of both kinds of pigment, one can measure the threshold and also the degree of discoloration; the latter is determined by the difference in reflectivity at a wave length typical for the particular discoloration (green in the case of erythema) [16]. The threshold dose required for erythema is essentially a value depending on the individual. The threshold

dose D of white persons is $50 < D < 600$ (D = ultraviolet intensity in units of the IG-Farben Dosimeter $\times$ time in minutes) [16, 54]. The mean value of all persons examined (mostly blond) was $D = 200$, a value which corresponds to a 15-min exposure to sun and sky at 60° solar altitude in middle latitudes. In all persons the reddening measured by means of a reflectometer increased by about 1 per cent for each 80 D units above the threshold value of D. Different climates (arctic to tropic), the temperature of skin or air, the origin of the radiation (sun or sky), wind, etc., do not influence the relation of D and the erythema produced [16, 64].

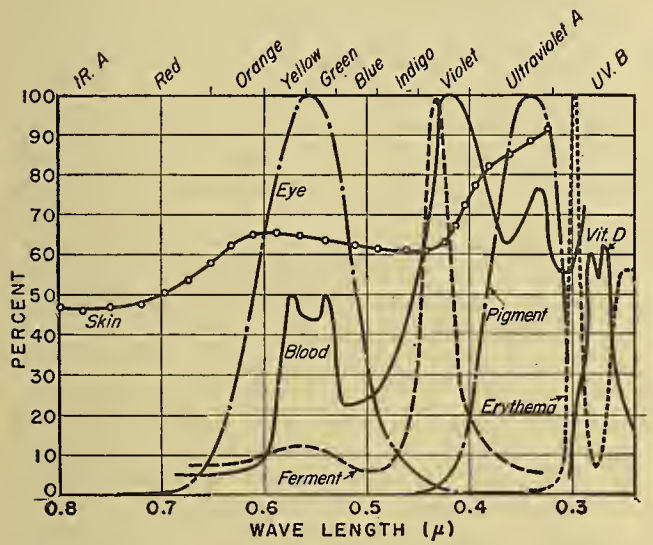


Fig. 2.—Threshold and absorption spectra for the human body. *Blood*—decrease in light intensity when passing through a layer of blood corpuscles 7.8 μ thick (about equal to the diameter of one blood corpuscle). *Skin*—loss in passing through a light-brown layer of skin 0.8 mm thick. *Ferment*—breathing ferment, according to Warburg, which is held responsible for the effect of violet light. *Vitamin D*—absorption coefficient of nonirradiated ergosterin. *Eye*—sensitivity curve of the normal eye, adapted to daylight. The outer limits lie at 800 and 320 mμ. *Pigment*—the curve of direct pigmentation due to long-wave ultraviolet, the so-called "immediate pigment," according to Henschke and Schultze [46]. *Erythema*—skin-reddening curve plotted according to the reciprocal values of the excitation threshold. In the curves "blood," "skin," and "ferment," light depletion is represented; 100 denotes opacity. In the case of "vitamin D," the numbers, which have been divided by 100, represent the absorption coefficients. All other curves are referred to the maximum of the curve in question. (*After Pfleiderer and Buettner* [69].)

Immunity against erythema can be attained by the "delayed pigment," by thickening of the epidermis, or by both. Ointments which do not transmit at $\lambda < 0.32\ \mu$ afford protection against erythema without preventing the formation of "immediate pigment." A light erythema may be a desirable stimulus for a healthy person; on the other hand, an acute form is associated with a dangerous activation of tubercular lung processes and may, in rare cases, even cause skin cancer. The cosmetic value of the pigment is a matter of fashion.

Vitamin D, which is indispensable for the formation of bones, is produced from the ergosterin of the skin by almost the same radiation (Fig. 2) and at about the same superficial skin layer as erythema. Since this vitamin nowadays can also be taken orally, rickets can no longer be considered a climatic disease.

The ultraviolet radiation below 0.32 μ varies more than that of any other part of the solar spectrum. It is absorbed by the atmospheric ozone, highly scattered by air and aerosol, and reflected by snow. The effect of dust above large cities is frequently overestimated; the loss for Boston [41] and Berlin [16] amounts to less than 20 per cent for the noontime sun, and even less for the sky. In Berlin the loss is no greater for the ultraviolet than it is for the other wave lengths.

Table II. Ultraviolet Intensity of Sun and Sky*

Region	Sun's elevation							
	10°	20°	30°	40°	50°	60°	70°	90°
Northern and central Europe, and central Africa, altitude < 500 m	1	3	6	8	11	14	—	—
Oceans in tropical latitudes	2	4	7	12	17	22	27	37
Central Europe, altitude 1150 m	—	4	8	12	16	20	—	—
Central Europe and central Africa, altitude approx. 3000 m	—	5	11	16	21	26	31	42

* Measured with the IG-Farben UV-Dosimeter, 1938 model, during the warm season and with a clear sky (summarized in [69]).

Table II shows measurements made with the IG-Farben Dosimeter (1938 model) from Lapland to the hills of the Belgian Congo. Table III shows the effect of clouds of medium thickness on the ultraviolet radiation of sun plus sky.

Table III. Effect of Cloudiness on Ultraviolet Radiation*

Cloudiness	0/10	3/10	5/10	9/10	8/10	10/10
Percentage of ultraviolet radiation transmitted	100	90	75	65	55	45

* For clouds of medium thickness, with average elevation of sun.

The brightness and color of the surroundings produce a number of psychological and physiological effects via the eye. The greatest change in this field, the introduction of electric illumination at the beginning of this century and the resulting elimination of biological darkness during the winter, is held responsible for the more rapid maturing of adolescents observed since that time.

Glare is typical of two climates: the arctic and the subtropic, where bright snow or sand, abundant sunshine (subtropic), and the scarcity of natural shadow coincide. Temperate and tropical zones have less sunshine as well as a darker background and numerous sources of shadow. Glare increases the contrast threshold of the eye, impairs space perception, and makes landscapes appear flat. Snow blindness may be due to glare as well as to an ultraviolet erythema of the eye.

The beneficial or, in the case of an excess, harmful effect of the natural radiation encountered in the open

is usually combined with thermal climatic stimuli, such as cold, wind, water, and snow. Good and bad physiological effects thus depend frequently on the totality of all climatic stimuli.

METEOROTROPISM

The preceding sections present information on the current status of *physical bioclimatology:* measurable weather factors produce measurable changes in man by biologically understandable mechanisms [16]. Such clear, causal relationships are rarely encountered in bioclimatology except in the fields of thermal and photochemical effects. Not even the effects of changes in the chemical composition of atmospheric air (CO_2, O_3, N_2O, I) fall into this category.

The deleterious effects of harmful industrial effluents are felt by the population only if the wind has the proper direction, or during prolonged stable air stratification, especially in winter. All our knowledge of micrometeorology and austausch has to be applied to check this disgraceful aspect of modern artificial climate. Hay fever is connected with the blossoming time of certain species of grass, as well as with dryness and wind direction. All of these phenomena can, to a certain degree, be reproduced in the climatic chamber; however, in the case of head colds, headaches, and associated diseases of doubtful origin, this is no longer entirely true. In considering daily and seasonal rhythms (except for those caused by radiation and dietary conditions) and the numerous "trigger" diseases initiated in the sympathetic nervous system, we have entirely left the field of physical bioclimatology for the field of "statistical bioclimatology."

The Effect of Periodic Weather Processes. Nearly every biological component from the rectal temperature to the mental disposition has a daily and a seasonal rhythm. Correlation of these biological elements with any meteorological factor proves nothing a priori, except that weather and human life take place on the rotating earth. Changes in personal climate caused by night work or work in a mine apparently fail to cause an immediate change in the phase of biological curves; on the other hand, in case of sudden change of longitude (such as by air travel) the biological phase seems to adapt itself within a few days (H. Strughold).

The nature of the seasonal rhythm can be illustrated by two diseases whose causes have been recognized. The high infant mortality rate during summertime, which was formerly observed in all temperate climates, was due to overheating as a result of an excess of blankets and clothing [31] and to a lack of cleanliness causing an increase of the number of all kinds of microorganisms. Many of these microorganisms find optimal living conditions during summer weather. Lack of ultraviolet-produced vitamin D, accompanied by dietary deficiencies, leads to a wintertime peak of rickets. If exposure to the sun from February to April (central Europe) causes excessively rapid healing, that is, calcification of the bones, the calcium level in the blood decreases, possibly resulting in tetany of infants [31, 61, 69].

Many infectious diseases show yearly variations that are due either to changes in the daily contact among people or to alterations in living conditions of microorganisms. Circulatory diseases are more frequent during the season of greatest stress owing to sudden weather changes. Certain skin diseases have their peak during the month of maximum perspiration. The sex ratio of births and the pH of blood vary with the season [67]. The number of deaths from heart failure, suicide, blood pH, and the ratio of births to deaths reveal a weekly as well as a lunar period, according to Petersen [67].

The Effect of Aperiodic Weather Processes. Many nonepidemic diseases begin abruptly, that is, with a well-defined onset. Some of them have a tendency to begin suddenly in number, that is, the distribution of initial manifestations over all days of the year is by no means statistically random. A trigger cause must therefore be present, and this has long been sought in the weather. The numerous attempts to hold individual weather elements (p, e, v, ϑ_a, or their time derivatives, or oxygen content of the air, precipitation) responsible must be considered as having failed. Even in the case of head colds which may perhaps be caused partly by thermal effects, it has not been possible to find a correlation with the daily temperature fluctuations, except perhaps for a decrease in frequency for very slight temperature fluctuations and for strong winds [36]. Extremely low temperatures have usually failed to produce colds in heroic self-experimentation or in the case of entire armies [59]; however, a sudden onset of thawing weather or rain may be coincident with an onset of colds. On the other hand, it has been made possible by the methods of Linke [55] and de Rudder [31] to find a correlation between the incidence of diseases and weather fronts, prefrontal subsidence, and orographic foehn.

The meteorotropic diseases have in common the prerequisite of a constitutional or acquired predisposition of the individual. The predisposition may be a latent infection or a permanent defect such as a scar. They may also be partly initiated by the sympathetic nervous system. The weather frequently served only as a trigger which numerous other stimuli could provide equally well. Some of these diseases are "pains due to the weather," laryngeal croup, thrombosis and embolism, acute glaucoma, hemoptysis, the onset of diphtheria, herpes, stroke, and attacks of asthma [9, 31]. Variations in the healthy body caused by weather (blood pressure, flushing of the skin, and a drop of the leucocyte count upon administration of $NaCl$) have also been observed. Experiments [32] have revealed correlations of psychological conditions with weather fluctuations as well as with solar activity.

However, it must be noted that not every front acts as a trigger and that, in addition to fronts, other factors may act in the same way. Sometimes cold and warm fronts and occlusions act alike; sometimes their consequences are diametrically opposed. For many diseases

correlation with a weather phenomenon is difficult to detect because the time lag between the given phenomenon and the first manifestation of the disease is unknown and perhaps variable; furthermore, almost all investigations have been made in central Europe and in the northern United States, that is, in regions where changes of air masses are too frequent to correlate one onset of illness with one front only. A number of the pertinent investigations have failed to meet statistical requirements.

According to Petersen [65], different fronts have different effects: cold fronts cause "anabolism," contraction of blood vessels, spasms, and biochemical reduction; warm fronts produce the opposite effects. Similarly, Curry tried to classify his patients into those sensitive to warm fronts and those sensitive to cold fronts [10, 29, 30]. A premonition of the weather, together with the foehn sickness, was explained [36] by the theory that descending air currents produce triggering action, particularly in the case of rheumatic attacks. The weather effect demonstrated by fronts may be nonspecific [69]; it may be a remote effect or, most probably, it may be caused by admixtures to the air. Pressure variations of all frequencies were found to be without effect [81]. On the other hand, the foehn illness could be avoided by purification of the air [81]. The decisive agent was assumed to be the oxidizing power of the air, the "*Aran*" [29, 30] (essentially ozone), which is brought down from high altitudes by descending air currents and perhaps produced by electric discharges. According to this, one could expect no causal relationship with fronts, but only a statistical relationship. The "*Aran*" hypothesis remains to be verified.

The vast number of publications in the past on effects of the static electric field of the atmosphere and of the terrestrial magnetism on man have not survived more recent critical tests. The daily occurrence of contact and friction electricity during extreme dryness, such as in heated rooms in winter, causes uncomfortable electric discharges and may occasionally cause an explosion of gasoline or ether vapor.

Theoretical calculations and some clinical results instigated the hypothesis of certain effects of inhaled atmospheric ions on man. Beside enhancement of specific electrochemical effects of the inhaled nucleus, a charge of the alveolar wall by unipolarly charged and absorbed ions may be anticipated. Open-air tests did not yield a reliable correlation coefficient between the number and unipolarity of large atmospheric ions or any other electrical factor of the inhaled air and the human health.

The very interesting correlation between sunspot activity and mental diseases, suicides [32–34], gestation eclampsia [6], the sedimentation rate of the healthy male blood serum [82], cholera in Russia, and meningitis (for a summary see [67]) belong in the field of bioastrophysics unless a meteorological agent can be found as an intermediate factor.

Aviation Medicine [3, 39, 48, 76]. At high altitudes, or on mountains, man is subject to specific effects of oxygen partial pressure and of changes of air pressure. Additional effects stemming from heat or cold, radiation, stress of flying or mountain climbing, acceleration, etc., might be aggravated by the effects of pressure. The blood, as the carrier of fuel and oxygen, must attain at least an 80 per cent saturation with oxygen in passing the lungs. To accomplish this, an oxygen pressure of 80 mb must prevail within the lungs, a fact which in turn presupposes a 133-mb oxygen pressure of the inhaled air. Ordinarily, 63 mb of H_2O, 53 mb of CO_2, as well as the N_2 pressure, have to be added to account for a real picture of the air inside the lungs.

At an average height of 3.5 km, the external O_2 pressure assumes the value of 133 mb. Above this level, a nearly complete physiological compensation is provided through various countermeasures: increase in breathing, increase in the amount of the oxygen-carrying vehicle (blood), and decrease in the alveolar CO_2 pressure.

Below 96 mb O_2 pressure (above 6 km of height), these compensations fail in a nonadapted man and the "engine" stops owing to lack of oxygen.

If pure oxygen is breathed, the afore-mentioned critical partial pressures of oxygen are reached at altitudes as high as 11.8 km and 14 km, respectively. For this reason, pressurized cabins are indispensable above these heights, even for a crew breathing O_2 by use of masks.

The reduction in barometric pressure in itself may cause discomfort and distress due to the expansion of gases trapped in body cavities (intestines, dental cavities, sinus).

Nitrogen is physically dissolved in the blood according to the existing partial pressure (Henry's law). Subsequent to a fast ascent to heights above 8 km, this N_2 tends to degasify from the blood, forming small bubbles which provoke pain in the joints and may even affect the brain. Owing to the hydrostatic pressure, the lower organs are usually less affected. These symptoms are known under the name "aeroembolism." The N_2 in the blood may be replaced by O_2 or *He* to prevent aeroembolism [48, 79].

A sudden drop of pressure may be caused by rupture of a pressurized cabin at great heights. This drop causes rapid expansion of the air in the lungs. This "explosive decompression" may even damage the walls of the lungs.

Space Medicine [4, 5, 23, 47, 52, 53, 77]. Space medicine comprises the human factor involved in flights beyond the lower stratosphere and in flights along trajectories equivalent to those of a celestial body. The main problems of this new field are:

1. With increasing altitude the protective atmospheric filter will gradually be lost. First solar X-rays are encountered at 100 km, but the existence of hard components, able to penetrate the hull of a craft, is dubious. Effects of cosmic-ray showers and of cosmic radio waves on man are being considered, but they are not very likely. Biological effects of cosmic-ray primaries prevalent above 25 km may be of prime importance. In passing through tissue, they produce very

thick and heavy ionization tracks which, compared with effects of alpha rays, should be strong enough to endanger lives exposed for more than one hour [53, 77].

2. The cabin climate of a craft at an altitude of more than 150 km, coasting along a celestial orbit after power shutoff, would be characterized by radiation equilibrium of the hull and lack of air convection inside. The surface temperature depends on the following radiations in addition to interior heating: sun, solar radiation reflected from the earth, and the radiation from the earth and cosmos. The main material constant is the ratio of the absorption coefficient at $\lambda < 3\,\mu$ to that of $\lambda > 3\,\mu$. This ratio equals 1.0, 0.1, and 6.0 for ideal black, white, and metallic surfaces, respectively. With a solar constant of 1200 Cal m^{-2} hr^{-1}, an average earth albedo of 0.42, and an average earth radiation of 174 Cal m^{-2} hr^{-1}, equilibrium temperatures are found [23] for a sphere, a plate always facing the sun, and a plate always facing the opposite direction (Table IV). The temperatures are higher than one might expect.

Table IV. Equilibrium Temperature of a Solid Surface Outside the Atmosphere but Close to the Earth

Form of surface	Day			Night
	black °C	white °C	nickel °C	all °C
Sphere	+63	−42	+239	−68
Plate facing the sun	+122	−51	+347	−29
Plate facing the opposite direction	+68	−13	+231	−270

If the earth's gravity is balanced by the centrifugal forces of the flight trajectory, gravity within the craft is absent and convection consequently ceases (see equation (3)). Exchange of heat, water vapor, respiratory gases, settling of dust, etc., can take place only by processes such as artificial ventilation and filtering. The thermal insulation of thick air layers would be equivalent to that of stagnant air of this thickness.

3. Lack of gravity may influence the human sensory system in a way and to an extent which cannot be predicted.

Effects of Extreme Heat and Cold [22, 45, 62, 73]. All the formulas on heat exchange of the human body, described in the first section, are based upon the assumption of a steady state. The skin and rectal temperature changes are too small to cause additional factors in our equations, except the simple one concerning any slow, temporal, linear increase of the body temperature as a whole.

If, however, heat transfer to the skin surpasses 500 to 1000 Cal m^{-2} hr^{-1}, the simplification of a steady state no longer holds, and the differential equation of heat conductivity has to be applied in full.

When heat or cold is applied during a short period of time, generally less than one minute, the skin temperature changes according to the amount and kind of heat transfer (Newtonian, constant nonpenetrating heat, constant penetrating radiant heat, contact heat, etc.) and according to the material and initial temperature of the skin. The temperature change of the surface proper generally depends on only one constant, the product of three characteristics (heat conductivity × density × specific heat) of the skin at the surface. This product was shown experimentally to lie between the values of water and fat, and not to depend on the blood flow. With this experimental fact sudden temperature changes of the skin can be calculated (formulas in [22, 25, 45, 62, 73]), for example, for the rapid cooling in very cold wind, according to measurements of Siple [79] and Buettner [22], or heating by solar radiation [73].

PROBLEMS FOR FUTURE RESEARCH

Bioclimatology is essentially an applied science. As such, it can be expected to experience changes in its major problems as a consequence of progress made in related fields. Taking tropical medicine as an illustrative example, we may note that it may one day find itself in the position of being confined to problems of sultriness and sunstroke as soon as progress in epidemiology will have eliminated the major objectives currently being pursued. From this point of view, heat and cold merit serious considerations as basic and isolated problems. Few thermal measurements have been made on clothed people exposed to the actual environment of various climates, whereas ample data exist for nude subjects under the artificial conditions of a climatic chamber. These conditions fail to account for the variation of some important factors, chief among them being rapid changes in the wind speed. These variations may cause variations in the skin temperature, but to a varying degree, depending on the situation, because cooling power depends on $\sqrt{v}$, air transport on v, and wind pressure on v^2. Temperature and radiation are also subject to rapid meteorological changes.

It is believed that there exists a pronounced disproportion between competitive formulas and instruments designed to determine the "dry" cooling and their application in the practice of physicotherapy, comparative climatology, clothing design, and building construction. In fact, the excellent work carried out on clothing by the Office of The Quartermaster General, U. S. Army, had to be accomplished mainly without the use of "cooling power." There is much left to be done in the study of the effects of strong winds on clothed bodies. More open-air studies can be expected to yield stress data, integrating the effects of heat and humidity.

Even if all these formulas were on hand, their climatological use would be impaired by the lack of data on wind velocity near the ground, radiation, vapor pressure, and soil temperature. Typical climatological data are more useful to a botanist than to a physician. On a climatological map the summer climate of, say, El Paso and San Antonio, Texas, look nearly alike. But the difference in vapor pressure makes the two localities very different in their habitability.

As far as solar ultraviolet radiation is concerned, it is still uncertain whether it is beneficial in all cases or not. Its absence has sometimes been observed to coin-

cide with uncleanliness, malnutrition, darkness, cold, and similar conditions. A simple recording ultraviolet radiation meter might further the use of sharply defined ultraviolet doses in climatotherapy.

Pollens and industrial aerosols are well known and frequent forms of noxious constituents of the air. Detailed studies, especially of the dependency of air pollution on weather, would facilitate the protection of the community.

In contrast to the situation outlined above, outdoor experiments made in the field of "statistical bioclimatology" are far more numerous than those made in the climatic chamber. The solution of these problems might not be found by additional simultaneous observations of the outbreak of illness and weather elements or of weather complexes such as fronts. There are too many unknown factors which might influence the result. If one single element, such as ozone, is thought to be responsible, every effort should be made for the study of its effects, using physiological methods and climatic chambers. Generally, we urgently need physiological methods indicating small discomforts from weather and clinical methods for the detection of the slightest and earliest traces of the various illnesses referred to above. With such data available, a statistical correlation of meteorological and medical phenomena could become fruitful.

After completion of this manuscript the excellent monograph, "Physiology of Heat Regulation and the Science of Clothing" [64], appeared. It describes in full the field of research which is briefly discussed here. The monograph is based mainly on the large number of investigations in laboratories in the United States during World War II.

REFERENCES

1. Adolph, E. F., and Associates, *Physiology of Man in the Desert*. New York, Interscience, 1947.
2. Amelung, W., "Abhängigkeit der Erkältungskrankheiten vom Klima und Wetter." *Dtsch. med. Wschr.*, 66:85–89 (1940).
3. Armstrong, H. G., *Principles and Practice of Aviation Medicine*, 2nd ed. Baltimore, Williams & Wilkins, 1943.
4. —— Haber, H., and Strughold, H., "Aero Medical Problems of Space Travel." *J. Aviat. Med.*, 20:383–417 (1949).
5. Armstrong, H. G., and others, *Symposium on Space Medicine*. Univ. of Illinois Publ. (in press).
6. Bach, E., und Schluck, L., "Untersuchung über den Einfluss von meteorologischen, ionosphärischen und solaren Faktoren sowie der Mondphasen auf die Auslösung von Eklampsie und Präeklampsie." *Zbl. Gynäk.*, 66:196–221 (1942).
7. Barcroft, H., and Edholm, O. G., "The Effect of Temperature on Blood Flow and Deep Temperature in the Human Forearm." *J. Physiol.*, 102:5–20 (1943).
8. Belding, H. S., *Protection against Dry Cold*. Quartermaster, Climatic Res. Lab., Lawrence, Mass., 1949.
9. Berg, H., "Sonnenflecke, solare Vorgänge und biologisches Geschehen." *Grenzgeb. Med.*, 2:328 (1949).
10. —— u.a., "Dr. Manfred Currys *Bioklimatik*, eine Diskussion." *Grenzgeb. Med.*, Bd. 2 (1949).
11. Blockley, W. V., and Taylor, C., *Studies on Human Tolerance for Extreme Heat*. First Summary Rep., Univ. of California, Los Angeles, 1948.
12. Blum, H. F., "The Physiological Effects of Sunlight on Man." *Physiol. Rev.*, 25:483–530 (1945).
13. Buettner, K., "Die Wärmeübertragung durch Leitung und Konvektion, Verdunstung und Strahlung in Bioklimatologie und Meteorologie." *Veröff. preuss. meteor. Inst.*, Bd. 10, Nr. 5, 37 SS. (1934).
14. —— "Über die Wärmestrahlung und die Reflexionseigenschaften der menschlichen Haut." *Strahlentherapie*, 58:345–360 (1937).
15. —— "Erythembildung durch Sonnen- und Himmelsstrahlung." *Strahlentherapie*, 61:610–615 (1938).
16. —— *Physikalische Bioklimatologie*. (*Probl. kosm. Phys.*, Bd. 18.) Leipzig, Akad. Verlagsges., 1938.
17. —— "Neue physikalische Ergebnisse zum Wärme- und Wasserhaushalt des Menschen." *Wien. med. Wschr.*, 94:373–374 (1944).
18. —— "Bioklimatologie" in *FIAT Rev. Germ. Sci., 1939–1946, Meteorology and Physics of the Atmosphere*, R. Mügge, sen. ed. Off. Milit. Govt., Germany, Field Inform. Agencies, Tech. Wiesbaden, Dieterich, 1948. (See pp. 38–45)
19. —— "Protective Clothing for Heat and Cold" in *German Aviation Medicine, World War II*, Vol. 2, pp. 876–886. Prepared under the auspices of the Surgeon General, U. S. Air Force. U. S. Govt. Printing Office, Washington, D. C., 1950. "Physical Heat Balance in Man." *Ibid.*, pp. 766–791.
20. —— "Conflagration Heat." *Ibid.*, pp. 1167–1187.
21. —— "Effects of Extreme Heat on Man; Protection of Man against Conflagration Heat." *J. Amer. med. Ass.*, 144:732–738 (1950).
22. —— "Effects of Extreme Heat on Human Skin." *J. appl. Physiol.* (in press).
23. —— *Bioclimatology of Manned Rocket Flight*. Univ. of Illinois Publ. (in press).
24. Burton, A. C., "Human Calorimetry II. The Average Temperature of the Tissues of the Body." *Nutr. Abstr. Rev.*, 9:261–280 (1935).
25. Carslaw, H. S., and Jaeger, J. C., *Conduction of Heat in Solids*. Oxford, Clarendon Press, 1947.
26. Coblentz, W. W., "Bioclimatic Measurements of Ultraviolet Solar and Sky Radiation in Washington, D. C., 1941–1944." *Bull. Amer. meteor. Soc.*, 26:113–117 (1945); "Measurements of Biologically Effective Ultraviolet Solar and Sky Radiation in Washington, D. C., 1941–1946." *Ibid.*, 28:465–471 (1947); "Correlation of Bioclimatic Ultraviolet and Total Solar Radiation in Washington, D. C., 1941–1948." *Ibid.*, 30:204–207 (1949).
27. Court, A., "Wind Chill." *Bull. Amer. meteor. Soc.*, 29:487–493 (1948).
28. Courvoisier, P., "Luftdruckschwankungen und Wetterfühligkeit." *Arch. Meteor. Geophys. Biokl.*, (B) 1:115–126 (1949).
29. Curry, M., (pseudonym Maria Hesters), *Schriften der deutschen Kriegsmarine*. Zirka 1943–44.
30. —— *Bioklimatik*, 2 Bde. Riederau, Ammersee, 1946.
31. de Rudder, B., *Grundriss einer Meteorobiologie des Menschen*, 2. Aufl. Berlin, J. Springer, 1938.
32. Düll, B., *Wetter und Gesundheit*, Teil 1. Leipzig, T. Steinkopff, 1941.
33. —— und Düll, T., "Zur Frage solaraktiver Einflüsse auf die Psyche." *Z. ges. Neurol. Psychiat.*, 162:495–504 (1938).
34. —— "Kosmisch-physikalische Störungen der Ionosphäre,

Troposphäre und Biosphäre." *Biokl. Beibl.*, 6:65–76, 121–134 (1939).

36. Ficker, H. v. und de Rudder, B., *Foehn und Foehnwirkungen.* Leipzig, Akad. Verlagsges., 1943.

37. Flach, E., *Atmosphärisches Geschehen und witterungsbedingter Rheumatismus.* Dresden, T. Steinkopff, 1938.

38. Frankenburger, W., und Hammerschmid, H., "Ultraviolett-Messungen natürlicher und künstlicher Strahlungsquellen mit dem UV-Dosimeter." *Beitr. Geophys.*, 53:88–110 (1938).

38. Gagge, A. P., "Standard Operative Temperature, a Single Measure of the Combined Effect of Radiant Temperature, of Ambient Air Temperature and of Air Movement on the Human Body" in *Temperature, Its Measurement and Control in Science and Industry.* New York, Reinhold, 1941. (See pp. 544–552)

39. —— and Shaw, R. S., "Aviation Medicine" in *Medical Physics*, Vol. II, O. Glasser, ed. Chicago, The Year Book Publishers, 1950. (See pp. 41–65)

40. Hall, F. G., and others, *The Significance of Thermal Balance in Testing Gloves and Boots.* Aeromedical Lab., HQAMC USAF, Wright Field, 1948.

41. Hand, I. F., "Atmospheric Contamination over Boston, Massachusetts." *Bull. Amer. meteor. Soc.*, 30:252–254 (1949).

42. Hardy, J. D., "The Radiating Power of Human Skin in the Infrared." *Amer. J. Physiol.*, 127:454–462 (1939).

43. Hausser, I., "Über spezifische Wirkungen des langwelligen ultravioletten Lichts auf die menschliche Haut." *Strahlentherapie*, 62:315–322 (1938).

44. —— "Über Einzel- und Kombinationswirkungen des kurzwelligen und langwelligen Ultravioletts bei Bestrahlung der menschlichen Haut." *Naturwissenschaften*, 27:563–566 (1939).

45. Henriques, F. C., Jr., "Studies in Thermal Injuries, V." *Arch. Pathol. Lab. Med.*, 43:489–502 (1947).

46. Henschke, U., und Schultze, R., "Untersuchungen zum Problem der Ultraviolett-Dosimetrie; über Pigmentierung durch langwelliges Ultraviolett." *Strahlentherapie*, 64:14–42 (1939); "Untersuchungen zum Problem der Ultraviolett-Dosimetrie; Wirkung der Sonnenstrahlung auf die Haut." *Ibid.*, 64:43–58 (1939).

47. Hess, V. F., and Eugster, J., *Cosmic Radiation and Its Biological Effects.* New York, Fordham University Press, 1949.

48. Hornberger, W., "Decompression Sickness" and "Pressure Breathing and Its Physiological Effect" in *German Aviation Medicine, World War II*, Vol. I, pp. 354–394, 482–486. Prepared under the auspices of the Surgeon General, U. S. Air Force. U. S. Govt. Printing Office, Washington, D. C., 1950.

49. Houghten, F. C., and Yagloglou (Yaglou), C. P., "Determining Lines of Equal Comfort." *Trans. Amer. Soc. Heat. Vent. Engrs.*, 29:163–176 (1923); "Cooling Effect on Human Beings Produced by Various Air Velocities." *Ibid.*, 30:193–212 (1924). Yaglou, C. P. and Miller, W. E., "Effective Temperature with Cooling." *Ibid.*, 31:89–99 (1925).

50. Humboldt, A. v., *Kosmos.* Stuttgart, J. G. Cotta, 1845. (See p. 30)

51. Huntington, E., *Mainsprings of Civilization.* New York, Wiley, 1945.

52. Kiepenheuer, K. O., Brauer, I., und Harte, C., "Über die Wirkung von Meterwellen auf das Teilungswachstum der Pflanzen." *Naturwissenschaften*, 36:27–28 (1949). Brauer, I., "Experimentelle Untersuchungen über die Wirkung von Meterwellen verschiedener Feldstärke auf das Teilungswachstum der Pflanzen" (in press).

53. Krebs, A. T., *The Possibility of Biological Effects of Cosmic Rays in High Altitudes, Stratosphere and Space.* Med. Dept., Field Lab. of U. S. Army, Fort Knox, Kentucky, 1950.

54. Langen, D., "Über die Vergleichbarkeit der mit dem Dosimeter gemessenen UV-Werte." *Strahlentherapie*, 66:530–534 (1939).

55. Linke, F., "Die Luftkörperanschauung." *Z. ges. phys. Ther.*, 37:217–220 (1929).

56. Mills, C. A., *Medical Climatology.* Springfield, Ill., C. C. Thomas, 1938.

57. —— "Climatic Factors in Health and Disease" in *Medical Physics*, Vol. I, O. Glasser, ed. Chicago, The Year Book Publishers, 1944. (See pp. 232–244)

58. Missenard, A., *Étude physiologique et technique de la ventilation.* Paris, L. Eyrolles, 1933.

59. —— *Der Mensch und seine klimatische Umwelt.* Stuttgart, 1938.

60. Mörikofer W., "Meteorologische Strahlungsmessmethoden." *Hand. biol. ArbMeth.*, Bd. 2, Nr. 3, SS. 4005–4245 (1939).

61. —— "Zur Methodik medizinisch-meteorologischer Untersuchungen an Hand der Frage der Beziehungen zwischen Anfällen von Spasmophilie und ultravioletter Sonnenstrahlung." *Experientia*, 5:86–88 (1949).

62. Moritz, A. R., and others, "Studies of Thermal Injury." *Arch. Pathol. Lab. Med.*, 43:466–488 (1947).

63. Nelson, N., and others, "Thermal Exchanges of Man at High Temperatures." *Amer. J. Physiol.*, 151:626–652 (1947).

64. Newburgh, L. H., ed., *Physiology of Heat Regulation and the Science of Clothing.* Philadelphia, W. B. Saunders, 1949.

65. Petersen, W. F., *The Patient and the Weather*, Vols. 1–3. Ann Arbor, Mich., Edwards Brothers, 1935.

66. —— "Human Organic Reactions to Weather Changes." *Bull. Amer. meteor. Soc.*, 21:170–175 (1940).

67. —— *Man, Weather, Sun.* Springfield, Ill., C. C. Thomas, 1947.

68. Pfleiderer, H., Verh. 3. Int. Lichtkongresses. Wiesbaden, 1936.

69. —— und Buettner, K., *Bioklimatologie, Handbuch der Bäder und Klimaheilkunde.* Berlin, J. Springer, 1940.

70. Pfleiderer, H., und Less, L., "Die klimatischen Ansprüche an die Atemwege des menschlichen Körpers." *Biokl. Beibl.*, 2:1–4 (1935).

71. Pinson, E. A., "Evaporation from Human Skin with Sweat Glands Inactivated." *Amer. J. Physiol.*, 137:492–503 (1942).

72. Prandtl, L., *Abriss der Strömungslehre.* Braunschweig, F. Vieweg & Sohn, 1931.

73. Reuter, H., "Über die Theorie des Wärmehaushaltes einer Schneedecke." *Arch. Meteor. Geophys. Biokl.*, (A) 1:62–92 (1948); "Zur Hyperthermie der menschlichen Haut." *Ibid.*, (B) 1:9–16 (1948); "Zur Theorie des Wärmehaushaltes strahlungsdurchlässiger Medien." *Tellus*, Vol. 1, No. 3, pp. 6–14 (1949).

74. Riemerschmid, G., "The Climatic Strain in Human Beings as Indicated by Cooling-Ball Temperature Measurements in the Union of South Africa." *S. Afr. med. J.*, 15:267–275 (1941).

75. Robinson, S., Turrell, E. S., and Gerking, S. D., "Physiologically Equivalent Conditions of Air, Temperature and Humidity." *Amer. J. Physiol.*, 143: 21–32 (1945).

76. Ruff, S., und Strughold, H., *Grundriss der Luftfahrt-Medizin*, 2. Aufl. Leipzig, J. A. Barth, 1944.

77. Schaefer, H. J., "Evaluation of the Present-Day Knowledge of Cosmic Radiation at Extreme Altitude in Terms of the Hazard to Health." *J. Aviat. Med.*, 21:375-394 (1950).

78. Schickele, E., "Environment and Fatal Heat Stroke." *Milit. Surg.*, 100:235-256 (1947).

79. Siple, P. A., and Passel, C. F., "Measurements of Dry Atmospheric Cooling in Subfreezing Temperatures" in "Reports on Scientific Results of the United States Antarctic Service Expedition, 1939-1941." *Proc. Amer. phil. Soc.*, 89:1-389 (1945). (See pp. 177-199)

80. Stone, R. G., "On the Practical Evaluation and Interpretation of the Cooling Power in Bioclimatology (1)." *Bull. Amer. meteor. Soc.*, 24:295-305 (1943).

81. Storm van Leeuwen, W., u. a., "Studien über die physiologische Wirkung des Föhns." *Beitr. Geophys.*, 44:400-439 (1935).

82. Takata, M., und Murasugi, T., "Flockungszahlstörung im gesunden menschlichen Blutserum 'Kosmo-terrestrischer Sympathismus.'" *Biokl. Beibl.*, 8:17-26 (1941).

83. Wezler, K., "Physiological Fundamentals of Hyperthermia and Pathological Physiology of Heat Injury" in *German Aviation Medicine, World War II*, Vol. II. Prepared under the auspices of the Surgeon General, U. S. Air Force. U. S. Govt. Printing Office, Washington, D. C., 1950.

84. Winslow, C.-E. A., Herrington, L. P., and Gagge, A. P., "New Method of Partitional Calorimetry." *Amer. J. Physiol.*, 116:641-655 (1936); "Physiological Reactions of the Human Body to Varying Environmental Temperatures." *Ibid.*, 120:1-22 (1937); "Heat Exchange and Regulation in Radiant Environments above and below Air Temperature." *Ibid.*, 131:79-92 (1940).

SOME PROBLEMS OF ATMOSPHERIC CHEMISTRY*

By H. CAUER

Institute for Chemical Climatology, Hohenberg a. d. Eger, Germany

INTRODUCTION

Although a good basis for chemical investigations in the atmosphere was established during the past century by Smith [53], and individual investigations such as those by Schmauss and Wigand [51] have been of great value, a full incorporation of chemical concepts and procedures into the consideration of meteorological phenomena has not, up to the present, become general. This is due primarily to the fact that satisfactory methods for the chemical analysis of air do not yet exist, and furthermore that none of the available methods lend themselves to the routine procedures which have been possible in the measurement of physical phenomena. At present, experimental work of this kind requires a good chemical-analytical knowledge and many years of experience if passably reliable results are to be obtained. Since the pertinent methods consist of special techniques for detecting traces, their mastery cannot be expected by physicists, even those whose minor field is chemistry or physical chemistry. On the other hand, although the analytical chemist has the necessary specialized knowledge, including the requisite information in the physico-chemical field, he lacks a fundamental insight into the problems of physical meteorology.

Therefore, the most important task in the development of chemical meteorology and climatology appears to the author to be the training of analytical chemists who study physical chemistry and physical meteorology as secondary fields. A second requirement, and one which will necessitate the active participation of the analytical chemist, is the further development of methodology. Not until the development has been more or less completed will it be possible to undertake a third major task, that of making many parallel investigations in micrometeorological as well as world-wide networks under the most varied atmospheric conditions and at various heights above the ground.

METHODS

As a survey of the chemical methods available today, two groups of methods developed in Europe will be discussed briefly: (1) the absorption-tube method for detecting the presence of secondary gaseous substances, and (2) the condensation method for the analysis of water-soluble constituents of the air which form condensation nuclei.

Absorption-Tube Methods for Detecting Traces of Gaseous Materials. The iodine of the air is determined, in principle, according to the method developed by von Fellenberg [25], that is, by absorption in a very dilute aqueous K_2CO_3 solution and final determination by colorimetric or titrimetric techniques. This method has been greatly simplified by the author with the apparatus shown schematically in Fig. 1. It consists of a motor-driven pump, dry-gas meter, and Cauer absorption tube. As a result of control investigations by Quitmann [44], this equipment has been simplified and can be used for serial and uninterrupted tests. It is made in three sizes [7]: an apparatus mounted on a stand equipped with an absorption tube (circulation capacity 10 m^3 hr^{-1}); a readily portable apparatus provided with an absorption tube (circulation capacity 0.6 m^3 hr^{-1}); and a smaller apparatus for measuring small amounts. The use of this equipment has been described elsewhere [5]. The chemical analysis requires 60–90 min.

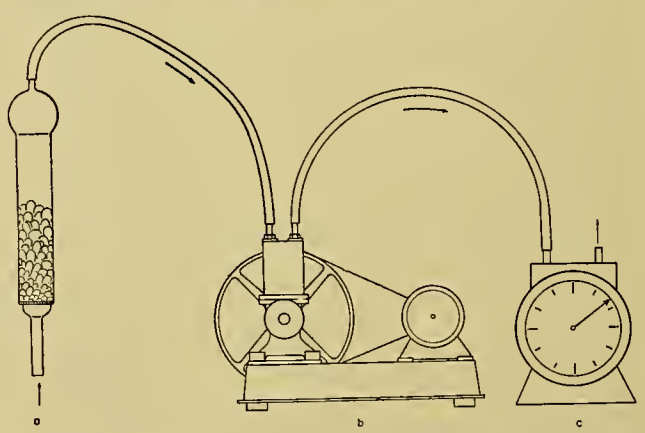


FIG. 1.—Schematic representation of apparatus for detecting traces of gaseous materials; (*a*) Cauer absorption tubes without separate charging vessel; (*b*) motor and pump; (*c*) dry-gas meter.

Depending on the concentration present, the total chlorine (chloride, free chlorine, chlorine dioxide) is determined, according to Cauer [6], with one of the devices mentioned above, using a 1.5 per cent solution of KOH. Final determination is made by titration with $AgNO_3$. According to the data given by Nitschke [41], best results are obtained in alcoholic solution after the method of Bang and Blix. One-sixth of the free chlorine escapes analysis. The air sampling requires 10–60 min, the chemical analysis about 1 hr.

The sulfuric acid or the sulfates and sulfites are determined by two methods. The first is a nonquantitative standard method developed by Liesegang [38], which measures the quantity of sulfates plus sulfites precipitated during a period of 100 hr on a porcelain bell equipped with an extraction thimble. The end determination is made gravimetrically as $BaSO_4$. The method

* Translated from the original German.

gives an insight into the distribution of the sulfate up to 50 km leeward of intense sources of gaseous effluents. Exact instantaneous measurements can, of course, be made only with the apparatus discussed above in the paragraph on the determination of iodine. For this purpose from 0.05 to 20 m^3 of air (depending on the concentration) are washed in aqueous solutions of alkali hydroxides or carbonates, a process which requires 5–120 min. Subsequent analysis requires about 60 min.

Hydrogen sulfide is determined, according to Quitmann [45], by absorption in 2–3 cc of acetic cadmium acetate solution, with the smallest Cauer apparatus. For this measurement, 300–500 l of air are washed in about 60 min. The end determination is performed with *KI* solution in approximately 30 min.

Nitric acid is determined in the same way as sulfuric acid, in dilute aqueous solutions of alkali hydroxides, using the same equipment. After reduction with metallic zinc, the nitrate is determined as nitrite with Griess's reagent [47]. The time required corresponds to that needed for the determination of H_2SO_4.

Ozone is determined, according to Cauer [6], with the aid of the intermediate-size apparatus described previously, which is provided with an absorption tube (circulation capacity 0.6 m^3 hr^{-1}). The classical method of detection, based on reaction with *KI* in a neutral aqueous solution, is used. From 100 to 300 l of air are washed within 10–30 min in 10 cc of 0.2 to 0.6 per cent aqueous $NaC_2H_3O_2$ solution (*p*H 6.7–7.1) which contains 50 μg iodine ions as *KI*. The iodide is oxidized to I_2 by O_3 and the I_2 is removed by the air stream. Therefore, in the end determination the residual nonoxidized iodide is analyzed according to the method of von Fellenberg [25], and the difference is computed in terms of O_3. For experienced chemists, this analysis requires 20 min, but necessitates great care, and is difficult for nonchemists. By working in shifts, four trained operators can make two or even three parallel series of hourly measurements for several weeks at a stretch. The method is inapplicable if the air contains relatively large quantities of nitrite and of free halogens which, because of their acidic character, nullify the buffer effect. Unpublished experiments, made on the North Sea, have shown that even at a high natural content of total Cl_2 the buffer effect is not impaired, because the dilute free chlorine present is evidently converted very rapidly into chlorine dioxide. A corresponding absence of significant errors has also been noted in the presence of large quantities of natural nitrite [13]. A similar method, published later by V. Regener [49], must be considered a close parallel to the foregoing technique because of the addition of the buffer effect contributed by Ehmert. Unfortunately, because of difficulties inherent in the apparatus, this procedure is even more awkward to carry out than Cauer's method.

Simplification of the method for O_3 determination appears to be an important problem. In the author's opinion, the approach to this problem is hardly to be found in deep cooling and absorption in silica gel (because of increased possibility of error), but perhaps in the oxidizing effect of O_3 on aldehydes, as attempted by Briner and Perrottet [4]. The approach used by Dirnagl [22] and Curry [19] for developing automatic recording of the *KI* method has much to recommend it. This apparatus is a development of the automatic recorder for the determination of H_2S, according to Kraus [32]. According to their technique, a large unmeasured quantity of air is blown mechanically on a filter paper soaked with a neutral *KI* solution. At stipulated time intervals a fresh portion of the filter paper is exposed by clockwork. The resulting coloration can be calibrated by the exact methods discussed previously. An improvement of this method and a reduction of its cost would offer new possibilities for more extensive comparative investigations. In this respect, to be sure, technical deficiencies cannot be justified by assuming that an additional oxidizing factor, such as O_x, might be present, or even that the total oxidation value is being measured [22]. The latter is certainly unlikely, for NO_2 and ClO_2 oxidize in acidic solution at the *p*H value used, but by no means completely.

The total oxidation value is determined, according to Cauer [6], in the same way as is O_3; not in neutral solution, but in H_2SO_4 solution (0.05 per cent) having a *p*H below 2.8. The oxidized I_2 is calculated in terms of active *O*, in which two iodine atoms correspond to one oxygen atom. The time required for sampling and analysis is the same as for O_3.

The total reduction value, according to Quitmann [44], is determined with the aid of the smallest of the Cauer absorption tubes, by washing 20–50 l of air in 3 cc of a concentrated $K_2Cr_2O_7$–H_2SO_4 solution for 15–30 min. The final determination consists of the addition of *KI* and titration of the liberated iodine with thiosulfate. The difference between this iodine value and the one which would have been determined if no chromate had been reduced by substances in the air corresponds to the reduction effect of the air and can again be computed (according to the formula: 2 iodine = 1 oxygen) in terms of the active oxygen missing for the oxidation of the reducing substance (see Table III). In this way the total oxidation values can be compared directly with the reduction values. If the reduction value is divided by the oxidation value, the quotient represents the *Redox value* of the air. During World War II a simple, rapid method for determining the reduction value, designed for hygienic purposes, was developed by the author. For meteorological purposes, this method, as yet unpublished, might also be applicable to air that is not too humid.

Condensation Method for Determining Traces of Substances Which Form Condensation Nuclei. Condensation nuclei are obtained, according to Quitmann and Cauer [47], or according to H. Cauer and G. Cauer [15], by cooling the highly polished metal surfaces of a Cauer condensation sphere below the dew point but not to the freezing point. The water in the thin layer of air very close to the metallic surface is rapidly attracted by chemical substances suitable as nuclei, and droplets are

subsequently deposited.[1] As a result of thermal convection, this process is continuous in the newly advected air, and fog particles, proportional in number to the mass of air flowing close to the metallic surface, adhere to it and finally drip off into a graduated tapered vessel. The condensation apparatus (Fig. 2) consists of an egg-shaped dew-cup which is plated with *Cr* or *Ni*, is well-polished, and has a capacity of two liters of ice water.

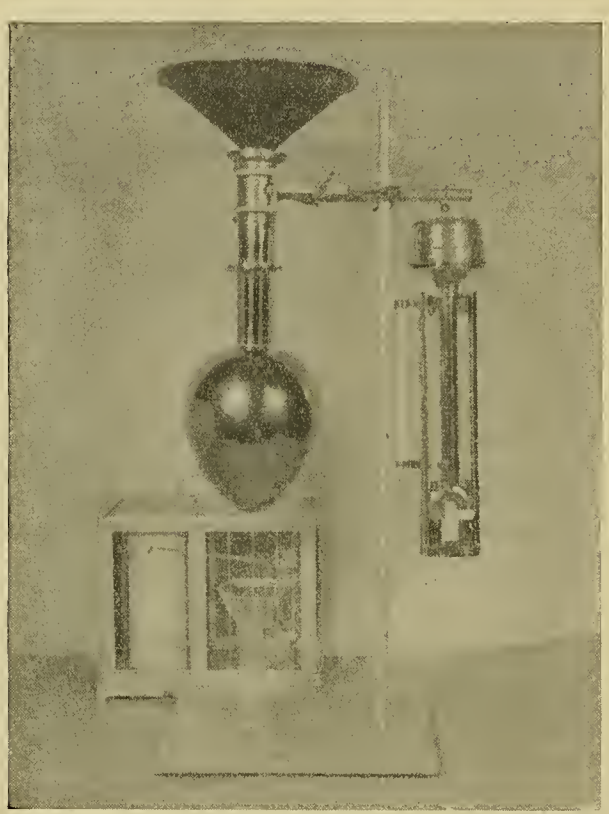


FIG. 2.—Apparatus for the analysis of traces which form condensation nuclei. Center, Cauer condensation dew-cup; at right, Assman psychrometer; below, thermohygrograph.

During the time required for the dew deposit, 20 min in most cases, the mean absolute water content of the air must be determined in the immediate vicinity with the aid of an Assmann psychrometer or a thermohygrograph. Depending on the number of analyses and the substances to be investigated, amounts of condensate ranging from 1 to 20 cc are collected. During ventilation the sphere must be shielded from direct sunlight and dust. Hence it cannot be used during storms; likewise, it cannot be used when the absolute humidity of the air is small. During cold weather a coating of hoarfrost is produced by charging the sphere with a refrigerant.

Insofar as possible, the chemical substances are analyzed by the methods usual in water analysis [47]. For other substances special trace-detection methods are used, for example, spot analyses [15, 24]. The number of analyzable substances can be increased greatly with the help of a spectrograph. For this purpose a measured quantity of dew is carefully evaporated in a platinum dish and the residue is picked up with the heated carbon rods of the spectrograph. Several standard solutions of different concentration are treated in the same way. The subsequent spectrophotographs show the presence and order of magnitude of substances in the dew by the spectrographic lines and their intensity. From the measured concentration of substances in the dew, and on the basis of the simultaneous measurement of the mean absolute humidity, the quantity of chemical substances capable of forming condensation nuclei is computed for one cubic meter of air. It was unexpectedly found that the highly hygroscopic compounds can be determined more accurately in this way than by the absorption-tube methods mentioned previously. When frost deposits are used, however, the results must be multiplied by a factor of three [15], a fact which was still unknown at the time of the first publication of the method [47], and for which no adequate theoretical explanation has been found.

By this method the halides, sulfates, ammoniacal substances, soluble alkali compounds, and alkaline-earth compounds as well as organic materials such as formaldehyde can be quantitatively determined. Metals can be determined particularly well with the aid of the spectrograph. Other substances such as nitrates and nitrites can also be determined with the aid of the condensation sphere, but only that portion of the substance which can act as condensation nuclei, depending on the relative humidity and pressure in a given case. The physico-chemical reasons for this phenomenon are still unexplained. The nonpolar gaseous components of the air, such as the two principal components, O_2 and N_2, cannot be analyzed by this method.

RESULTS FROM ABSORPTION-TUBE METHODS

Iodine of the Air. Contrary to previous assumptions, the iodine in the air over central Europe [10] does not originate directly from the ocean but comes principally from the combustion of iodine-containing ocean products for the extraction of iodine and for domestic heating in coastal towns. The principal sources of the iodine in the air over Europe consequently are Brittany, Scotland (Shetland Islands), Scandinavia (Stavanger region), and to a minor extent Spanish Galicia (Finisterre).

The iodine is lifted off the ground by convection cells and is transported in the air masses as far as the Carpathians; in the case of intense storms this distance is occasionally covered in less than twenty-four hours. The dilution occurring during this process is at first very marked, but later it proceeds at a continuously decreasing rate. Under normal conditions, the mean iodine content along the coast of Brittany is 0.4 $\mu g\ m^{-3}$; during the combustion of seaweed (*laminaria* and *fucus*) it runs to 6000 $\mu g\ m^{-3}$. In central Europe it

1. The direct condensation of steam on cooled, polished, metal surfaces without the existence of condensation nuclei amounts to almost one per cent of the total possible precipitation, according to Mayer [39] (this percentage also includes the direct condensation on the surface of the water). This occurrence is so low that the validity of the following chemical analysis is not affected.

normally ranges up to 0.05 $\mu g\ m^{-3}$, but during the intrusion of air masses from districts where the smoldering of seaweed is practiced it is 0.6 $\mu g\ m^{-3}$ (von Fellenberg's mean [25]). In eastern central Europe (Szepes, Slovakia) the normal is 0.08 $\mu g\ m^{-3}$, but the intrusion of air from seaweed-burning regions raises this value to 0.13 $\mu g\ m^{-3}$. If air from the interior of the continent (Russia) reaches central Europe, the mean there drops to only 0.1 $\mu g\ m^{-3}$. If the natural mean, that is, that quantity of iodine which directly and solely originates from the sea, is calculated, the value found for central Europe is 0.0025 $\mu g\ m^{-3}$. The higher values actually found reveal the effects of industrial processes, of coal combustion, and of decomposition processes on the earth's surface.

Under the influence of light, and to a lesser extent of O_3, or by autocatalytic processes, there exists an iodine cycle between the gaseous phase and a state wherein the iodine is dissolved in droplets. If the iodine is dissolved as iodide ion, it is easily oxidized to molecular iodine, escapes as gas from its solution, and when further oxidized forms iodine pentoxide (which is intensely hygroscopic and nucleogenic). However, by reduction of the free iodine to hydrogen iodide and subsequent formation of condensation nuclei, the iodine can again return to the dissolved state.

Ozone of Air Strata near the Ground. Repeated measurements [8, 9, 13, 14, 50, 54] show that the concentration of O_3 in the air near the ground is, on the average, independent of the height of the sampling point above sea level. Means and deviations shown in Table I support this conclusion. Only a mean of individual measurements made by Ehmert [23] during flights at altitudes from 1000 to 8000 m above the ground is essentially higher than the other means. The peak values were found by Ehmert at the level of the cumulus and stratus cloud cover. By using optical methods, E. Regener [48] also found a few high values at greater heights in the troposphere during thunderstorm conditions, that is, at a time when air was not only being transported downward from above, but also upward out of the thunderstorm region where electrical equalization processes between clouds were taking place. For this reason, O'Brien, Mohler, and Stewart [42, 43] were unable to confirm these findings for stable anticyclonic conditions. These authors found smaller concentrations of O_3 in the lower stratosphere than at the ground. In such cases, Regener visualizes a damming effect of the O_3 flow at the ground, a phenomenon that, however, would be possible only if the reducing substances in the air were present in far smaller quantities, and if the oxidation power of O_3 were small enough to resemble that of O_2. In the "pure" mountain air of Ober Schreiberhau, Cauer [12] found reduced substances to be in excess by a factor of 160, that is, there was an active oxygen deficiency. Laboratory experiments have shown that the O_3 from a limited O_3 source does not decompose rapidly, but reacts almost instantaneously with the reducing substances of the air so that after a brief time nothing further can be detected save the odor, which evidently characterizes the reaction products rather than the O_3. Since the reducing substances originate principally from anaerobic decomposition processes of organic material in soils and waters whence they rise upward almost continuously, the O_3, assuming its exclusively stratospheric origin, could appear at ground level only at intervals and for short periods during conditions of subsidence from greater altitudes. Likewise, an accumulation of O_3, whatever its source, can occur only where the reducing substances in the air or the decomposing organic substances in the soil are lacking. This would probably be the case in extended desert regions, such as the Atacama Desert.

Table I. Surface Ozone Concentrations in Europe

Location	Altitude msl (m)	Mean ($\mu g\ m^{-3}$)	Range ($\mu g\ m^{-3}$)	Number of analyses
Friedrichshafen, Lake of Constance	1000–8000	38*	23.8– 77.8*	14
Jungfrau, Switzerland	3450	11.9	4.0– 30.0	20
Weszterheim, Tatra, Slovakia	1010	30.0	5.0–102.0	412
Ober Schreiberhau, Riesengebirge	750	16.8	0.3– 36.0	351
Glatzer Bergland, Silesia	400–800	14.5	0.0– 40.0	448
Koenigstein, Taunus:	400			
More than 20 m from lightning rod				
General		15.9	0.0– 99.9	472
Cloudy weather		18.3	1.3– 99.9	339
Fair weather		10.6	0.0– 35.0	133
Not more than 0.02 m from lightning rod				
General		31.7	1.9–189.1	128
Cloudy weather		46.9	5.2–189.1	77
Fair weather		11.4	1.9– 87.5	51
Island of Norderney, North Sea:	20			
General		24.8	5.5– 73.4	181
Near a grounded conductor tip (~0.02m)		35.5	3.1–126.1	48
Berlin, radio antenna mast	175	6.0	0.0– 20.0	50

* Approximate values.

Therefore, it is hard for a chemist to conceive of the generally assumed presence of an "ozone flow" from the stratosphere, whose richest O_3-bearing strata at 22–26 km altitude contain no more than about 200–300 $\mu g\ m^{-3}$, especially because the short-wave insolation causes, in addition to O_3 formation, an almost equally intense dissociation at high altitudes. Hence, the chemist is compelled to seek other continuously active and especially more copious O_3 sources than that which the stratosphere can offer, even under optimum conditions (except in the case of foehn). Moreover, the results of a series of measurements made in the Riesengebirge [50, 54] and in the Taunus [13] agree with these considerations. According to these results, strong fluctuations are usual even for stable weather conditions without high-reaching turbulence; unusually low or even zero values were found for air that had subsided from a considerable altitude (above 3000 m) during foehn conditions or as a consequence of subsidence in an anticyclone over the plains. In other words, the zero values occur in air with few large nuclei, but with relatively considerable quantities of motile small nuclei, that is,

during conditions of sharply increased conductivity or low electric potential. This phenomenon is strikingly parallel to the low O_3 values in air that has been enriched by radioactive emanations, that is, the highly conductive air at radioactive spas [9, 19]. The objection could be raised here that O_3 is produced during the decay of radium. This is, indeed, the case for the principal members of the decay series, or rather, O_3 is produced by their electrons which are either not decelerated or decelerated only slightly. However, the quantity of O_3 thus produced is extraordinarily small, although numerous excited and ionized atoms and molecules will simultaneously be formed by secondary radiation. Therefore the principal effect of the emanations from the ground is probably not an increase of the O_3 content but an increase of the ionization of the air. The strongly decreased values observed during foehn can thus be explained by the fact that air subsiding from greater altitudes is already deficient in O_3 as well as by the fact that the O_3 is almost completely eliminated by reduction processes in strata near the ground because, as a result of the high conductivity, no electron accelerations sufficiently great for local O_3 formation can occur from grounded point dischargers in the air near the surface, or between clouds. Equally low values of O_3 are observed at the inception of condensation phenomena during the formation of fog and protracted precipitation with mist, and likewise in the presence of tobacco smoke and cooking vapors indoors. The reaction mechanism for the elimination of O_3 in such cases is still unexplained. It must be assumed that O_3 during fog formation reacts with reducing substances such as ammoniacal compounds. This appears to be the reason for the low values and small fluctuations of O_3 in forests, where particularly intensive exhalations of reducing anaerobic decomposition products occur.

During fair weather, the mean values increase. The air behind a cold front, which has few middle- and large-sized nuclei because of their removal by rain, is frequently still poor in O_3. The O_3 values increase only when the number of small nuclei decreases and that of the large nuclei increases, or, in other words, concurrently with the intermittent increase of the potential occurring frequently during such weather. Similar phenomena may be observed when rather low clouds or high fogs move by and when brief precipitation sets in (Lenard effect). The abrupt increase of the values during subsidence phenomena from *lower* altitudes (height of the lower clouds) is also very informative. This increase, in the author's opinion, is caused by O_3 which originated during electrical discharge processes in nonhomogeneous fields between clouds. However, in addition to this, the formation of O_3 at grounded point dischargers might be of considerable local importance. In fact, it has been shown by investigations on lightning rods [13] that, even in fair weather, values are found around these devices which are three times larger than those recorded farther away. Thus, the quantities of O_3 that emanate from such discharger tips into the surrounding air are considerable as compared with the average total O_3. Ozone from grounded dischargers was recently found by the present author in investigations over a limited measuring field in Wyk on the island of Föhr. In these investigations it was particularly striking to note that considerable absolute differences occurred over distances of a few meters, but that the trend in the fluctuations was everywhere the same without exception. At constant wind the values were consistently higher at one measuring station than at the other, a behavior pattern which was sometimes reversed when the wind changed. From the foregoing observations one gets the impression that sometimes the one measuring station and sometimes the other was subject to the effect of O_3 from a more intense source in the field.

A rough calculation by Reiter,[2] based on findings by Schottky, shows that, with a potential of 130 v m^{-1} on a hemispherical conductor tip of 0.5 cm^2 surface area one meter above the ground, the potential may reach 1,000,000 v m^{-1} over partial conductor points (microscopically small elevations). However, a potential of only 500,000 v m^{-1} is sufficient to produce electrons having the acceleration of 8 ev required for O_3 formation, assuming an average of 10^5 electron impacts before adhesion to gas molecules. The formation of O_3 at grounded conductor tips such as on lightning rods therefore appears possible even in a normal electrostatic field. To what extent this occurs at less elevated conductor tips has not been investigated sufficiently to permit any definite statements. However, chemical measurements indicate that O_3 is formed even with lower points during rather brief intensive fluctuations of the electric field. It is not possible with our present knowledge to appraise the significance of such O_3 formation near the ground as compared with the O_3 production in nonhomogeneous electrostatic fields between clouds, augmented perhaps by O_3 from the stratosphere, although the subsidence of O_3 from great heights has never been satisfactorily demonstrated. For the solution of this problem extensive investigations of O_3 are needed, involving not only measurements from field stations and aircraft, but also analyses with respect to the companion reactions and the reaction products of O_3. Furthermore, the rapid fluctuations (within seconds) of the electrostatic field, and the change of the number of nuclei must be measured simultaneously, as was done in the extensive work by Landsberg in collaboration with Amelung [1].

Total Oxidation of the Air. Fluctuations of the oxidation values are preponderantly parallel to those of O_3, but opposing trends may also occur. In the Taunus region [13] the mean was 15 μg active oxygen per cubic meter, in Norderney [14] it was about 24 $\mu g\ m^{-3}$. Previous oxidation values [11, 12] were always lower as compared to the values derived from the deficiency in active oxygen, or from the reduction values. Where reducing substances are lacking, as in desert regions or in the Arctic, the total oxidation values are probably large and effective. The total oxidation values are produced by the oxidizing effects of O_3, NO_2, ClO_2, Cl_2, Br_2, I_2, and oxides of the latter two halogens. Despite thousands of analyses by means of titanium sulfate,

2. Personal communication.

hydrogen peroxide could be detected only in the immediate vicinity of fireplaces and therefore cannot be considered an oxidizing factor of practical importance.

RESULTS FROM THE CONDENSATION METHOD

This method takes into account merely those condensation nuclei around which water droplets form by condensation, not only during the slight water-vapor supersaturations and pressure fluctuations possible in the free atmosphere, but also even before saturation is attained. The condensation nuclei thus represent distinctly hygroscopic substances, that is, those secondary constituents of the air which possess the most intense molecular attractive forces (capillary forces). Their radii range in order of magnitude from 10^{-7} cm to 10^{-4} cm. In the following discussion the averages and fluctuations obtained thus far will be considered; the data are given in Table II.

Table II. Surface Concentrations of Nuclei-Forming Substances in Europe

Substance	Region	Mean ($\mu g\ m^{-3}$)	Range ($\mu g\ m^{-3}$)
Mg	Mainland and coast	3.1	0.0– 65.2
Cl	Mainland and coast	32.2	0.0–964.8
SO_4	Mainland and coast	2.6	0.0–732.0
NH_3	Mainland	7.9	2.5– 54.4
NH_3	Island of Norderney	5.5	0.0– 25.0
NO_2	Mainland	1.0	0.0– 21.6
NO_2	Island of Norderney	0.14	0.0– 1.2
H_2O_2		0.0	0.0– 0.0
$HCHO$	Mainland	0.5	0.0– 16.0

Chlorine of the Air. Contrary to previous concepts, it is not the sulfates but the chlorides that represent the principal mass of nuclei-producing substances in Europe, except in the more populous regions.

In contrast to the spray theory, according to which sea salts are transported chemically unchanged in droplets of sea water, lose their water in dry air, and function subsequently as condensation nuclei far inland, it was possible to determine that a sharp chemical disintegration of the droplets takes place very soon after lifting from the ocean surface [16, 17]. The ratio of the Cl^- to the Mg^{++}, K^+, and Na^+ on the high seas, as determined during submarine voyages, does not correspond even approximately to the ratios in the sea water, where, for example, a constant weight ratio, $Mg/Cl = 1/15$, is found. The chlorine in the gaseous phase apparently escapes rapidly from the droplets by autocatalytic processes and under the influence of O_3 [6]. However, the chlorine evidently remains in this state only a short time [16, 17]. Under the influence of light quanta it is partly converted with hydrogen into HCl, which, being hygroscopic, produces condensation nuclei, and is partly oxidized to ClO_2, according to data obtained on islands in the North Sea. This ClO_2 is likewise productive of condensation nuclei and dissociates in the dissolved droplets to form $HClO$, or HCl and $HClO_3$. These alternations between the dissolved and gaseous phases probably occur in the air in the nature of a continuous cycle. In the eastern regions of central Europe Cl_2 is occasionally absent from the air, even in the form of chlorides. The invasion of maritime air into these regions (Tatra and Riesengebirge) can often be detected by the fact that chlorides appear and that their amounts increase rapidly. The alkali ions and alkaline-earth ions, which are enriched by the liberation of Cl_2 from the droplets, do not show such a cyclic phenomenon. They combine with CO_2 and diminish rapidly by precipitation.

Nitrogen Compounds of the Air. Nitrites and nitrates appear to be quantitatively dependent on electrical discharge phenomena, on the heating and humidity of the ground (bacterial activity), and on pressure fluctuations [11, 12]. In the air of high mountains they represent important condensation nuclei (nitrite production from rock formations in the Andes or in the Atacama Desert). In contrast, they exist only to a minor degree in maritime air.

Ammonia or ammoniacal compounds appear to depend quantitatively on anaerobic processes of decomposition, combustion processes, and periods of inflorescence (*e.g.*, of the *Prunus* genus) [12], and, next to chlorides, represent the largest group of substances which form condensation nuclei over the European continent. They occur over the entire continent without exception. Only on sand islands having sparse vegetation and human population are zero values found in maritime air. In the case of land winds that have traversed the sea, low values are also encountered.

Formaldehyde of the Air. Formaldehyde was found to be an index of the presence of incomplete combustion processes. No indication of a transportation by subsidence from high altitudes, a process assumed by Dhar and Ram [21] and by Groth and Suess [27], could be found up to the present time. Some of the results have been published elsewhere [28].

Table III. Average *p*H Value of the Aerosol and Mean Reduction Values of the Air

Location	Altitude msl (m)	Mean *p*H value	Mean reduction value (active oxygen deficiency in $\mu g\ m^{-3}$)
Nebelhorn, Alps	1000-2000	3.4	—
High Tatra, Slovakia	>1000	4.0	—
Ober Schreiberhau, Riesengebirge	50	4.7	160
Berlin	25	6.2	1600
Island of Norderney, North Sea	10	4.6	—
Diesel engine rooms	—	5.2	18,800

The *p*H Value of the Aerosol. The mean *p*H value of the condensation nuclei seems to be a function of the mass ratio of the hygroscopic acidic and basic secondary constituents in each case (see Table III). In high mountains (1000 m above sea level), the strongly acidic average *p*H value (bactericidal for pathogenic bacteria) is due to the decrease in ammoniacal substances, so that without the attendant buffer effects the nitrite present in the droplets is able to dissociate strongly. In the mountains at an altitude of only 700 m the mean is forced into the weakly acidic range (which

arrests virulence), because of the increase in the concentration of ammoniacal substances in the nuclei, whereas in Berlin the mean *p*H value is forced almost to the neutral point (which promotes virulence). Conversely, the mean *p*H value on the Island of Norderney again lies in the weakly acidic range, because of the lack of ammoniacal substances. In this case, however, the weakly buffered acids are not nitrites but hydrogen compounds of the halogens originating from the sea.

WORKING HYPOTHESES

Significance of Molecular Forces for Condensation. The unexpected results of quantitative analyses obtained by the condensation method gave the first stimulus to the theoretical consideration of condensation [12]. If gaseous substances subsequently dissolve in suspended drops of pure water, an equilibrium must occur between the gaseous and the dissolved portion of the gas. This precludes a total quantitative determination by analysis of the dissolved portion only. The adequacy of a quantitative determination rests therefore on the assumption that prior to the condensation a certain orderly arrangement exists between the corresponding gaseous chemical substances and the water-vapor molecules. The hygroscopic substances accordingly represent an integral component of numerous condensation nuclei, owing to their attractive forces which are inherently stronger than those which the water molecules exert upon one another. Wilson [59] and Wegener [57] were aware that the laws of thermodynamics are insufficient to explain cloud formation. According to contemporary concepts, as treated by Wolf [60], a gas (water vapor) is not condensable at finite temperatures if only elastic impact forces exist between its molecules in accordance with the Boyle-Mariotte law. To explain condensation (cloud formation), the chemist must therefore have recourse to intermolecular forces and, in the last analysis, to processes of atomic energy. This approach was initiated by Lenard [37] prior to 1914. He was followed by Volmer [56], who, however, did not recognize as clearly as Lenard the great significance of the short-range forces manifest in the nuclei formation *before* water-vapor saturation (*e.g.*, fog formation by NH_4NO_3 produced by O_3 and other factors, or even by artificial acidic fogs $Cl \cdot SO_2 \cdot Cl$). He believed that it was possible to explain water-vapor aggregation and condensation into droplets by relatively strong supersaturation alone without the aid of nuclei. The four types of short-range molecular forces will now be discussed in the light of our present knowledge.

Polar Forces. These forces, which are of electrostatic nature, give rise to mutual orientation of the molecules, particularly that between electrical charge carriers and the dipole. As in the case of the ion atmosphere in solutions treated by Debye [20], it is obvious to assume that neutral dipoles of the water vapor become attached in a more or less oriented manner to charged gas molecules and form a "cluster." At any rate, this clustering can be considered as the origin of the so-called "small ions" in air (water-vapor molecules, $r = 1.4 \times 10^{-8}$ cm; small ions, $r = 7 \times 10^{-8}$ cm). Spatial considerations may require a certain supersaturation before the small ions can grow sufficiently large to become visible. There is room around the ion for only a restricted number of oriented water molecules. Only when supersaturation occurs, or in the course of an increased number of collisions, is it possible for additional, similarly oriented dipoles to penetrate this configuration. In this process the centers of the molecules approach each other and as a result the induction forces, discussed below, become effective and further increase the attractive forces of the molecules. The free rotation due to the thermal motion is thereby considerably diminished and is even nullified during sufficient, say adiabatic, cooling. If charged gas molecules were to adhere to varyingly large groups of water-vapor molecules scarcely, if at all, affected by molecular forces, the so-called "small ions" would increase in size irregularly and at a much lower supersaturation. In the free atmosphere the majority of the small, charged nuclei do not grow into large ones, since small ions vanish by recombination and by association with medium and large nuclei which are produced by more strongly hygroscopic substances (tangling of chains and rings of dipolar substances with water-vapor molecules, resulting in condensation—see below) and which possess an electrical double-layer.

Dipole Forces. These forces, which are also of electrostatic nature, orient the neutral, permanently dipolar water-vapor molecules with respect to each other. This orientation is achieved without the aid of extraneous ions and solely by the polar attractive forces between dipoles. As the temperature of the system is lowered, the free rotation ceases here also. Depending on whether the water dipoles come in contact with each other or with different dipoles (1) a simple super molecular formation may occur with dipoles nonpolarly parallel to each other, (2) a ring formation may occur, or (3) a chain formation may occur when the dipoles are in polar arrangement with subsequent tangling. Tangling phenomena involving dipoles of supercooled liquids (rubber) were observed by Debye. In a conversation with the present author, Hückel and Eucken advanced the idea that water-vapor molecules unaffected by an ion and not in combination with hygroscopic substances occur principally in groups of eight (rings), which are in equilibrium with groups of four (rings) and with single molecules. Because of their closed form, the pure water rings, in the present author's opinion, are probably less suited to formation of nuclei than are the chains having admixtures of foreign substances. In the latter, a more rapid growth and tangling can occur as a result of the greater proximity of the centers of the molecules. Furthermore, for the same reasons, these chains probably resist dissolution from an addition of heat somewhat more strongly than do the ring-shaped, readily dissolvable, pure water rings. Naturally, dipoles of the most varied types, such as NH_3, HCl, H_2SO_4, and H_2CO, unite with the dipole H_2O. The decisive factor for the foregoing process is the dipole moment (the unit of which is 1 "debye" (D) $= 10^{-18}$ electrostatic units) of the two substances, since the attraction at sufficient proximity is propor-

tional to the product of the two values of the dipole moment. The effect of two dipoles upon each other is one of attraction, repulsion, or relative cancellation, depending on their mutual positions. It could be assumed that in gases with freely rotating molecules (thermal motion) these forces mutually cancel in the mean. Actually, however, according to Wolf [60], the rotation is not uniform when dipoles move past each other, since the positions of attraction (smaller potential energy) are preferred to those of repulsion. In addition the dipole movements are probably somewhat affected by the electrostatic field, since by means of X-ray techniques Német [40] found a relatively large effect of the electrostatic field on the structure of ice.

The behavior of azeotropic mixtures is further evidence for the approximate correctness of these concepts. For example, if *HCl* of any concentration is allowed to evaporate at normal pressure, an equilibrium at a concentration of approximately 20 per cent acid is reached in which the ratio of acid molecules to water molecules is 1/6, that is, the same ratio as for an hydrochloric acid hydrate which has been arranged by dipole forces. In effect, therefore, all the molecules that were not retained by a small excess of attractive forces are vaporized. For the condensation of water molecules on *HCl* it is significant that the recombination at normal pressure takes place in the same molecular ratio 1/8, regardless of whether a concentrated or a dilute acid is subjected to distillation. According to our present knowledge, this constancy of molecular ratios can be explained only by short-range forces pre-orienting the dipoles prior to condensation.

Computations from the relatively small Cl^- values in dew and hoarfrost at the North Sea by Köhler [31], and in Ober Schreiberhau by the author, showed that the hydrochloric acid nuclei—which were still assumed by Köhler to be *NaCl* nuclei—could have originated from a few molecules of *HCl* with eight times as many molecules of water.

Induction Forces. These forces, which are of electrostatic nature, originate when one pole induces a dipole moment in a neighboring pole; this dipole moment points in the direction of the inducing pole so that an attraction results between the inducing and induced molecules. The attraction is proportional to the product of the exciting pole moment (either a permanent dipole or an ionized gas molecule) and the induced dipole moment. This type of attraction, however, is far less significant than the dipole forces of permanent dipoles. These induction forces are probably important in condensation on walls by inducing a mirror dipole and are important for a similar reason in the case of condensation on solid suspended particles. However, for meteorological phenomena this is not too important, since many types of solid particles require a higher water-vapor saturation to produce condensation than is customarily encountered in the atmosphere. Other particles, such as ash dust with its hydrophilic compounds, for example, K_2CO_3 or freshly powdered quartz crystals, which are characterized by a highly active surface, possess readily inducible dipoles.

Dispersion Forces. These forces, which are of electromagnetic plus electrostatic nature, are based on the fact that alternating electric fields are produced as a result of the motion of electrons in an excited state. Under the influence of these fields, dipole moments, whose interaction produces a net attraction, are induced in each of two or more mutually approaching molecules. Their effect is far greater than that of induction forces. The dispersion forces are not a function of temperature; they are effective for relatively large molecules, at rather high temperatures, and also for nonpolar molecules. Their lifetime is extremely short. However, with adequate stimulus (ultraviolet radiation), the same atoms and molecules can be continuously re-excited.

Importance of the Order of Hydration for the Mode of Charging Nuclei. The electrical double-layer, a consequence of the hydration order, was recognized clearly for the first time during the observation of the chemical separation in the spray of salt sols already mentioned [18]. A description of a separation experiment of this type follows.

The mother liquor used for the atomization in this experiment contained *Mg*, *Cl*, and *K* in the following proportions: $Mg/Cl = 1/440$; $Mg/K = 1/4$. The *p*H value was 7.6. Fine droplets (of the magnitude of condensation nuclei) were produced by a jet of compressed air; at various distances from the nozzle they had the mean values given in Table IV.

Table IV. Aerosol Characteristics in a Spray Experiment

Distance of droplets from nozzle (m)	0.30	1.50	3.00
Cl/Mg	83	30	37
*p*H	6.9	6.3	6.3
Concentration*			
Cl	9	2.72	2.72
Mg	48	40	32
K	25	15	10

* In per cent of concentration in mother liquor.

At 1.50 m from the nozzle, after 45 min of atomization, 15 per cent of the total chlorine dissolved in the droplets was still present in the form of condensation nuclei, and 85 per cent in the gaseous phase, probably predominantly as ClO_2, Cl_2, and *HCl*.

The results show a strong escape of the Cl_2 (produced autocatalytically), as well as a separation of the Cl_2 component which has remained dissolved from the nonvolatile alkalis, and a lower *p*H value (more acidic) in the far-drifting, atomized component. This is possible only if there is an excess of cations in the interior of the coarse drops and an excess of *Cl* anions in the outermost molecular layer. In the atomization process, portions of this layer are removed as most finely divided, far-drifting water droplets, while the coarser drops containing the bulk of the mass fall to the ground near the nozzle. Such an arrangement of the chemical substances in the solution can be explained only by hydration, according to Born [3]. That is, neutral water molecules are attracted by the chemical ions; moreover, this

attraction increases as the size of the ion decreases. Water molecules thus attracted form the so-called water coating. Owing to the particular spatial conditions which cause a stronger pull of the large ions toward the interior, the hydrated ions now become arranged toward the center in the order of increasing size of their coating, whereas the "bare" ions (in this case the Cl^-) accumulate on the surface. Jebsen-Marwedel [30] states that for similar phenomena in fluid glass the "bare" ions, that is, those with smaller molecular forces, are squeezed into the surface. According to Benson [2], this surface activity as a preliminary stage of molecular separation is a familiar phenomenon in the case of foam. However, according to Whitney and Grahame [58], it also represents the basic reason for the occurrence of the electrical double-layer. Undoubtedly, this surface activity is also the basis for the electrical double-layer of droplets suspended in the atmosphere.

Ballo-Electricity. The precipitation electricity (Lenard effect, that is, negative charge of the fine particles and positive charge of the somewhat coarser ones) must also be affected by the order of hydration, according to the discussion above, because of its dependence on the double-layer. To prove this, experiments with atomization of drops in thunderstorm updrafts and experiments with falling drops (rain) were undertaken using dilute aqueous solutions of $MgCl_2$ and $[Mg(NH_3)_6]Cl_2$, corresponding to the concentration in raindrops. The atomized droplets were chemically analyzed and tested for charge at various distances up to 3 m from the nozzle. These tests were made in the same way with the falling of the central drops and the small secondary drops [12]. The investigations yielded the following information:

In addition to the appearance of the Lenard effect, a chemical separation occurred in the atomized drops such that the fine particles, which drifted to some distance after being torn from the bulk surface, were enriched with Cl^- ions and showed a decreased pH value (more acid), whereas the coarser particles which settled more rapidly were enriched with cations and showed a higher pH value. In the case of the falling drops the same difference could be found between the coarse main drop and the fine secondary drop which represents the outermost shell of the total drop.

Therefore, the separation of pairs of ions is to be regarded as an almost confirmed cause for the balloelectric charge in the case of the Lenard effect. According to the concepts of Simpson [52] and the presentation of Israël [29], the repeated breakup of the water drops in the cloud by updrafts is, indeed, the fuel of the storm machine. However, the then unknown electrical agent which is effective during this drop disintegration and which is responsible for a continuously renewed occurrence of the Lenard effect, can now be considered as recognized in the order of hydration that repeats itself after each drop breakup. The mechanism of the processes has not yet been explained in all its details, nor is the significance of this effect for thunderstorm formation completely understood.

Electrical Charge by Adhesion. In continuation of the concepts outlined above, this process is also linked to the order of hydration. This process takes place on suspended droplets and hence is not associated with drop disintegration, but is produced by the adhesion of a gas molecule, ionized by cosmic radiation, secondary radiation, etc., or by adhesion of a small ion. The adhesion to droplets is in all probability a selective process. If positive chemical ions are accumulated in the boundary layer of the droplet, negative particles are attracted and positive ones are repelled. If some other type of chemical phenomenon is responsible for the accumulation of negative ions in the boundary layer, the reverse occurs. This concept of the control of the charging process by chemical phenomena leads also to the explanation of the unipolarity of clouds. According to this working hypothesis, droplets of like origin and hence of like chemical composition must assume charges of the same sign. A selective association of small charge carriers was observed by Tyndall [55]. According to him, alcoholic gases capture principally negative carriers and this effect increases with an increase in the length of the carbon chain. If the dipole group OH is removed from such compounds, any effect on carriers ceases. Thus, one can imagine that clouds whose elements are of uniform chemical composition act like a filter on the smallest charge carriers of one sign. If, for example, over extended areas of blossoms (*Prunus*), heated air cools adiabatically as it rises to a greater altitude and forms droplets with an excess of the NH_4^+ cation in the surface layer, these droplets must become negatively charged. Conversely, droplets suspended directly above (as in inversions) with an excess of NO_2^- in the surface layer (condensation on nitrites) would become positively charged. Observations pointing in this direction (specifically, data from chemical experiments) were obtained in the High Tatra (Szepes, Slovakia) during the development of areal thunderstorms [11, 12]. Perhaps such a phenomenon is the explanation for the fact that frequently the upper portions of the thunderheads were positively charged while the bottom portions were negatively charged. During the penetration of droplets into clouds having a different type of chemical make-up, or opposite charge, electrical balancing processes and coagulation result. The consequence of this, in turn, is a different chemical make-up of the new droplets and thus new, rapid, layerwise coagulation. It is possible that the rapid fluctuations of the electrostatic field are correlated with such filter and coagulation processes. If this is true, perhaps these processes also control the stronger or weaker production of O_3 at the tips of grounded conductors as well as in nonhomogeneous fields between clouds. Their possible significance for thunderstorm formation cannot yet be evaluated from the available experimental data.

Ice-Particle Charging. This process was recognized as one of the most important sources of thunderstorm electricity in the informative studies by Findeisen [26] and might, in the final analysis, have no other origin than a chemical separation. Although in the case of ice a hydration equal to that in droplets cannot exist, at least a boundary layer arrangement with a double-

layer may occur. Particularly in the production of graupel, in which the liquid phase of the water customarily occurs in conjunction with the solid phase, such behavior must be expected, as is attested to by the very valuable physical experimental work of Lange and collaborators [33–36] on the Volta potential between the phases. They refer repeatedly in their papers to the significance of the molecular forces and the traces of chemicals in the orientation processes which run parallel to the crystallization and the melting. The orientation processes consist of wandering of ions, formation of double-layers, hydration, and dehydration. Exhaustive analyses of trace elements, especially concerning the chemical difference between broken-off crystal tips and the main mass of the crystal, might lead a step further in this case. In this connection, the information reported by Wall at the Geophysical Institute of the University of Frankfurt a. M. in 1948 is also important. With the aid of rotating mirrors he was able not only to ascertain the dipole character of snow crystals but also to observe, even at a low temperature, a liquid layer in their interior. The latter phenomenon is possible only if the liquid consists of a highly concentrated salt solution whose substances serve as condensation nuclei. An explanation of the physico-chemical phenomena at the interface between ice and mother liquid should, in the opinion of the present author, give more fundamental information concerning the charging of the particles than does the usual assumption of the presence of piezoelectricity.

CONCLUSION

The scope of research in the field of atmospheric chemistry could only be sketched in this highly compressed survey. Several promising lines for research in the immediate future have been suggested, with a brief discussion of modern methods, results, and working hypotheses to serve as a general background. The overall development of a practical "applied atmospheric chemistry" is an undertaking which will require two or three research teams of not less than eight or nine scientists well schooled in analytical and technical methods, equipped with first-class laboratory facilities, and working in collaboration. They could within a reasonable time develop experimental methods of a high degree of accuracy and speed. Simultaneous serial observations of various interrelated chemical substances in the atmosphere are badly needed, but very small groups will not be in a position to undertake the work required. Complementary physical measurements, such as nuclei counts and the measurement of ultraviolet radiation, should be undertaken simultaneously.

It is not unreasonable to expect that in a few years chemical concepts and techniques will not only be available for atmospheric hygiene, but will also be useful in general weather forecasting. The first rapid and usable techniques, important in the field of dynamic meteorology, possibly may not be connected with the determination of O_3 but with determinations of the pH value of aerosols and the reduction power. Such uses would be particularly valuable in that they would provide a rapid and accurate indication of the motion of air masses and equally reliable information about such factors as the degree of turbulence.

REFERENCES

1. Amelung, W., und Landsberg, H., "Kernzählungen in Freiluft und Zimmerluft." *Biokl. Beibl.*, 1:49–53 (1934).
2. Benson, F., und Perrin, J., in Marcelin, A., *Oberflächenlösungen, zweidimensionale Flüssigkeiten und monomolekulare Schichtungen.* Dresden, Steinkopff, 1933. Sonderabdruck aus Kolloid-Beihefte, Bd. 36. (See pp. 280 ff.)
3. Born, M., "Über die Beweglichkeit der elektrolytischen Ionen." *Z. Phys.*, 1:221–249 (1920); *Z. Elektrochem.*, 26:401–403 (1920).
4. Briner, E., et Perrottet, E., "Méthode d'analyse de l'ozone, à l'état très dilué, fondée sur l'action catalytique exercée par ce gaz dans l'oxydation des aldéhydes." *Helv. chim. Acta*, 20:293–298 (1937); "Méthode d'analyse de l'ozone très dilué. II—Détermination de la concentration de l'ozone dans l'air à Genève." *Ibid.*, 20:458–461 (1937).
5. Cauer, H., "Bestimmung des Jodes der Luft." *Z. anal. Chem.*, 104:161–169 (1936).
6. —— "Bestimmung des Gesamtoxydationswertes, des Nitrits, des Ozons und des Gesamtchlorgehaltes roher und vergifteter Luft." *Z. anal. Chem.*, 103:321–334, 385–416 (1935).
7. —— "Entnahmeapparatur für chemisch-klimatologische und technische Luftuntersuchungen." *Z. anal. Chem.*, 103:166–180 (1935).
8. —— "Chemisch-bioklimatische Studien in der Hohen Tatra und ihrem Vorland." *Der Balneologe*, 3:7–23 (1936).
9. —— "Chemisch-bioklimatische Studien im Glatzer Bergland." *Der Balneologe*, 4:545–565 (1937).
10. —— *Schwankungen der Jodmenge der Luft in Mitteleuropa, deren Ursache und deren Bedeutung für den Jodgehalt unserer Nahrung.* Berlin, Verlag Chemie, 1939.
11. —— "Chemie der Atmosphäre," *Naturforschung und Medizin in Deutschland, 1939–1946* (*FIAT Rev.*). Wiesbaden, Dieterich, 1948. (See Vol. 19, p. 277)
12. —— "Ergebnisse chemisch-meteorologischer Forschung." *Arch. Meteor. Geophys. Biokl.*, (B) 1:221–256 (1949).
13. —— "Chemisch-bioklimatische Studien in Koenigstein, Taunus." *Arch. phys.-diätet. Ther.*, Bd. 1, Heft 2, SS. 87–103 (1949).
14. —— "Chemisch-bioklimatische Studien auf der Nordsee-Insel Norderney." (To appear in *Arch. phys.-diätet. Ther.* (1951).)
15. —— und Cauer, G., "Die Bestimmung des Magnesiums in den Nebelkernen und Niederschlägen der freien Atmosphäre." *Z. anal. Chem.*, 124:81–85 (1942).
16. —— "Das Magnesiumchlorid der Nebelkerne." *Der Balneologe*, 9:301–309 (1942).
17. —— "Studien über den Chemismus der Nebelkerne in Ober Schreiberhau." *Der Balneologe*, 8:345–353 (1941).
18. —— "Die Gradierhausluft von Bad Dürrenberg." *Der Balneologe*, 10:1–12 (1943).
19. Curry, M., *Bioklimatik*, 2 Bde., Riederau/Ammersee, 1946.
20. Debye, P., "Molekularkräfte und ihre elektrische Deutung." *Phys. Z.*, 22:302–308 (1921).
21. Dhar, N. R., and Ram, A., "Formaldehyde in Rain Water." *Nature*, 130:313–314 (1932).
22. Dirnagl, C., "An Instrument That Records the Oxidizing Effect of the Air." *Bull. Amer. meteor. Soc.*, 30:214–217 (1949).

23. Ehmert, A., "Über das atmosphärische Ozon." *Ber. dtsch. Wetterd. U. S.-Zone*, Nr. 11 (1949). (See pp. 26–28)
24. Feigl, F., *Qualitative Analyse mit Hilfe von Tüpfelreaktionen*, 2. Aufl. Leipzig, Akad. Verlagsges., 1935.
25. Fellenberg, T. v., *Das Vorkommen, der Kreislauf und der Stoffwechsel des Jods.* (*Ergebn. Physiol.*, Bd. 25.) München, Bergmann, 1926. (See pp. 176–363)
26. Findeisen, W., "Über die Entstehung der Gewitterelektrizität." *Meteor. Z.*, 57:201–215 (1940).
27. Groth, W., und Suess, H., "Bemerkungen zur Photochemie der Erdatmosphäre." *Naturwissenschaften*, 26:77 (1938).
28. Hadamczik, E., *Bericht über Untersuchungen des Formaldehydes der Luft.* Diss., Universität Kiel, 1947.
29. Israël, H., "Die elektrischen Erscheinungen des Gewitters." *Wiss. Abh. D. R. Reich. Wetterd.*, Bd. 8, Nr. 4 (1942).
30. Jebsen-Marwedel, H., "Qualitative Lösungsmechanik." *Angew. Chem.*, Reihe B, 19:7, 186–190 (1947).
31. Köhler, H., "Untersuchungen über die Elemente des Nebels und der Wolken." *Medd. meteor.-hydr. Anst. Uppsala*, Vol. 2, No. 5, 73 pp. (1925); "Studien über Nebelfrost und Schneebildung und über den Chlorgehalt des Nebelfrostes, des Schnees und des Seewassers im Halddegebiet." *Bull. geol. Instn. Univ. Upsala*, 26:279–308 (1937).
32. Kraus, R., "Ein einfaches Gerät zur automatischen quantitativen Schwefelwasserstoffanzeige in der Atmosphäre." *Chem. Fabr.*, 9:241–242 (1936). "Über ein neues Gasspürgerät I. Anzeige von Schwefelwasserstoff." *Z. anal. Chem.*, 112:1–6 (1938).
33. Lämmermann, H., und Lange, E., "Das äussere elektrische Potential von in einem Temperaturgefälle befindlichen Metallen." *Z. phys. Chem.*, 49(B):219–234 (1941).
34. Lange, E., "Voltapotentiale an H_2O-Phasen als Quelle der Gewitterelektrizität." *Meteor. Z.*, 57:429–436 (1940).
35. —— "Meteorologisch interessierende Voltapotentiale an H_2O-Phasen." *Z. Elektrochem.*, 47:867–876 (1941).
36. —— und Weidemann, M., "Elektrostatische Potentiale an elektrochemisch wichtigen Systemen starrer Ladungsschichten." *Z. Elektrochem.*, 47:568–580 (1941).
37. Lenard, P., *Wissenschaftliche Abhandlungen.* Leipzig, S. Hirzel, 1942.
38. Liesegang, W., "Über die Verteilung schwefelhaltiger Abgase in freier Luft." *Gesundheitsing.*, 54:705–709 (1931); "Die Reinhaltung der Luft." *Ergebnisse angewandter physikalische Chemie*, Bd. III. Leipzig, Akad. Verlagsges., 1935.
39. Mayer, H., *Aktuelle Forschungsprobleme aus der Physik dünner Schichten.* München, R. Oldenbourg, 1950. (See p. 277)
40. Német, A., "Untersuchung über Strukturänderung der Kristalle im elektrischen Feld." *Helv. phys. Acta*, 8:97–151 (1935).
41. Nitschke, A., "Eine Micromethode zur Bestimmung der Chloride in Körperflüssigkeiten." *Biochem. Z.*, 159:489–490 (1925).
42. O'Brien, B., "Vertical Distribution of Ozone in the Atmosphere. I—Spectrographic Results of 1934 Flight." *Nat. Geogr. Soc. Contrib. Tech. Papers, Stratosphere Ser.*, No. 2, pp. 49–70, Washington, D. C. (1936).
43. —— Mohler, F. L., and Stewart, H. S., "Vertical Distribution of Ozone in the Atmosphere. II—Spectrographic Results of 1935 Flight." *Nat. Geogr. Soc. Contrib. Tech. Papers, Stratosphere Ser.*, No. 2, pp. 71–93, Washington, D. C. (1936).
44. Quitmann, E., "Über die Bestimmung von Gasspuren in der Luft mit dem Waschrohr nach H. Cauer." *Z. anal. Chem.*, 103:258–261 (1935).
45. —— "Über die Bestimmung von Schwefelwasserstoffspuren in der Luft." *Z. anal. Chem.*, 109:241–246 (1937).
46. —— "Über den Reduktionswert der Luft." *Z. anal. Chem.*, 114:1–8 (1938).
47. —— und Cauer, H., "Verfahren zur chemischen Analyse der Nebelkerne der Luft." *Z. anal. Chem.*, 116:81–91 (1939).
48. Regener, E., "Ozonschicht und atmosphärische Turbulenz." *Meteor. Z.*, 60:253–269 (1943).
49. Regener, V. H., "Messungen des Ozongehaltes der Luft in Bodennähe." *Meteor. Z.*, 55:459–462 (1938).
50. Renger, F., *Tagesgang von Nitrit und Ozon in Ober Schreiberhau.* Diss., Universität Breslau, 1943 (unpublished).
51. Schmauss, A., und Wigand, A., *Die Atmosphäre als Kolloid.* Braunschweig, F. Vieweg & Sohn, 1929.
52. Simpson, G. C., and Scrase, F. J., "The Distribution of Electricity in Thunderclouds." *Proc. roy. Soc.*, (A) 161: 309–352 (1937).
53. Smith, R. A., *Air and Rain.* London, Longmans, 1872.
54. Tichy, H., "Gleichzeitige Messungen von Ultra-violett und bodennahen Ozon." *Der Balneologe*, 6:125–130 (1939).
55. Tyndall, A. M., "Carriers of Electricity in the Atmosphere." *Nature*, 122:16–17 (1928).
56. Volmer, M., und Weber, A., "Keimbildung in übersättigten Gebilden." *Z. phys. Chem.*, 119:277–301 (1926).
57. Wegener, A., "Über die Eisphase des Wasserdampfes in der Atmosphäre." *Meteor. Z.*, 27:451–459 (1910).
58. Whitney, R. B., and Grahame, D. C., "Modified Theory of the Electrical Double Layer." *J. chem. Phys.*, 9:827–828 (1941).
59. Wilson, C. T. R., "Condensation of Water-Vapour in the Presence of Dust-Free Air and Other Gases." *Phil. Trans. roy Soc. London*, (A) 189:265–307 (1897).
60. Wolf, L., *Theoretische Chemie.* Leipzig, J. A. Barth, 1943.

ATMOSPHERIC POLLUTION

Atmospheric Pollution *by E. Wendell Hewson* 1139

ATMOSPHERIC POLLUTION

By E. WENDELL HEWSON

Massachusetts Institute of Technology

INTRODUCTION

Contamination of the atmosphere by human activities has occurred in many countries and from early times [86]; with increasing industrialization such pollution is becoming more widespread and severe. Atmospheric pollution presents a problem the solution of which requires contributions from many fields, such as chemistry [60], city planning [87], legislation [93], public health [46, 93, 99], mechanical engineering [77, 93], meteorology [9, 19, 24, 39, 104], and plant physiology [45, 93, 97, 106]. Thus the analysis, measurement of concentrations, and methods of removal at the source of certain offending contaminants are chemical problems; city planning, by locating industrial plants most advantageously relative to population centers, makes its contribution; legal authorities draw up enforceable codes; industrial hygiene and public health authorities may specify acceptable upper limits to contaminants; mechanical engineers play their part through development of more efficient combustion processes and design of dust collectors; meteorology describes the manner in which contaminants are transported and diffused by the atmosphere and offers methods of alleviation; and plant physiology examines the effects of pollution on vegetation. General discussions of atmospheric pollution are available, either brief [1, 8, 34, 36, 51, 63], or comprehensive [56, 93]. For a very complete bibliography of meteorological literature on atmospheric pollution, see [48].

Before launching into the discussion of the strictly meteorological aspects of atmospheric pollution, it is desirable to survey briefly several closely related topics, namely types of atmospheric contaminants and methods of measuring them.

TYPES OF CONTAMINANTS

Man-made contaminants in the atmosphere may be divided broadly into two categories, particulate matter and gases.

Particulate Matter. The largest of the atmospheric impurities are certain of the dusts, which tend to settle out in a relatively short time. Grinding for size reduction is a prolific source. Dusts are also frequently formed by industrial combustion processes and ejected into the atmosphere by flue gases of sufficiently high velocity to carry heavier particles up the stacks; such high stack velocities are characteristic of industrial furnaces. These impurities consist of coal and coke dusts and fly ash. The smaller particles tend to remain suspended and are frequently termed "aerosols." Fumes, formed by volatilization and condensation of solids, and smokes are typical; many of these smaller particles are composed of tar and other combustible matter; some of them are hygroscopic and act as condensation nuclei [50]. Finally, there are the mists, consisting of liquid droplets which are often of a permanent nature. For detailed discussions, see [24, 36, 56, 98, 102].

Gases and Vapors. The presence of such contaminants, although invisible, may be revealed by their smell, or by a stinging or burning, as of the eyes, or by their corrosive effect on materials or damage to vegetation, or, in special cases, by the radiation emitted by them. Sulfur dioxide is one of the most prevalent of such gases. Acid vapors have a corrosive action and certain organic compounds in extremely small concentrations are malodorous [36, 80, 81, 98]. Radioactive waste gases from nuclear reactors present special features [8, 76]. Radiation hazards from such gases decrease with time because of their decay, argon-41, for example, having a half-life of 110 minutes.

MEASUREMENT OF ATMOSPHERIC POLLUTION

Various methods of measuring pollution have been devised: some of these methods, along with an extensive bibliography of 120 references, are given in [13]; the basic considerations which should be kept in mind when designing instruments are described in [35]. The more widely used methods of measuring pollution are described briefly below; fuller treatments are available elsewhere [24, 93], and suggested improvements in instrumentation have also been outlined [24, pp. 132–136].

The Deposit Gauge. This instrument, designed for the purpose of collecting the total amount of material deposited on a given area during a given time, is one of the first widely used for the measurement of atmospheric pollution [66]. One form is similar to a rain gauge, and consists essentially of a funnel with a collecting bottle below; the impurities and rainwater collected are analyzed by chemical methods. Its limitations are also similar to those of the rain gauge [61], since its "catch" is very sensitive to exposure, and depends not only on pollution present but also on rainfall, wind direction, and turbulence [4; 25, pp. 18–21; 59]. A detailed description of the significance of deposit-gauge measurements has been given [25, pp. 8–11]. Other more refined deposit gauges have been designed, but they all suffer from the above-mentioned limitation.

The Smoke Filter. Samples of the solid matter suspended in the atmosphere may be obtained by drawing a measured volume of air through a filter paper by means of a pump. The solids are filtered out, leaving a gray stain on the paper; their amount is determined by visual comparison with a standard scale of shades or by photometric methods. An automatic filter, which has been widely used, was designed by Owens

in 1918 [26]; a modified design more suitable for the North American climate has been produced [18]. A variety of more elaborate instruments for the study of particulate matter have recently been used [98]; these include the condensation-nuclei counter [94], the midget impinger [15], the cascade impactor and the electron microscope [57].

The Ringelmann Chart. In the United States the emission of solids from stacks has been estimated by the Ringelmann chart. The observer compares the shade of grayness of the smoke with a series of shade diagrams consisting of black horizontal and vertical lines on a white ground; the thickness of these lines increases progressively through the series. The method is highly subjective and open to criticism, the results obtained being influenced by such factors as background, illumination, etc. [33, 55].

Sulfur Dioxide Apparatus. Two methods are widely used. In one, air is drawn through a bubbler containing dilute hydrogen peroxide. Sulfur dioxide in the air combines with the hydrogen peroxide to form sulfuric acid. The acidity of the solution is determined either by titration [24] or by measuring its electrical conductivity [89, 90, 91]; the concentration of the sulfur dioxide is then computed from the measured acidity of the solution. A recording instrument utilizing the titration principle provides a measure of materials, such as sulfur dioxide, that exert a reducing action on an acid solution of bromine [96, 98].

The other method, developed by Wilsdon and McConnell [105], depends on the corrosive action of sulfur dioxide. A prepared surface of lead peroxide is exposed to the atmosphere for a specified period of time, usually a month; sulfur dioxide reacts with the peroxide to form lead sulfate. The yield of sulfate is proportional to the mean concentration of sulfur dioxide. Of the various meteorological elements, only rainfall influences the reaction rate [24].

The Ultraviolet Daylight Integrator. The amount of ultraviolet light from the sun cut off is a measure of pollution by particulate matter and is significant for the health of human and plant life. A number of methods of measuring ultraviolet radiation received at the earth's surface have been devised [56], the best of which consists of two coaxial translucent fused silica bulbs, one inside the other, and with necks downward. With this arrangement, a nearly constant intensity of diffuse light passes down the neck of the inner bulb for any position of the light source over a hemisphere; a silver filter transmits only ultraviolet radiation; the intensity of the ultraviolet light is varied by an optical wedge; and finally, the light falls on suitable photographic paper [24].

Visibility. Visibility data permit a determination of pollution by solid and liquid particles [27]. Objective methods of measuring the attenuation of light by particulate contaminants in the atmosphere have been devised [29, 78, 79, 98], and laboratory studies have been made. In one important laboratory study [10] two types of suspensoids were used, one consisting of transparent liquid drops produced by burning phosphorus, and the other of black opaque carbon particles produced by burning camphor. For the carbon particles, the glare and diffusion were negligible, and the logarithm of the transmission was proportional to the total amount of carbon. For a water cloud there was considerable scattering of light and the transmission did not fall off with concentration nearly so rapidly as for carbon. Mixed clouds showed intermediate properties. In another laboratory study [98], measurements were made of the opacity of pollutants such as sulfur trioxide which combine with water to form sulfuric acid mists. Studies of the interrelationship of visibility and various parameters have been made; in one, it was deduced that the product of the number of particles per unit volume, the relative humidity, and the visibility is a constant [47]; in another, that visibility is inversely proportional to smoke concentration [24, pp. 118–119]. A start has been made in the application of theory to the problem [10, 30, 78], but there are uncertainties involved. For example, since some of the particles are hygroscopic, their size at higher relative humidities may increase by condensation, thus decreasing the visibility even though there may have been no increase in pollution [62]. The contribution of visibility studies has to date been strictly limited.

POLLUTION BY SINGLE SOURCES

Atmospheric pollution by a single source is frequently of importance, as in the case of an industrial plant emitting contaminants from its stack. A number of theoretical studies of this problem have been made.

Basic Equations. One of the first theoretical treatments was that by Roberts [70] in 1923. He developed an equation for the concentrations downwind from a continuous point source using as a parameter the coefficient of eddy diffusivity. Subsequent studies [69] have indicated the inadequacy of an approach based on the assumption that the coefficient is constant and suggest that other parameters are required.

In 1932, Sutton [82] developed an equation for a point source at the earth's surface by assigning an expression, deduced from dimensional considerations, to the correlation coefficient introduced by Taylor in his theory of diffusion by continuous movements.

Using statistical concepts fundamentally very similar to those of Taylor and Sutton, Bosanquet and Pearson [12] and Bosanquet [11] developed a number of useful expressions for the pollution from a point source. One such expression gives the concentration which is effective in cumulative processes, such as the blackening of a neighborhood by soot and the attack of structures by acid constituents of the stack effluents. Thus, if the fraction of the year during which the wind direction falls within an arc θ is $a\theta$, then the average value of the concentration for the whole year is given by

$$\bar{\chi} = \frac{Qa}{pux^2} e^{-h/px}, \tag{1}$$

where Q is the mass of contaminant emitted per unit time, u is the wind speed, x is the distance downwind from the source, h is the height of the stack, and p is a

numerical parameter with an average value of 0.05 and a range from a minimum of 0.02 for low turbulence to a maximum of 0.15 for very turbulent air flow. Bosanquet and Pearson also give a formula for the ground-level concentration χ_0 during a brief interval of time due to a continuous point source, in the form

$$\chi_0 = \frac{Q}{\sqrt{2\pi}pqux^2} \exp\left(-\frac{h}{px} - \frac{y^2}{2q^2x^2}\right), \quad (2)$$

where y is the distance crosswind from the axis of the smoke cloud, q is a second numerical parameter with an average value of 0.08, and the other symbols are as defined for equation (1).

In 1947, Sutton [84] extended his treatment to cover elevated point sources and assigned values to the various parameters involved. The general expression derived is

$$\chi = \frac{Q \exp\left(-y^2/C_y^2x^{2-n}\right)}{\pi C_y C_z u x^{2-n}} \cdot [\exp\{-(z-h)^2/C_z^2x^{2-n}\} + \exp\{-(z+h)^2/C_z^2x^{2-n}\}], \quad (3)$$

where z is distance upward; C_y and C_z are virtual diffusion coefficients for the crosswind and vertical directions, respectively; n is a numerical parameter whose value, lying between 0 and 1, is related to the diffusing power of the turbulence; and the other symbols are as given above. With $z = 0$, (3) gives the concentration at the surface as

$$\chi_0 = \frac{2Q}{\pi C_y C_z u x^{2-n}} \exp\left\{-\frac{1}{x^{2-n}}\left(\frac{y^2}{C_y^2} + \frac{h^2}{C_z^2}\right)\right\} \quad (4)$$

and its maximum as

$$\chi_m = \frac{2Q}{e\pi uh^2}\left(\frac{C_z}{C_y}\right), \quad (5)$$

which occurs at a distance x_m from the source, given by

$$x_m = (h^2/C_z^2)^{1/(2-n)}. \quad (6)$$

The parameter n, obtained by fitting the theoretical profile $u = u_1 z^{n/(2-n)}$ to the observed change of wind speed with height, has the value ¼ under average conditions of lapse rate, with a range from ½ for a large inversion to ⅕ for a large lapse. For average conditions, Sutton gives values of C_y varying from 0.21 for a source at the surface to 0.07 for one at 100 m and values of C_z from 0.12 to 0.07 for the same range of heights of sources; for sources at 25 m and higher the values of C_y and C_z are the same. For inversions the coefficients are smaller; for large lapse rates they are greater. The values of the numerical parameters n, C_y, and C_z given above have been obtained by Sutton from basic meteorological data, but their validity has been checked through measurement of concentrations from surface sources only. Their use does not give instantaneous values of concentrations, but rather the average at a fixed point for a period of not less than three minutes; Sutton states that in the absence of systematic changes in wind direction near the surface the time-mean values for longer intervals are not significantly different. The validity of the numerical values of the parameters for elevated sources has not been established. Figure 1 gives maximum concentrations at the ground and the distance from the stack at which such maxima occur as functions of stack height and of groups of vertical temperature gradients; the curves are derived from equations (5) and (6) above.

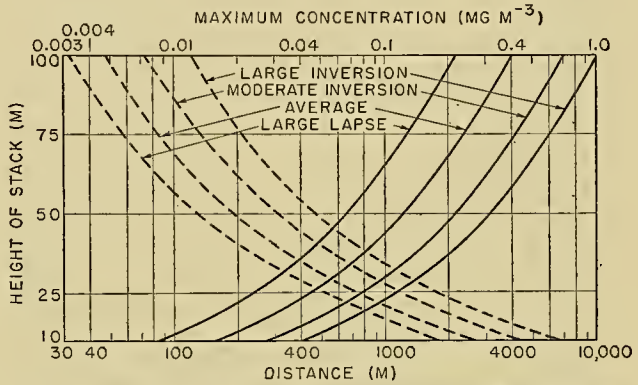


FIG. 1.—Maximum ground concentrations (dashed curves) and distances from stack base (solid curves) at which they occur with various vertical temperature distributions for stacks of different heights, according to Sutton. Rate of emission: 1 g sec^{-1}.

Lowry [52] has given an equation for maximum surface concentrations which embodies features of both (1) and (5). It has the form

$$\bar{\chi}_m = \frac{2Q}{e\pi h^2}\left(\frac{a_m}{u}\right), \quad (7)$$

where $\bar{\chi}_m$ is the time-mean concentration at a fixed point for a period of one hour; a_m is the frequency during the hour of the most frequent wind direction at the top of the stack; and the other symbols are as defined previously. The location of the maximum ground concentration is also specified: its direction from the stack is given by the most frequent wind direction at the top of the stack during the hour; its distance is given by the empirical expression

$$x_m = h \csc \sigma, \quad (8)$$

where σ is the standard deviation of the wind direction in degrees during a period of 10 or 15 minutes. Equation (8) applies only for sources at a height of the order of 100 m and when the range of direction is greater than 10 or 15 deg. The general validity of equations (7) and (8) has been verified by measurements of surface concentrations of oil-fog emitted from the 108-m experimental stack at Brookhaven National Laboratory. For further details and a suggested use of these two equations in climatological planning, see page 1152.

An equation for concentrations under very stable conditions has been given by Barad [5]. His approach derives generically from Roberts' work [70], but incorporates certain modifications. Barad assumes that in the horizontal layer of air near an elevated source, the variation with height of the eddy coefficient for vertical transfer is sufficiently small in very stable

conditions to warrant treating the coefficient as a constant with a value equal to the mean for the layer; similarly, a mean value for the eddy coefficient for horizontal transfer is also taken. Instead of assuming, as in previous treatments, a point source of infinite concentration, an area source of finite concentration is postulated. The source is taken to be an elliptical area whose major axis lies horizontally crosswind and whose minor axis is vertical; the area is located in the vertical plane at which the effluent flow first becomes horizontal after leaving the top of the stack. With these assumptions and specified boundary conditions the fundamental differential equation is solved. Computed concentrations are substantially smaller near the source than those obtained from Roberts' equation, but the difference becomes progressively less with increasing distance from the source. As the size of the vertical areal source increases, appreciable differences occur for greater distances downwind. The theory has not yet been checked by measurements of concentrations under inversion conditions. The emphasis on areal rather than on point sources is valuable, and further investigations of diffusion from such sources under various atmospheric conditions may well lead to significant advances of a general nature. Barad's analysis suggests that studies of the effluent from industrial stacks of large diameter, considered as vertical areal sources, should be made under inversion conditions.

Relatively few measured values of concentrations from individual elevated sources are available to permit a comparison with theory; those taken in the course of the Trail investigation [19] are of limited applicability because they were taken in the Columbia River valley; those taken at Brookhaven National Laboratory [53] have not been generally available long enough to permit a full analysis of them; the publication of other analyses [16, 92] is in such a form that it is difficult to apply the results in an evaluation of the several equations which have been proposed, as given above. It is thus possible at the present time to make only a few tentative remarks concerning the relative merits of the various approaches.

A comparison of the equation of Bosanquet and Pearson, (2), with that of Sutton, (4), brings out the fact that the expressions are similar in form but different in detail. The numerical parameters p and q are closely related to Sutton's C_z and C_y, a point emphasized by the fact that under average conditions $p = 0.05$ and $q = 0.08$ whereas $C_y = C_z = 0.07$ for a source at a height of 100 m. On the other hand, Sutton's equations involve an additional parameter n which gives them a greater flexibility for portraying faithfully the diffusion in complex meteorological conditions. Sutton [84] points out that his treatment supports a number of the main conclusions of Bosanquet and Pearson.

It is stated by Lowry [52] that, in the light of measured concentrations at Brookhaven National Laboratory, Sutton's equation (5) gives concentrations which are too high if a time-mean of an hour is used. It appears that the measured concentrations may be less than one-tenth of the values predicted from (5) using Sutton's values of the parameters, although generally the discrepancy is less marked. If we take (5) as applying for a time-mean of an hour, then a comparison of (5) and (7) indicates that $a_m = C_z/C_y$. Sutton assumes that, at heights of 25 m and above, turbulence is isotropic, in which case it should follow that $a_m = 1$, but Lowry finds that $a_m < 1$. Thus it appears that when time-means of an hour are taken, rather than those of several minutes, as considered by Sutton, the effective turbulence is not isotropic but $C_y > C_z$. It has been pointed out by Barad [5] that under very stable conditions at Brookhaven the plume of oil-fog travels nearly horizontally for miles with very little increase in vertical thickness but with a gradual lateral widening and a meandering as of a river. It is not yet clear whether the predominance of lateral spreading of the smoke is primarily a result of greater horizontal than vertical turbulent mixing or a result of the variation of the mean wind direction with height, a factor mentioned by Church [16]. The fact that the rate of variation of mean wind direction with height in the surface layers is a maximum in slowly moving and very stable air suggests that this factor should be taken into account under such atmospheric conditions. Further investigation on this point is required. The observed meandering of the smoke plume in an approximately horizontal plane recalls to mind the work of Parr [67], who suggested that the intensity of vertical mixing decreases and that of lateral mixing by large eddies increases as the vertical stability increases. When such meandering is marked, it is clear that the time-mean concentration for a period of the order of an hour at a fixed point relative to the earth will in general be substantially smaller than that for a period of a few minutes or the instantaneous maximum value. The work of Lowry and Barad at Brookhaven indicates that the magnitude of time-mean concentrations is a function of the length of the sampling period, and that such time-mean concentrations have limited significance unless the periods are specified.

It is not yet possible to compare the analysis of Barad [5] with other treatments, because observational data are lacking. However, it may be that an adequate analysis of the behavior of smoke from an elevated source in very stable air will require an assessment of the influence of the variation of mean wind direction with height, mentioned above.

To summarize briefly, each of the proposed equations must be used with caution, and applied only when certain conditions are met. When, under average conditions, a general estimate of the mean concentration for several minutes is sufficient, either Bosanquet and Pearson's equation (2) or Sutton's expressions will be adequate. On the other hand, if mean concentrations for a similar period under other meteorological conditions are required, where quantities such as the vertical temperature gradient differ markedly from the average, then Sutton's equations are to be preferred. When mean concentrations at a point for longer periods are required, Lowry's expression will be more reliable. With strong inversions, Barad's analysis merits atten-

tion. It must be emphasized, however, that although the above analyses are the best available, they still await experimental verification. Only Lowry's treatment has any direct experimental backing, and further verification of it is needed. Many more measurements of concentrations and of the associated meteorological variables are required in order to check the forms of the expressions and, more particularly, to evaluate the various parameters under a wide range of conditions. Furthermore, under certain atmospheric conditions described in the next two sections, the expressions given above break down completely.

Concentrations Near the Source. Whether or not significant concentrations of suspended impurities occur at the surface near a stack depends largely on the stability of the atmosphere. Etkes and Brooks [21] have described the behavior of smoke from a stack during various conditions of stability and wind speed. Their analysis indicates that smoke comes to the surface near a stack only with light winds and a superadiabatic lapse rate. Such conditions occur often on clear summer afternoons or occasionally after the passage of a cold front, especially during the early autumn over continental areas. This behavior of the smoke near the source is confirmed in a study by Church [16], who classifies it as "looping," a descriptive term for the oscillating motion of the narrow plume of smoke. He finds that the large convective eddies which produce looping occur only with winds of 20 mph or less. On the average, the plume first reaches the ground 80 ft away from a 200-ft stack with a wind speed of 1 mph; the distance is proportionately greater for higher speeds up to 20 mph. For a 200-ft stack, Church gives the following tentative average values of the dilution at the surface during unstable conditions: 2000 at 450 ft; 4000 at 900 ft; and more than 10,000 beyond 1700 ft. The dilution is defined as the number of volumes of air with which a unit volume of stack effluent has been mixed; a dilution figure of 500, for example, signifies a concentration $\frac{1}{500}$ of that in the stack. Lowry [52] gives three stack heights as the approximate distance at which maximum concentrations occur in unstable air and with light winds. Bosanquet and Pearson [12] point out that under such conditions mean surface concentrations for periods from a few minutes to an hour are given by equation (1) if a suitable value of a is used.

Large particles, such as fly ash, will be deposited near the source irrespective of weather conditions; the lighter the wind the nearer to the source will the deposit occur.

Some general conclusions concerning the occurrence of high concentrations at the surface near a stack may be reached by inference. The diurnal variation of such concentrations will show a maximum frequency of occurrence during the early afternoon and a minimum during the night and early morning. The annual variation will consist of a maximum during the summer and a minimum during the winter and early spring.

A considerable amount of atmospheric pollution comes not from stacks but from sources at or near the earth's surface as a result of various operations near industrial plants. The most reliable data available are measured concentrations downwind from point and line sources taken during average meteorological conditions on downland on Salisbury Plain, England [83]. Sutton's theoretical expressions [83] for such concentrations are in satisfactory agreement with these data, but measured values neither of concentrations nor of Sutton's parameters are available for extreme conditions, such as large temperature lapses or inversions, or for various types of terrain. The solution of a number of pollution problems thus awaits further measurements of concentrations from known surface sources and of the various meteorological parameters required for an evaluation of theoretical approaches. Such measurements would be most valuable if made at various locations and heights among and around typical industrial plants under both average and extreme meteorological conditions.

Concentrations Distant from the Source. The general picture of concentrations distant from a stack is given by the theoretical and semitheoretical analyses of Bosanquet and Pearson, Sutton, Lowry, and Barad. Church's empirical analysis [16] also provides information of interest, especially concerning the physical behavior of the smoke under various conditions and the degree of its dilution as related to wind speed and stability. The latter for a 200-ft stack is portrayed in Fig. 2; the degree of stability is expressed by $1/\theta \times d\theta/dz$. It will be noted from the figure that the smallest dilutions (greatest concentrations) at the ground occur with light winds and unstable air, as indicated in the previous section. The least dilutions increase with the wind speed for a specified degree of stability. For a given wind speed, the least dilutions diminish with decreasing stability until a minimum is reached and then increase; the minima occur at progressively greater values of the stability as the wind speed increases.

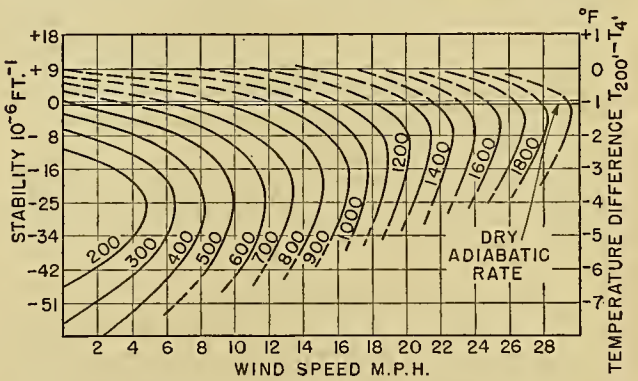


Fig. 2.—Least dilutions (maximum concentrations) at the ground as related to stability and wind speed, according to Church. A dilution of 800 denotes a concentration which is $\frac{1}{800}$ of that in the stack. Stack height: 200 ft. (*Reproduced by permission of the American Chemical Society.*)

When considering the range of usefulness of these approaches, the assumptions on which they are based should be kept in mind. These approaches provide a picture of the diffusion of smoke or gas from a stack under any one of a given set of conditions, with the

implicit assumption that the controlling meteorological conditions do not change significantly during the period under consideration; these treatments may be termed then, in a certain sense, static. But the atmosphere is a dynamic entity whose characteristics may and often do change radically within short periods of time. These changes sometimes lead to surface concentrations which are not predicted by current theories. Several examples may be given. Dean, Swain, Hewson, and Gill [19] describe concentrations of sulfur dioxide found in the Columbia River valley in the vicinity of the smelter at Trail, British Columbia. During the summer the highest surface concentrations occur with considerable regularity at about 8 A.M., the surprising feature being the fact that the onset and progressive development of these fumigations occur practically simultaneously at stations as far as 35 miles downvalley from the smelter. Bosanquet and Pearson's discussion was available at the time of the investigation, but it provided no assistance in this case. The probable explanation was given by Hewson [38]. During the early morning hours the gas flows downvalley in a thin layer which does not reach the valley floor because turbulence is inhibited

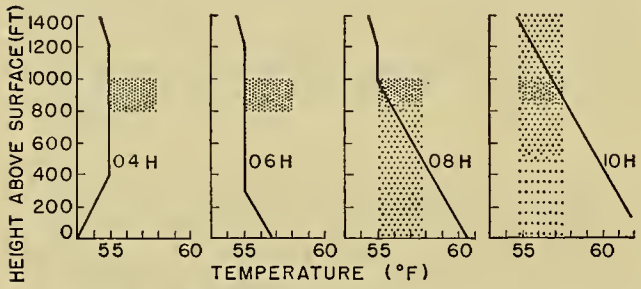


FIG. 3.—Various stages in the process by which contaminants from an elevated source reach the ground in high concentrations during a summer morning, according to Hewson [38].

by the stability of the air. After sunrise, solar radiation heats the surface which in turn heats a layer of surface air, producing in it a superadiabatic lapse rate and marked turbulence. The thickness of this turbulent layer increases with time; when its upper boundary reaches the layer of highly concentrated gas aloft, the gas diffuses rapidly downward producing sudden, high, and nearly simultaneous fumigations along the valley floor. As the upper boundary of the turbulent layer continues to rise, upward diffusion of the gas proceeds, leading to exponentially decreasing surface concentrations thereafter. The several phases involved are illustrated in Fig. 3. From the theoretical standpoint a simplified two-stage mechanism may be visualized. Stage one: concentrations aloft during the early morning are those appropriate for a continuous point source (the top of the stack) and a highly stable atmosphere. Stage two: subsequently the plume of smoke or gas acts as an instantaneous line or narrow area source in a highly turbulent atmosphere, with diffusion occurring initially downward and sideways but not upward; since the instantaneous line source moves with the wind, there is no net component of air flow normal to the line source. Investigations near other smelters on level terrain have shown that such simultaneous fumigations during the morning in summer are of frequent occurrence; the effects noted above are therefore not confined to valleys. According to Lowry [52], experiments at Brookhaven show that the average surface concentration during the fifteen-minute period with peak values is twenty times the maximum specified by Sutton's equations. The findings at Trail, that the greatest concentrations at the surface during the summer occur several hours after sunrise, are thus confirmed. The fact that high concentrations of gas or smoke may come to the surface in this manner miles from the stack must be borne in mind when peak values likely to be reached are of interest.

The effect of an isothermal or inversion layer a short distance above the smoke plume is a second factor which is not allowed for by the general theory. On June 3, 1939, a quasi-stationary frontal surface was located in the Columbia River valley near Trail [19, 38]. As shown by airplane soundings, there was a frontal isothermal layer at a level below the top of the valley sides but well above the top of the stacks, with a large lapse rate below the isothermal layer. Measured turbulence near the surface was large, but unusually and unexpectedly high surface concentrations of sulfur dioxide were measured at stations downvalley from the smelter. The low level of turbulence in the isothermal layer aloft prevented significant upward diffusion through it from occurring, and as a result the gas flowed downvalley as in a giant pipe. Over level terrain the result would not be so serious, since the gas would be free to diffuse laterally. However, there is little doubt that the presence of frontal or subsidence inversions and isothermal layers just above the smoke stream causes pollution conditions of serious concern which are not considered by present theories.

It is clear that there is a great need for both theoretical and experimental investigations of such special conditions which often lead to the most serious instances of atmospheric pollution.

The Influence of Topography. If the terrain in the vicinity of a source of pollution is not essentially level, the difficulty of specifying what the concentration of a contaminant will be at a given point under various meteorological conditions is greatly increased. The problems encountered when the source is in a valley have been fully described [19], and are twofold. When the winds are moderate or strong, local eddies caused by the configuration of the land produce a distribution of smoke which is well nigh impossible to predict from theoretical considerations. An evaluation of this factor requires a detailed investigation, the results of which can rarely be extrapolated to apply to other sites. Secondly, when the winds are light, local winds such as mountain and valley winds predominate, and even the prediction as to whether the wind will be up- or downvalley presents a major problem. The problem is further complicated if a number of side valleys and ravines lead into the main valley. The same factors, but to a

lesser degree, provide complications over any area of irregular terrain. Near a shore line, land and sea breezes occur and must be taken into account when considering the probable areal distribution of airborne pollution. The effect of topography will be considered further in the next section.

The Deposition of Particulate Matter. Calculations of the rate of deposition on the ground of particulate matter from a continuous point source have been presented by Baron, Gerhard, and Johnstone [6]. The rate of deposition is given by the product of the concentration of particles adjacent to the laminar layer and their free settling velocity through this layer, that is, by $\chi_0 w_s$. The quantity χ_0 is given by appropriate forms of either Bosanquet and Pearson's or Sutton's equations, and w_s by Stokes' law for small spheres with a specific gravity of 1.0, in air at 70F. An expression for the fraction ϕ of the total cloud which is deposited per unit time per unit distance downwind is obtained by integrating the product in the crosswind direction from $-\infty$ to $+\infty$. Thus, for example, for Bosanquet and Pearson's equation, the authors obtain

$$\phi = \frac{w_s(1.78h/p)^{w_s/up}}{pux^{(1+w_s/up)}} e^{-h/px}. \tag{9}$$

For Sutton's equations a more involved analysis is used. The values computed with Sutton's equation agree well with those for Bosanquet and Pearson's equation for lapse conditions only, since the latter use an average value for the vertical diffusion coefficient. The essential results may be summarized as follows: a decrease in turbulence increases the maximum rate of deposition and shifts the point at which it occurs downwind from the source; an increase in stack height decreases the rate of deposition and shifts the point of maximum deposition downwind—the maximum rate of deposition is approximately inversely proportional to the stack height. A similar analysis for continuous line sources is available [44].

The Coagulation of Particulate Matter. If smoke particles coagulate at all rapidly in the atmosphere, the effect will be important in increasing the rate of deposition as the particles grow in size. It is known that coagulation by molecular movements is small [101]. The coagulation of smoke particles in the surface layers of the atmosphere has been studied both experimentally and theoretically by Teverovsky [88]. By means of ultramicroscopes the change in the size of smoke particles in a plume was investigated: the number of particles and the mass concentration of the smoke were measured at several points along the plume and the average radius of the particles was calculated from the values obtained. It was found that with 10^4 to 10^6 particles per cubic centimeter the average radius increased from 0.2 μ to 0.4 μ while the smoke traveled a distance of 1 km. This coagulation is attributed to the action of microeddies which causes neighboring particles to converge. Teverovsky gives the orders of magnitude of various diffusion coefficients as follows: the coefficient for molecular movements is 10^{-6} cm^2 sec^{-1}; the coefficient for the microeddies to which coagulation is ascribed is 10^{-3} cm^2 sec^{-1}; and the coefficient for large eddies which are negligible as coagulating agents is 10^4 cm^2 sec^{-1}. The application of theories based both on dimensional analysis and on considerations of energy dissipation leads to results in good agreement with the observations made. These conclusions suggest that the possible effect of turbulent coagulation on the accuracy of photoelectric measurements of smoke concentrations should be examined.

Wind-Tunnel Investigations. Two types of wind-tunnel studies have been made, those to determine the effect of eddies induced by the stacks themselves and by nearby buildings [40, 58, 73, 74], and those to determine the distribution of a contaminant in undisturbed flow [58]. Because the effect of large-scale semipermanent eddies induced by buildings may be much more significant near the source than that of the prevailing field of turbulence in the atmosphere, experiments of the former type can be counted on to give information that is, in the main, reliable. Wind-tunnel studies of undisturbed flow are open to more serious question, however; in these it is assumed that the effective diffusing agency is the turbulence induced by the jet itself, and no allowance is made for the variations of the natural turbulence of the ambient atmosphere, which may be considerable for a given wind speed. Even near the stack the distribution of a contaminant will be significantly affected by the degree of natural turbulence present. The main objection to studies in present-day tunnels is that the degree of turbulence cannot be varied so as to simulate the range of natural turbulence in the atmosphere which is associated with a wide range of stability conditions.

It is possible that various actual stability conditions could be simulated by using a wind tunnel in which the lower surface of the tunnel could be heated and the upper surface cooled and vice versa. Because of the bounding surfaces above and below and the small vertical extent of the air flowing in the tunnel, it is probable that a lapse rate such as the dry adiabatic would not have the basic significance that it exhibits in the atmosphere. However, a lapse rate in the tunnel such that the air density increases with height, that is, a lapse rate greater than 0.034C m^{-1}, might simulate a superadiabatic lapse rate in the atmosphere and produce marked thermal turbulence in the tunnel. Similarly, by cooling the bottom of the tunnel and heating its top, strong inversions could be produced which would reduce turbulence to a degree characteristic of lesser inversions in the atmosphere. The technical difficulties in such a study are considerable, but do not appear to be insuperable. Thus, inversion conditions have been obtained in the Göttingen "hot-cold" tunnel by heating the roof of the tunnel by steam and cooling the lower surface by running water. Detailed comparisons of concentrations from a given stack under various atmospheric conditions with those obtained with exact replicas of stack and surrounding terrain in such a wind tunnel should serve to put model studies on a firm basis. An excellent discussion of wind-tunnel

techniques as applied to meteorological problems, including those posed by atmospheric pollution, is available.[1]

POLLUTION BY MULTIPLE SOURCES

The problems of pollution by multiple sources include those of single sources and a large number of others as well. The interplay of meteorological and other factors is extremely complex in the most important instance, that of a city, and only a start has been made in the study of such factors and their various roles. The main conclusions of the few studies of city pollution which have been made are set forth below.

Effect of Wind. The role of the main air currents of the atmosphere's circulations in transporting pollution is an obvious one. The effects of large-scale features of the general circulation and of local winds on pollution in the Los Angeles area have been described in detail by Beer and Leopold [7]. Of the latter type, both land and sea breezes and mountain and valley winds play a part in alleviating or contributing to the nuisance in the Los Angeles region. More detailed studies of the effects of wind on city pollution have been made, such as that at Leicester, England [24]. There the average effect of wind in shifting the center of surface pollution was surprisingly small. Pollution was naturally greater downwind from the center of the city, but the points of maximum surface concentration of smoke and sulfur dioxide never moved more than half a mile from the city's center as a result of wind; as the wind speed increased, the nearly circular contour lines of equal average pollution at the surface moved downwind by distances up to one mile, without any change in their radii. These results suggest that the upward diffusion of pollution by turbulent mixing is a major factor in limiting surface pollution, a conclusion confirmed by Fig. 4, which shows the effect of winds of various speeds on ultraviolet daylight in Leicester during the winter. The figures in the circles give the loss of ultraviolet light, expressed in logarithmic units, as a percentage of that received in the country districts around. Since the loss of ultraviolet light will be an approximate measure of the smoke above that portion of the city, the diagram indicates that the maximum of total smoke in winter may be found as far as two miles downwind from the city's center. The surface maximum was never more than half a mile downwind. A comparison between daily mean smoke and daily mean wind speed at Leicester was also made. In general the pollution decreases with increasing wind speed, the decrease being progressively more marked with increasing stability, and more pronounced in winter than in summer.

Less elaborate studies of city pollution have been made, such as that of Davidson [17] for New York City. The main meteorological conclusion of this investigation is that dust concentrations at the surface vary inversely as the square root of the wind speed. Other studies suggest that surface smoke concentration varies inversely as the wind speed [31, 47].

1. Consult "Model Techniques in Meteorological Research" by H. Rouse, pp. 1249–1254 in this Compendium.

Effect of Atmospheric Stability. The immediate effect of stability on surface city pollution depends on whether the stable layer is at the surface or aloft. At Leicester the connection with surface stability was studied. It was found that daily mean smoke increases with stability, the increase being small with winds of 10 mph but becoming progressively more pronounced with decreasing winds. The increase was more marked in winter than in summer. There is evidence that diurnal variations of stability, and hence of turbulence, influence the pollution from multiple sources. An investigation of pollution in fourteen of the largest cities in the United States [42] showed a low level of contamination in the early morning, followed by a peak in the forenoon, which occurs earlier in summer than in winter, at a

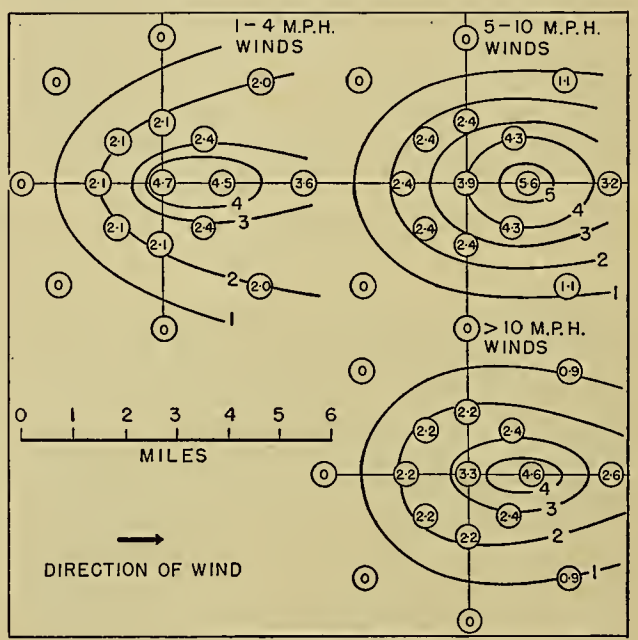


FIG. 4.—The effect of wind speed on ultraviolet daylight losses during winter weekdays in Leicester. Logarithmic units: one unit represents a loss, due to smoke, of 9.4 per cent of the possible ultraviolet daylight. (*Reproduced by permission of H. M. Stationery Office.*)

time ranging from about 6:30 to 8 A.M.; this maximum is followed by a rapid fall to a minimum about 2 P.M. when there is another rise to a peak about 8 P.M. The afternoon minimum is as low or lower than the early morning one. Similar results have been found in studies of pollution in British cities. These observations suggest strongly that the pollution which causes the morning maximum is carried downward from aloft by increased turbulence associated with increased lapse rates caused by solar heating, in much the same manner as described in the preceding section for elevated single sources [38]. The seasonal shift in the time of the morning maximum is clearly related to the seasonal variations in the solar heating.

The stability of air aloft has a considerable influence on summer pollution in the Los Angeles area, as pointed out by Beer and Leopold [7], and by Magill [54]. During that season the air at higher levels is relatively warm as

a result of subsidence occurring over the eastern portion of the Pacific anticyclone. Near the surface the air is relatively cool because of its recent oceanic trajectory, and mechanical turbulence has produced in it a marked lapse rate. Between these two bodies of air an inversion is usually present, extending perhaps from 2000 to 4000 ft. Thus, although there may be turbulence and convection near the surface, the inversion aloft inhibits the upward diffusion of contaminants by eddy processes, and extensive pollution develops.

Effect of Rain. It is not yet clear whether rain tends to alleviate surface pollution by washing impurities out of the air. At Leicester it was found that from 27 to 37 per cent of dissolved matter deposited on the surface and from 16 to 19 per cent of insoluble matter deposited there was brought down by rain. It is not known, however, whether these materials were present in the cloud particles from which the rain developed or whether they were washed out of the contaminated atmosphere by the rain drops as the latter fell. An investigation in fourteen large American cities [42] indicated that rain had no effect in cleansing the air, but rather that it tended to increase the concentration of smoke near the ground. The evidence at Leicester indicates that rain does not reduce the concentrations of sulfur dioxide in the air at all rapidly.

The Annual Variation. Near cities in extratropical latitudes there is a marked annual variation, with a maximum in winter and a minimum in summer. A large proportion of the winter pollution is due to products of combustion released from heating plants in buildings, but industrial contaminants emitted are not subject to the same degree of annual variation. Because of this complication, the effects of the annual variation of atmospheric conditions on pollution are not at all well known. The winter maximum and summer minimum are well marked at New York City [17] and at Leicester. At the latter, smoke has the largest range of annual variation, and insoluble deposit the smallest.

The Weekly Variation. This is due, not to meteorological factors, but to the shutting down, partially or completely, of industrial plants during the week end. It will therefore not be discussed here.

The Daily Variation. Surveys of smoke and dust in American cities, such as those for New York [17] and for a group of fourteen large cities [42], and in British cities, such as London, Glasgow, Blackburn, and Stoke-on-Trent [28], show a maximum at times ranging from 6:30 to 9 A.M. and a secondary maximum in the late afternoon. Increased industrial activity and home and office heating in the morning taken in conjunction with increasing turbulence associated with solar heating, as outlined earlier, account for the morning maximum; decreasing turbulence in the late afternoon before industrial emission of pollution has been reduced causes the secondary maximum.

Effect of Topography. As with single sources, the pollution from multiple sources, as represented by an industrial city, is markedly affected by topographic features. In a valley, for example, under stagnant atmospheric conditions, serious concentrations may and sometimes do develop; the valley sides act as physical barriers which prevent the lateral diffusion which normally occurs over more level terrain. Two extreme examples may be quoted. During the period December 1–5, 1930, severe pollution occurred in the Meuse valley in Belgium [22]. Anticyclonic conditions prevailed during the period, with fog and very light winds which carried pollution from the city of Liége and from industrial plants nearby into a narrow portion of the valley; an inversion accentuated the pollution. The conditions were so severe that several hundred persons suffered from acute respiratory troubles and sixty-three died on December 4 and 5. A similar disaster occurred at Donora, Pa., during the last few days of October, 1948 [23, 99]. Donora lies in the relatively deep and narrow valley of the Monongahela River at a distance of about twenty miles from Pittsburgh. During the five-day period of light winds associated with anticyclonic conditions, the pollution became severe and twenty persons died and hundreds were stricken in and near Donora. It appears from these instances that if stagnant anticyclonic conditions persist for four days or more, an occurrence most probable in middle latitudes in the autumn, fatal concentrations of pollution may develop. Topographic effects on pollution near some great cities are important. For example, the mountain ranges near Los Angeles are a factor at certain times in increasing the pollution problem there [54].

Because of the proven danger of lethal concentrations of contaminants to valley communities, there is an urgent need for both theoretical and observational programs of study of valley pollution. Given a knowledge of certain meteorological quantities in and just above a valley, of the rate at which contaminants are entering the valley, and of the topography of the valley, it should be possible to compute from approximate expressions the rate of accumulation of impurities in a given section of the valley during prolonged stagnant atmospheric conditions. Such data would forewarn of approaching critical concentrations and of the necessity for preventive action. As outlined in a later section, the climatological use of such data would also be most valuable.

Serious accumulations of contaminants are more likely to occur at inland locations where, in general, surface winds are lighter than near the coast line [104].

Further Research. There is a need for many more measurements of pollution in, near, and above industrial cities and of the associated pertinent meteorological parameters. Relatively little attention has been paid to concentrations at large distances from industrial cities; in the Leicester report [24, p. 126] it is suggested that downwind from a city the size of Leicester the smoke diminishes as the distance for the interval 4 to 10 miles; farther from the city, from say 10 to 100 miles, it diminishes as the square of the distance; and beyond about 100 miles it diminishes more slowly again, and ultimately as the distance. Such a tentative formulation of large-scale aspects is useful, but observations are needed to permit confirmation or modification

of the thesis. An analysis by Brunt [14] of surface dust counts for one year taken over a distance interval of ten miles downwind from the center of Norwich, England, shows that the pollution is approximately inversely proportional to the distance from the center of the city and that it diminishes with increasing wind speed. Suggestions as to research on various aspects of the problem of pollution by cities have been made [24, pp. 139–145]. Studies of the following are suggested: the three-dimensional distribution of pollution under various meteorological conditions; the distribution and characteristics of country pollution, including the heights reached by it; the processes of natural removal of pollution; atmospheric turbulence in its association with pollution; pollution in city parks, where there is no emission of pollution, for the detailed information to be gained; and special aspects of the problem which could readily be investigated by local authorities without the services of a highly trained staff.

The theoretical aspects of pollution by multiple sources have as yet received little attention. The problem to be solved is essentially that of the pollution caused by a large number of point sources of a widely varying nature. A start has been made by Sherwood [75], who applied Sutton's concepts to study theoretically the screening produced by the smoke emitted from a number of equidistant continuous point sources arranged along a line lying across wind. The problem presented by an industrial city is of course extremely complex: the point sources are irregularly located over an area, at varying heights, are emitting a variety of contaminants at differing rates, which also vary with time, into an atmosphere whose diffusing properties vary with both periodic and aperiodic components. Such a statement may merely induce deep pessimism, but it also suggests that an attempt to handle the problem of pollution from such an array of point sources by statistical combinations of individual point sources may prove fruitful. It is obvious that a complete theoretical treatment cannot be expected in the near future, but a start could be made by singling out one of the numerous variables and determining, if possible, the patterns of pollution resulting from various statistical representations of the chosen variable. Some of the other variables might then be handled by other appropriate statistical simplifications, leading gradually to a theory of increasing scope and validity. The measurement of the distribution of tracer substances emitted according to plan from a number of suitably chosen sources would be invaluable for checking theoretical distributions.

SECONDARY EFFECTS ON METEOROLOGICAL ELEMENTS

The possibility that atmospheric pollution has an effect on some of the meteorological variables and on the microfeatures of weather must be considered. For example, the question of the influence of a pall of smoke over a city under stagnant atmospheric conditions on both solar and terrestrial radiation is an interesting one which merits attention. It is possible that under such conditions differential heating or cooling between the polluted city air and the cleaner air over surrounding country districts leads to local convergence or divergence which in itself may act so as to increase concentrations over certain areas and at certain heights while causing decreased concentrations elsewhere.

Rainfall. Climatological studies of rainfall at Rochdale, England, indicate that both the average amount and average rate of rainfall are slightly smaller on Sunday than on other days of the week [2, 3]. Since the factories were closed on Sunday, it is inferred that condensation on the nuclei produced by industrial processes leads to the increased rainfall on the other days of the week. It has been shown that the Ruhr region of Germany gets measurable rain or drizzle on twenty more days per annum than do nearby less industrialized areas [103].

Fog. Measurements of relative humidity in fog often give values substantially less than 100 per cent, particularly in industrial areas. From this we may infer that atmospheric pollution increases the incidence and duration of fogs. There is also direct evidence: a long series of observations in Prague, Czechoslovakia, show that there were, on the average, about eighty-two days with fog per year in the period 1800–1880, and that since that period the number has nearly doubled [41]. The observations were made continuously at the same point and by the same rules. The effect of increasing industrialization and its attendant pollution is thus marked. It has been suggested [43] that sulfur dioxide may contribute directly to fog formation under certain conditions when photochemical oxidization by ultraviolet light is sufficient to produce sulfur trioxide, which is highly hygroscopic. On the other hand, there are indications [62] that such combustion nuclei alone cannot be responsible for the difference between relative humidities in fog in industrial and rural districts. The possibility that radiational cooling of contaminating particles, and thus of the ambient air, is a significant factor has been considered [49]. Other phases of the behavior of fog in a polluted atmosphere have been discussed [20]: in industrial cities it has been observed that a dirty fog at, say, 9 P.M. changes to a clean one by about 5 A.M. next morning, without a clearing of the fog, suggesting that the dirty particles fall out and are replaced by clean ones; fog droplets in town air dissolve the sulfur dioxide present, forming sulfurous acid which will oxidize to sulfuric acid which in turn acts to retard the clearing of such a fog. In a laboratory study of artificial fogs [65] it was found that, as pollution increases, a corresponding increase of fog density occurs only until the pollution attains a value characteristic of a small town. On the other hand, there is a continuous increase in fog duration with increasing air pollution. Further laboratory investigations [64] have shown the relative effectiveness of various combustion products and of the quality of the combustion process in prolonging the fog and in increasing its density. Incomplete combustion tends to increase both fog den-

sity and duration. From the foregoing brief discussion it will be obvious that the processes of mutual interaction between pollution and fog need further study.

LIMITATION AND CONTROL OF ATMOSPHERIC POLLUTION

There are in existence various laws whose purpose it is to limit pollution at the source [32]. Thus, in the United States, California has an air-pollution-control law and several counties and many municipalities have ordinances. This regulation of pollution is being undertaken by the following methods: by the sampling and inspection of discharges, in the case of smoke mainly through the use of the Ringelmann chart; by permits to operate existing equipment; by permits to build or install new equipment from which atmospheric contaminants may be emitted; and by control of the quality and kind of fuel used. Limitation of sulfur dioxide emission has in a few cases been specified. Thus the regulations of the Los Angeles County Air Pollution Control District permit a maximum emission of 2000 parts per million. In Great Britain the limitation tends to be more restrictive: the London County Council permits a maximum concentration of 35 ppm. Some of the possible means of limitation are mentioned below.

Limitation of Contaminants Emitted. There are two possibilities: to change plant processes so as to reduce contaminants in effluents to acceptable levels; or to remove or treat contaminants before they reach the atmosphere. Since the latter procedure usually involves less dislocation of plant facilities and less expense, it has been much more widely adopted. Methods of removing sulfur compounds and dust and fumes from waste gases have been described by Johnstone [43] and elsewhere [1].

Effect of Stack Height. It is clear from the discussion of the theory that high stacks are useful in alleviating surface pollution. However, it has been pointed out by Bosanquet and Pearson [12] that any increase above a reasonable height [68] has no practical effect. In other words, the use of high stacks leads to amelioration up to a certain point; to obtain further improvement, other methods must be used.

Effect of Stack Temperature. When effluent gases are warmer than the ambient air they have a buoyancy which causes them to rise. Their ascent ceases when, or shortly after, their temperature drops to that of the surrounding air. In this connection adiabatic cooling is of secondary importance in comparison with turbulent transfer of heat from the effluent gases to the air through which they rise. For given temperatures of effluent and atmosphere, the height to which the gases rise depends on the volume emitted per unit time, on the turbulence of the surrounding atmosphere, and on the wind speed. With low turbulence and low winds the eddy transfer of heat is relatively small and the gases rise nearly vertically to considerable heights, the height attained being greater with greater volume emitted per unit time. With large volumes emitted the gases may rise 400 to 600 ft; with even greater volumes they may rise 1000 to 1200 ft. With a highly turbulent atmosphere and higher winds the initial rise may be only one or two hundred feet or less. The height to which gases rise before leveling off has been called the "effective stack height" by Beers [8].

The problem has been discussed theoretically by Schmidt [71] and by Sutton [85]. Schmidt used approximate methods of solution involving infinite series, whereas Sutton postulated that if temperature differences between the vertical jet and the surrounding air are not too large, the spread of a vertical stream of hot gas must bear a close resemblance to the diffusion in the atmosphere of a cloud of cold smoke from a continuous point source. No wind and a dry adiabatic lapse rate are assumed, and the turbulence of the general environment is considered negligible in comparison with that induced by the jet itself. For the upward velocity w of the warm effluent Schmidt obtains an expression of the general form

$$w = \text{const}\, z^{-1/3}, \tag{10}$$

the constant being evaluated in terms of infinite series, whereas Sutton's corresponding equation is

$$w = \left(\frac{7gQ}{3\pi c_p \rho C T}\right)^{1/3} z^{-0.29}, \tag{11}$$

where g is the acceleration of gravity, Q is the heat supplied at the source per unit time, c_p is the specific heat of dry air at constant pressure, ρ is the air density, C is a diffusion coefficient with an approximate value of 0.3 $cm^{1/8}$, and T is the mean absolute temperature of the undisturbed air. Strictly speaking, since the actual source will not be a point, but an area, the height z is not measured from the mouth of the stack but from some lower level whose position depends on the size of the orifice. Both theories give results of the correct order of magnitude, but detailed checks with actual observations are required. Furthermore, the important case of calm conditions associated with an inversion is not covered by the theories. Sutton also obtains an approximate solution of the problem when there is a wind but, as he points out, the uncertainties then are much greater.

Meteorological Control. The possibility of alleviating pollution by varying the emission of contaminants as the diffusing power of the atmosphere varies has as yet received relatively little attention. Such a procedure has been used successfully by the smelter in the Columbia River valley at Trail, British Columbia. There, in order to prevent damage to vegetation downvalley (to the south) in the state of Washington, an upper limit to emission of sulfur dioxide has been set: with marked turbulence and hence rapid diffusion of the sulfur dioxide the maximum permissible rate of emission is much greater than when turbulence and diffusion are small; the upper limit is also a function of wind direction and speed [19, 37, 38]. The upper limits as specified by the Trail Smelter Arbitral Tribunal are given in Table I.

The degree of turbulence is specified in terms of the bridled-cup turbulence integrator [19], as developed by Gill. This instrument may be described as a gust accelerometer. Each change of wind speed of 2 mph causes a deflection in the trace made by a moving pen. The degree of turbulence is specified as the number of deflections per half hour; if this number is multiplied by four, the horizontal gust acceleration in miles per hour per hour is obtained. The turbulence integrator thus measures a basic parameter of the turbulence field. It has been found that winds at Trail from the north, east, south, southwest, and intermediate directions, and with a speed of 5 mph or more, generally do not carry the smoke downvalley and into the state of Washington, and so are considered favorable. Winds from directions other than these and a wind in any direction with a speed of less than 5 mph are unfavorable. Because the degree of susceptibility of vegetation to damage by sulfur dioxide varies with the time of day and the season of the year, the permissible maxima have corresponding variations—high maxima for low susceptibility and vice versa. Overriding safeguards are provided in other clauses of the control regime. Because of the great difficulty of forecasting winds and turbulence in a valley such as that in which the smelter is located, the control is based on contemporary, not forecast, conditions. Over more regular terrain it should be possible to forecast the pertinent parameters to a satisfactory degree of accuracy.

Table I. Maximum Permissible Sulfur Emission
(tons of contained sulfur per hour)

Time	Season	Turbulence*							
		0–74		75–149		150–349		≥350	
Midnight to 3 A.M.	Growing	2	*6*	6	*9*	9	*11*	11	
	Nongrowing	2	*8*	6	*11*	9	*11*	11	
3 A.M. to 3 hr after sunrise	Growing	0	*2*	4	*4*	4	*6*	6	
	Nongrowing	0	*4*	4	*6*	4	*6*	6	
3 hr after sunrise to 3 hr before sunset	Growing	2	*6*	6	*9*	9	*11*	11	
	Nongrowing	2	*8*	6	*11*	9	*11*	11	
3 hr before sunset to sunset	Growing	2	*5*	5	*7*	7	*9*	9	
	Nongrowing	2	*7*	5	*9*	7	*9*	9	
Sunset to midnight	Growing	3	*7*	6	*9*	9	*11*	11	
	Nongrowing	3	*9*	6	*11*	9	*11*	11	

* Deflections of turbulence integrator per half hour. Sulfur emission for unfavorable winds (left columns) and for favorable winds (right columns).

Methods of forecasting smog for the Los Angeles area, to permit control of the amount of contaminating substances discharged into the air when smog is imminent, are being developed [54, 98]. A study of the relationship between past occurrences of smog and past weather conditions has led to the development of a smog index S which is expressed in terms of meteorological variables.

$$S = \frac{10(T_D + 10)}{RW}\left(\frac{I}{V}\right)^{1/2}, \tag{12}$$

where: T_D = deviation in degrees Fahrenheit of the 24-hour mean temperature from the mean temperature for that particular day of the year;
R = relative humidity at noon;
W = total 24-hour wind movement in miles;
V = noon visibility in miles; and
I = inversion intensity from the equation

$$I = \frac{(\Delta\theta)^2}{3 + z\Delta z}, \tag{13}$$

where: $\Delta\theta$ = change in potential temperature in degrees Kelvin through the inversion layer;
z = height of the inversion base, in hectometers; and
Δz = thickness of the inversion layer, in hectometers.

Agreement is marked between days for which the smog index is large (indicating smog conditions) and days on which more than the minimum of four complaints of smog were received by the Los Angeles County Air Pollution Control District. The success of this index led to a method of forecasting smog in terms of the relationship between the presence or absence of smog and upper-air pressures and air movements. Smog occurs only when the three-day sum of the deviations from the normal height of the 700-mb surface lies within the limits 0 and +1000 ft (700-mb surface higher than normal) and the three-day sum of the daily average wind speeds at 10,000 ft is less than 80 mph (light winds at 10,000 ft). When the two parameters lie outside these limits there is a high probability that smog will not occur on the day for which the computation was made nor on the two following days.

Investigations made by Smith [76] and associates at Brookhaven National Laboratory have laid the foundation for a program of meteorological control of effluents from the nuclear reactor there. Predominant gustiness is classified according to types denoted by the symbols A, B, C, and D, corresponding to marked thermal turbulence, a combination of thermal and mechanical turbulence, marked mechanical turbulence, and turbulence with small vertical components of motion, respectively. These four types are illustrated in Fig. 5. The occurrence of each type in terms of gradient wind speed, vertical temperature distribution, time of day, and season of the year is specified; such specifications simplify the forecasting of the gustiness. The effluent to be controlled is argon-41, with a half-life of 110 minutes; the stack concentrations are so low that no health hazard is encountered from instantaneous maximum concentrations. Rather, it is mean radiation dosage over a number of days which has been selected as the operating criterion. A series of templates, each giving contours of mean hourly radiation dosage at ground level for a particular combination of wind speed and gustiness, were prepared on the basis of experimental data on oil-fog concentrations. The forecast radiation dosage, obtained from the templates, is added to the accumulated dosage over a period of days obtained by use of the templates in conjunction with the

known past meteorological conditions. By this procedure the forecaster is able to determine whether or not there is any likelihood that continued operation of the reactor would result in a radiation dosage higher than that specified as the maximum operating level.

increasing production costs; or by the construction of auxiliary plants for the recovery and processing of commercially valuable materials now in the effluents, to which plants varying amounts of the effluent could be diverted in accordance with prevailing meteorological

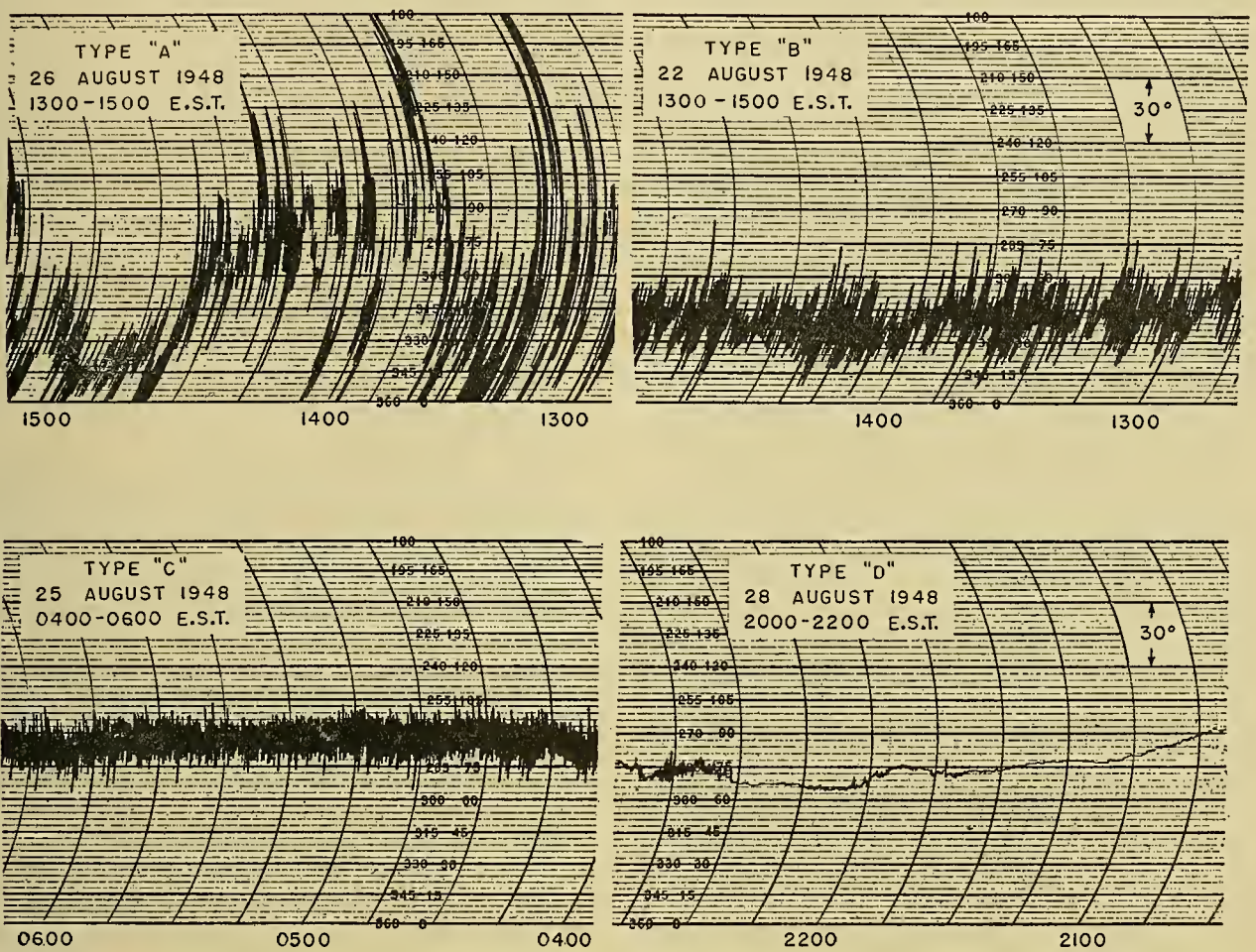


FIG. 5.—Four types of gustiness shown by the variability of wind direction as indicated by a wind vane at top of experimental stack at Brookhaven, according to Smith. *A*—great instability; *B*—moderate instability; *C*—moderate stability; and *D*—great stability.

In the report on the Donora investigations [99], it is recommended that production in local and nearby industrial plants be curtailed when stagnant atmospheric conditions develop in the valley. Such conditions are to be anticipated when an extensive, slowly moving anticyclone approaches the eastern United States and gives indications of stagnating; and when, in addition, the stability of the air in the valley exceeds a specified value, valley and upper winds are light and less than specified values, and moderate to dense fog in the valley continues for some time past noon.

The examples given above are sufficient to indicate the possibilities of forecasting future diffusion conditions. Such meteorological control as outlined above requires flexibility of plant operations. The requisite flexibility might be attained in several ways: by modifying plant processes so as to permit some variation of the rate of emission of contaminants but without unduly

conditions. It is probable that the cost of such a program applied to existing plants would in most cases be prohibitively high, but if the possibilities of meteorological control were seriously considered in the design of new plants and in the extensive modernization of older ones, a gradual amelioration of atmospheric pollution would be achieved over a period of years.

In urban areas the pollution by domestic heating units presents a serious problem. A wider use of central heating plants in new housing developments would be helpful. Such plants, with high stacks and a design which permits some degree of meteorological control, would reduce pollution. Such possibilities, and those of industrial control, should receive serious consideration by city planning commissions.

Although the alternative of removal of contaminants at the source has many attractive features, the problem of the disposal of such collected wastes may well become

increasingly acute; most rivers already receive more wastes than is desirable, and the dumping of wastes out-of-doors in large piles mars the landscape and raises other problems. However, these are not the only possibilities. For example, the problem of the disposal of economically worthless smokes and fine dusts collected at the source could be alleviated by turning off Cottrell precipitators or other precipitating equipment when atmospheric turbulence is high and thus allowing the atmosphere to disperse the contaminants in very low concentrations over a wide area. A possible extension of this method is to collect the fine particles during periods of low turbulence and then reintroduce them into the flue gases during periods when the diffusing action of the atmosphere is very high. A successful development of this method would thus permit atmospheric disposal of all but the large particles.

For a substantial fraction of the time the atmosphere is an excellent diffusing agency, spreading gases and aerosols over such wide areas in such low concentrations that they present no problem. It seems only reasonable to use this natural diffusing agency in the disposal of industrial wastes whenever possible.

Climatological Planning. Although the problem of waste disposal in the atmosphere is only one of a number to be considered when selecting a site for a new plant, considerable alleviation of pollution could be achieved if more weight were given to this factor. It is obvious that, if at all possible, a new plant should be located so that the prevailing winds will carry contaminants away from, not toward, populated areas, fertile agricultural land, valuable timber stands, etc. The prevalence of calm conditions, both persistent and transient, which lead to small lapse rates and inversions near the surface, taken in conjunction with local orographic conditions, should also be considered [104]. Thus, in the Los Angeles area, the records show that, on the average, inversions occur 260 days a year, of which 65 days are characterized by marked inversions. The persistence of such inversions is important in assessing the probable pollution nuisance. If it is found desirable to locate industrial plants in areas where natural diffusing conditions are poor, a knowledge of that fact would permit the design of plant equipment for the effective mitigation of the potential nuisance.

If a system of meteorological control is to be established, a study of the pertinent climatological data would provide information on the degree of curtailment of emission of contaminants desirable at various times of the year, the duration of periods of curtailment, and other facts of value in designing the requisite plant equipment. If substances are to be removed from the effluent and used in the manufacture of by-products, climatological data would permit an estimate of the required plant capacity and of its probable utilization throughout the year.

Methods of using climatological data for the solution of certain problems have been developed. Down-wash of contaminants in large-scale eddies in the lee of buildings may be marked when winds are high. Sherlock [72] has described a method of using wind data to determine how frequently such down-wash is likely to occur over a period of years. Lowry [52] has developed a basis for a pollution climatology. The necessary data are obtained from an anemometer and a wind vane at the proposed level of the top of the stack. Maximum ground concentrations are obtained from equation (7); the wind vane gives the appropriate value of the quantity a_m and the anemometer gives the value of the mean wind speed u, both for the period of one hour. The quantity σ in (8) is obtained from the wind vane. The two equations then give the mean maximum ground concentration for one hour and the distance from the source at which this mean maximum occurs. The longer the exploratory observations are continued the more accurate and reliable will be the results. To facilitate the use of the method, mean values are given of a_m and σ appropriate for the four types of gustiness shown by the wind vane and illustrated in Fig. 5. Allowance has been made for the high morning concentrations which often occur in summer, but the method is subject to the other limitations of current theories which have been described earlier.

Climatological planning is especially urgent in connection with industrial installations in valleys, as demonstrated by the disasters in the Meuse River valley in Belgium and at Donora in the Monongahela River valley which have been described earlier. Climatological studies of the relevant meteorological parameters would reveal how often dangerous accumulations in the valley are likely to occur, at what seasons of the year, and their probable duration. Such studies would also provide data for industrial plants in the valley to permit them to design units of the necessary capacity for collecting contaminants at the source.

THE PRESENT POSITION AND FUTURE REQUIREMENTS

A survey of atmospheric pollution studies indicates that advance in the meteorological aspects of the problem has been most rapid in Great Britain and the United States.

Systematic studies began earlier in Great Britain when, in 1912, a conference of delegates of municipal authorities and others was held in connection with the Smoke Abatement Exhibition in London. At that time a committee was appointed to "draw up details of a standard apparatus for the measurement of soot and dust and standard methods for its use." The committee's first report covered operations during the period April, 1914, to March, 1915. Since that time systematic measurements of contaminants in a number of cities have been made. However, detailed studies of the relationship between city pollution and meteorological conditions were not made until 1937, when a three-year survey of the city of Leicester [24] was commenced. Important experimental and theoretical contributions to the solution of problems presented by point and line sources have been made independently by Sutton and the group which worked under him at the Chemical Defence Experimental Station, Porton.

Uncoordinated work on a smaller scale has gone on for a number of years in the United States, but with a great increase in interest and activity during the past five years. Increased industrialization and, in particular, the problems of disposing of wastes from atomic energy installations have led to fresh attacks on atmospheric pollution. The former has produced such acute conditions in the Los Angeles area that a major investigation of all phases of the problem is now in progress. The latter have led to valuable studies of the behavior of effluents from experimental stacks.

Having considered the present position, it may be worth while to devote a little attention to some of the problems which may have to be faced in the future. In particular, general methods of procedure are worth discussing. For present purposes, possible future investigations may be classed either as extensive or intensive.

An extensive program of investigations is defined here as one in which a variety of problems are studied by a number of small and independent groups, coordination of effort among the different groups being limited to that achieved as a result of voluntary mutual consultation and discussion. The writer believes that such a program, because it provides the widest scope for individual initiative and inspiration, is in most cases the more desirable. Studies which have been made of pollution from single sources may be thought of as comprising an extensive program. The nature of the single-source problem is such that substantial progress may be made by such groups: the problems chosen for study may be delimited without sacrificing the essential usefulness of the results; relatively small groups of research personnel are sufficient, a few specialists in several fields at most being needed; instrumental requirements for gaining limited objectives are not excessive; and support and sponsorship by several interested organizations will be adequate. Furthermore, for these reasons the necessary financial support is not too great and can be obtained without undue difficulty.

However, there are other phases of the atmospheric pollution problem the successful solution of which requires resources and facilities of a higher order of magnitude. The outstanding example is that presented by multiple sources, as represented by an industrial city. Here the problem is one of great complexity and it should be studied simultaneously from a number of points of view if results of value are to be achieved. Such an investigation requires the services of groups of specialists and technicians equipped with facilities for measuring a variety of variables and for analyzing the observations in detail. In addition to substantial financial backing, the successful prosecution of such an intensive investigation of pollution in an industrial city requires the wholehearted cooperation of a number of organizations, such as municipal and county authorities, federal agencies, research institutions, and various industries whose plant operations contribute to the general city pollution. To give a specific example, the cooperation of radio stations in allowing meteorological instruments to be installed on their transmitting towers would be most helpful. Such an intensive program requires a higher degree of coordination and integration of effort than do the extensive studies with limited objectives referred to above.

The writer believes that both extensive and intensive programs of research will in the future make their contribution. The need for more coordination of effort in certain areas was clearly recognized at the U. S. Technical Conference on Air Pollution, held in Washington, D. C., on May 3–5, 1950. A Steering Committee was set up by the Conference to explore ways and means of extending the areas of cooperation; the meteorological representative on this Committee is Dr. H. E. Landsberg. During the meetings of the Meteorology Panel of the Conference the need for more research of both the extensive and intensive types was expressed. A report of the discussions of the Meteorology Panel was prepared by a group consisting of Dr. H. E. Landsberg and Dr. H. Wexler, the panel co-chairmen, Dr. E. W. Hewson, the panel chairman, and, by invitation, Dr. O. G. Sutton, chairman of the British government's Atmospheric Pollution Research Committee. This report was presented to the plenary session of the Conference [93]. Because of the fundamental nature of the problems considered, the report is reproduced in full below.

The panel meetings have revealed how extensive and how varied is the attack on the meteorological phases of the problem of air pollution. We are encouraged by the evidence of great activity and effort.

The present position may be summarized as follows. The distribution of pollution from a single source, either elevated or at the ground, has been studied both theoretically and experimentally for various meteorological conditions. Valuable information has been gained concerning the influence of special factors such as the following: diurnal variations of atmospheric diffusion; frontal surfaces and subsidence inversions aloft; deposition and coagulation of particulate matter in the atmosphere; and topography. Progress has been made in the study of pollution from multiple sources, such as represented by a city. The average distribution of contaminants in a city is governed by wind, rain, atmospheric stability, and topographic features. The contaminants in their turn influence rainfall and fog occurrence and persistence. However, our knowledge of city pollution and its dependence on weather conditions is less complete and detailed than that of pollution from a single stack.

Meteorological studies will aid in programs for minimizing pollution. The effects of increasing stack heights and temperatures have been determined. Rates of emission may, in cases where plant operations permit, be varied in accordance with clearly defined and precisely stated meteorological criteria. Methods of using climatological data to aid in determining the optimum location of a new industrial plant are available.

In short, there is a substantial fund of meteorological knowledge available for use in meeting present problems.

However, we find we are hampered in several ways. Our efforts are not as fruitful as they should be because of: a lack of coordination of effort; a lack of a basic theoretical doctrine by which to orient our efforts; and a lack of standardization of instruments and terminology. Our review of the present position has revealed a number of weak spots which require

attention and a number of lines of attack along which valuable progress may be anticipated.

The following are our recommendations for future action:

RECOMMENDATIONS

1. That standardization in methods of handling several aspects of the problem be achieved at the earliest possible moment. In particular, standard instruments and procedures of evaluation should be evolved and adopted for use. Such standardized instruments, *e.g.*, for direct measurements of turbulence, are required both for routine and research investigations. If it is necessary to use non-standard instruments, a full description of the operating characteristics of the equipment should be given. There is also an urgent need to adopt a standard terminology to ensure precision in presenting and conveying information.

2. That the theoretical development of the subject be promoted in every way possible. The development of basic turbulence theory is a prerequisite for fundamental advance in the study of atmospheric diffusion, and must be fostered.

3. That the problem of diffusion of effluents in a city be investigated intensively, since the pollution nuisance reaches the ultimate there. A program in which experiment and theory are both fully developed toward a common end is required. For example, it is suggested that intensive surveys of pollution and the associated meteorological conditions be made by a team which would survey ten cities well distributed over the country in annual succession, the team spending a year in each city. Each city would be chosen as characteristic for certain features of terrain and climate. At the end of the ten-year period ten cities would have been studied in sufficient detail to permit a classification of diffusion conditions in each in terms of specific quantities. It may then be possible to extrapolate the detailed conclusions to other cities with the aid of only routine measurements in the latter for comparison.

4. That micrometeorological and microclimatological surveys of cities and their suburbs be undertaken by local authorities with the assistance of a professional meteorologist. Instruments of standard type should be mounted on existing radio masts at standard heights with measured values presented for standard times in standard units. A detailed outline of the suggested procedure to be followed by local authorities in establishing such a survey should be prepared and distributed. Such a local survey would be relatively inexpensive.

5. That whenever a meteorological survey is in progress, a continuous record of the magnitude and location of pollution should be provided to permit correlation with meteorological variables. This will require the establishment of a standard pollution index so that different cities can be compared.

6. That the United States Weather Bureau be encouraged to take more observations of micrometeorological significance at its stations. Emphasis should be placed on obtaining regular observations of gradients of temperature and wind in the lowest 500 feet of the atmosphere.

7. That the organization be undertaken of existing and current literature on all phases of atmospheric pollution so that everyone may be aware of the work of those in other phases. Publication might be in toto or in abstract form.

CONCERNING SPECIFIC PROBLEMS TO BE ATTACKED

8. That, if possible, the relationship between wind-tunnel flow and natural atmospheric flow be established.

9. That the means by which contaminants are removed from the atmosphere be studied in detail. Is the natural cleansing process rapid enough to warrant increasing the rate of disposal of wastes in the atmosphere? At times the atmosphere is an extremely efficient disperser of gases and aerosols. In the long run, it may prove to be preferable to use this natural diffusing agency rather than removing contaminants at the source; with atmospheric dispersal the problem of disposal of economically worthless substances does not arise.

10. That an aerial survey of the distribution of inversion fog be undertaken. Atmospheric pollution tends to accumulate over terrain where such radiation fogs are of frequent occurrence.

11. That atmospheric electricity be measured more widely. The conductivity of the air depends on the degree of pollution, and affords a convenient method of studying the secular trends of the latter [95, 100].

With serious attention given to the above and related problems, meteorology will continue to make an increasing contribution to the solution of the problem of atmospheric pollution.

PROMISING LINES OF FURTHER RESEARCH

In portions of some of the preceding sections the writer has outlined his ideas on promising lines of attack on various aspects of the problem of atmospheric pollution. He is greatly indebted to a group of specialists in the field who have, at his request, stated briefly their ideas on how we may best advance our knowledge of atmospheric pollution. These suggestions will be most valuable as guideposts to give direction to future research. The statements are given below.

N. R. Beers. It is important that one be able to determine the significant turbulence by using the ordinary weather map and simple and inexpensive observations at his own station. The principles involved in sound refraction may be applied to investigate inversion conditions with both source and receiver on the ground.

Fixed base observations along a horizontal line, as between the meteorological towers at Brookhaven, may be made to investigate the structure of eddies. I have suggested specifically the possibility of hanging some 100 to 200 small balloons between the towers, to be fastened to the 900-foot-long "horizontal" cables. Motion pictures will illustrate the air movement between the towers at desired heights up to 100 feet above ground.

A general investigation should be made respecting trajectories of the air over a station, and nuclei of various kinds, such as sea salt, radon, pollen, dust, etc. The investigation must be carried out at several stations and at several elevations above ground.

Phil E. Church. I believe that man has the right to breathe pure, clean, unpolluted air as well as to drink unpolluted water. It is not likely, however, that the time will soon come when no man-made pollutants of any type will be discharged into the atmosphere or when atmospheric pollutant sewage systems will be constructed. Until such time, the concentration of contamination reaching the level where man lives is largely at the mercy of the characteristics of the air into which the contaminant is ejected. Hence, we must know more about these atmospheric characteristics, particularly in and near large and growing industrial centers. Especially important characteristics which must be thoroughly known at all levels into which contaminants are or will be ejected are duration, frequency, and magnitude of stable and unstable lapse rates, and wind direction, gustiness,

shear, and speed. These may in turn, after experimental work to determine quantitatively the amount of dilution a contaminant undergoes under various meteorological conditions, determine how much contaminant may be released without nuisance or physiological harm.

G. M. B. Dobson. As a result of much work in recent years, we now have a fair knowledge of the distribution of pollution within a town and the mechanism of its removal from the air at street level. On the other hand we know very little about the travel of the smoke cloud down wind from a large town though its effect is clearly seen in reduced visibility 50 or 100 miles away. A survey, possibly using aircraft, of the spreading, vertically and horizontally, and the change in concentration of the smoke at different distances would be of much interest. The results would need to be correlated with the meteorological conditions.

A. R. Meetham. Ventilation of towns. The concentration of smoke and gaseous pollution at street level is only partly determined by wind, being less closely correlated with wind speed than with other factors related to atmospheric turbulence. There is no orthodox method of measuring ventilation by turbulence; but it can be studied in towns where there is a large central park or smokeless zone by measuring the ratio of pollution in the center of the park to that a short way upwind. In Hyde Park, London, the ratio varied from 0.27 in high turbulence to 0.85 in low turbulence [24, p. 71].

M. E. Smith. Progress in meteorological research associated with atmospheric pollution requires the accumulation of certain data which are currently difficult to obtain without elaborate facilities. It is therefore desirable that relatively inexpensive equipment and simple techniques be developed to obtain these data. A thorough investigation of the usefulness of radio-wave propagation for the measurement of vertical temperature and humidity structure is particularly promising, since the equipment presumably would not involve the use of fixed towers or masts.

The determination of the extent to which wind tunnels can be used to simulate atmospheric turbulence by variation of vertical density structure, as well as other parameters, is also worthy of consideration. Such an investigation would establish more clearly the significance of numerous tests which have been made, and might provide an invaluable research tool for the investigation of proposed industrial installations.

O. G. Sutton. The situation in Great Britain at present is as follows. There is in existence a well-coordinated system of observations in the vicinity of sources of pollution, using standard instruments (deposit gauges, lead peroxide cylinders, etc.) which are provided by the Fuel Research Station of the Department of Scientific and Industrial Research and operated by the local authorities (town and borough councils). This procedure ensures that observations taken in different localities are comparable, a very important point for the subsequent statistical investigation. The results are analyzed by the Fuel Research Station and the final statistics, with maps of distribution, are issued in an official publication.

The main effort of the past decade, apart from research on instrumentation, has been the study of pollution at a selected site. The best known example is the Leicester Survey [24]. Although attempts were made in this survey to investigate the underlying physics of the distribution of pollution, it is probably fair to describe this work as primarily concerned with establishing the facts (in a statistical sense) of the distribution of pollution around a center of industry. The nature of the raw data and the heterogeneous character of the source made any really detailed physical study impossible.

It is hoped that it will now be possible to make more individual studies. The mathematical theory, although incomplete and by no means fully established, is now sufficiently advanced to give direction to the necessary experimental work. This should take the form of wind-tunnel investigations, supplemented by full-scale observations, of the distribution of pollution from a specific source, such as a power plant. This technique should be useful, for example, in revealing any tendency towards dangerous pockets of pollution in heavily contoured areas, by employing relief maps of the district with chemical indicators. Secondly, it should prove possible to establish a technique of experimentation for the prior investigation of a proposed factory or power plant layout to ensure that the effluent has a reasonable chance of getting away without being caught in some local eddy system and thus brought down to ground level in a high concentration.

To sum up, attention should now be directed more towards the mathematical-physical and experimental study of individual sources, and less towards the statistical survey of a polluted area. The routine observations should continue at all costs, since they form the essential factual basis of the problem.

REFERENCES

1. Anonymous, "Air Pollution." *Mod. Ind.*, 18: 46–50 (1949).
2. Ashworth, J. R., "The Influence of Smoke and Hot Gases from Factory Chimneys on Rainfall." *Quart. J. R. meteor. Soc.*, 55: 341–350 (1929).
3. —— "Smoke and Rain." *Nature*, 154: 213–214 (1944).
4. —— "Atmospheric Pollution and the Deposit Gauge." *Weather*, 3: 137–140 (1948).
5. Barad, M. L., "Diffusion of Stack Gases in Very Stable Atmospheres." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
6. Baron, T., Gerhard, E. R., and Johnstone, H. F., "Dissemination of Aerosol Particles Dispersed from Stacks." *Industr. engng. Chem.*, 41: 2403–2408 (1949).
7. Beer, C. G. P., and Leopold, L. B., "Meteorological Factors Influencing Air Pollution in the Los Angeles Area." *Trans. Amer. geophys. Un.*, 28: 173–192 (1947).
8. Beers, N. R., "Stack Meteorology and Atmospheric Disposal of Radioactive Waste." *Nucleonics*, 4: 28–38 (1949).
9. —— *Meteorology as a General Factor in Air-Pollution Problems*. Proc. First Nat. Air Pollut. Sym., pp. 86–89, Stanford Res. Inst., Stanford, Cal., 1949.
10. Bennett, M. G., "The Physical Conditions Controlling Visibility through the Atmosphere." *Quart. J. R. meteor. Soc.*, 56: 1–29 (1930).
11. Bosanquet, C. H., "Eddy Diffusion of Smoke and Gases in the Atmosphere." *J. Inst. Fuel*, 8: 153–156 (1935).
12. —— and Pearson, J. L., "The Spread of Smoke and Gases from Chimneys." *Trans. Faraday Soc.*, 32: 1249–1264 (1936).

13. Brown, C. E., and Schrenk, H. H., "Standard Methods for Measuring Extent of Atmospheric Pollution." *Inform. Circ. U. S. Bur. Min.* No. 7210 (1942).
14. Brunt, D., "The Distribution of Pollution around Norwich" in "Report on Observations in the Year Ended 31st March, 1931." *Rep. Obsns. atmos. Pollut., Lond.* No. 17, pp. 50–58 (1931).
15. Chen, W. L., and Charmbury, H. B., "Location of Dust-Producing Areas." *Industr. engng. Chem.*, 41: 2400–2402 (1949).
16. Church, P. E., "Dilution of Waste Stack Gases in the Atmosphere." *Industr. engng. Chem.*, 41: 2753–2756 (1949).
17. Davidson, W. F., "A Study of Atmospheric Pollution." *Mon. Wea. Rev. Wash.*, 70: 225–234 (1942).
18. —— and Master, W., "Automatic Dust Sampling and Analyzing Instruments for Atmospheric Pollution Surveys." *Mon. Wea. Rev. Wash.*, 69: 257–260 (1941).
19. Dean, R. S., and others, "Report Submitted to the Trail Smelter Arbitral Tribunal." *Bull. U. S. Bur. Min.*, No. 453 (1944).
20. Dobson, G. M. B., "Some Meteorological Aspects of Atmospheric Pollution." *Quart. J. R. meteor. Soc.*, 74: 133–143 (1948).
21. Etkes, P. W., and Brooks, C. F., "Smoke as an Indicator of Gustiness and Convection." *Mon. Wea. Rev. Wash.*, 46: 459–460 (1918).
22. Firket, J., "Fog Along the Meuse Valley." *Trans. Faraday Soc.*, 32: 1192–1197 (1936).
23. Fletcher, R. D., "The Donora Smog Disaster—A Problem in Atmospheric Pollution." *Weatherwise*, 2: 56–60 (1949).
24. Great Britain, Department of Scientific and Industrial Research, "Atmospheric Pollution in Leicester: A Scientific Survey." *Tech. Pap. atmos. Pollut. Res., Lond.* No. 1 (1945).
25. Great Britain, Department of Scientific and Industrial Research; The Investigation of Atmospheric Pollution, "Report on Observations in the Year Ended 31st March, 1937." *Rep. Obsns. atmos. Pollut., Lond.* No. 23 (1938).
26. Great Britain, Meteorological Office, Advisory Committee on Atmospheric Pollution, "Report on Observations in the Year 1917–1918." *Rep. adv. Comm. atmos. Pollut., Lond.* No. 4. (See pp. 20–23)
27. —— "Report on Observations in the Year Ending March 31st, 1921." *Rep. adv. Comm. atmos. Pollut., Lond.* No. 7, M. O. 249 (1922). (See pp. 27–31)
28. —— "Report on Observations in the Year Ending March 31st, 1921" and later Reports. *Rep. adv. Comm. atmos. Pollut., Lond.* No. 7 and following (1922 and later).
29. —— "Report on Observations in the Year Ending March 31st, 1923." *Rep. adv. Comm. atmos. Pollut., Lond.* No. 9, M. O. 260 (1924). (See pp. 46–59)
30. —— "Report on Observations in the Year Ending March 31st, 1924." *Rep. adv. Comm. atmos. Pollut., Lond.* No. 10, M. O. 270 (1925). (See pp. 52–53)
31. —— "Report on Observations in the Year Ending March 31st, 1925." *Rep. adv. Comm. atmos. Pollut., Lond.* No. 11, M. O. 280 (1925). (See p. 42)
32. Hartmann, M. L., "Legal Regulation of Air Pollution." *Industr. engng. Chem.*, 41: 2391–2395 (1949).
33. Hebley, H. F., "Factors Rarely Considered in Smoke Abatement." *Mech. Engr.*, 69: 281–287 (1947).
34. —— *Atmospheric Pollution Caused by the Diffusion of Waste Industrial Gases.* Proc. First Nat. Air Pollut. Sym., pp. 97–102, Stanford Res. Inst., Stanford, Cal., 1949.
35. Hemeon, W. C. L., "Instruments for Air Pollution Measurement." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
36. —— and Hatch, T. F., "Atmospheric Pollution." *Industr. engng. Chem.*, 39: 568–571 (1947).
37. Hewson, E. W., "Atmospheric Pollution by Heavy Industry." *Industr. engng. Chem.*, 36: 195–201 (1944).
38. —— "The Meteorological Control of Atmospheric Pollution by Heavy Industry." *Quart. J. R. meteor. Soc.*, 71: 266–282 (1945).
39. —— "Meteorological Aspects of Atmospheric Pollution" in *United States Technical Conference on Air Pollution.* New York, McGraw, 1951.
40. Hohenleiten, H. L. von, and Wolf, E. F., "Wind-Tunnel Tests to Establish Stack Height for Riverside Generating Station." *Trans. Amer. Soc. mech. Engrs.*, 64: 671–683 (1942).
41. Hrudička, B., "Zu den optischen und akustischen Eigenschaften des Klimas einer Groszstadt." *Beitr. Geophys.*, 53: 337–344 (1938).
42. Ives, J. E., and others, "Atmospheric Pollution of American Cities for the Years 1931 to 1933." *Publ. Hlth. Bull., Wash.*, No. 224 (1936).
43. Johnstone, H. F., "Technical Aspects of the Los Angeles Smog Problem." *J. industr. Hyg. Tox.*, 30: 358–369 (1948).
44. —— Winsche, W. E., and Smith, L. W., "The Dispersion and Deposition of Aerosols." *Chem. Rev.*, 44: 353–371 (1949).
45. Katz, M., "Sulfur Dioxide in the Atmosphere and Its Relation to Plant Life." *Industr. engng. Chem.*, 41: 2450–2465 (1949).
46. Kehoe, R. A., *Air Pollution and Community Health.* Proc. First Nat. Air Pollut. Sym., pp. 115–120, Stanford Res. Inst., Stanford, Cal., 1949.
47. Kimball, H. H., and Hand, I. F., "Investigations of the Dust Content of the Atmosphere." *Mon. Wea. Rev. Wash.*, 59: 349–352 (1931).
48. Kramer, H. P., and Rigby, M., "Cumulative Annotated Bibliography on Atmospheric Pollution by Smoke and Gases." *Meteor. Abstr. & Bibliogr.*, 1: 46–71 (1950).
49. Kraus, E. B., "Note on Fog and Atmospheric Pollution." *Quart. J. R. meteor. Soc.*, 73: 188–190 (1947).
50. Landsberg, H., "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, Supp. 3: 155–252 (1938).
51. Lessing, R., "The Industrial Aspects of Disperse Systems in Air and Gases. Introductory Paper." *Trans. Faraday Soc.*, 32: 1223–1233 (1936).
52. Lowry, P. H., "Microclimate Factors in Smoke Pollution from Tall Stacks." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
53. —— Mazzarella, D. A., and Smith, M. E., "Ground-Level Measurements of Oil-Fog Emitted from a Hundred-Meter Chimney." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
54. Magill, P. L., "The Los Angeles Smog Problem." *Industr. engng. Chem.*, 41: 2476–2486 (1949).
55. Marks, L. S., "Inadequacy of the Ringelmann Chart." *Mech. Engr.*, 59: 681–685 (1937).
56. Marsh, A., *Smoke.* London, Faber, 1947.
57. McCabe, L. C., and others, "Industrial Dusts and Fumes in the Los Angeles Area." *Industr. engng. Chem.*, 41: 2486–2493 (1949).
58. McElroy, G. E., and others, "Dilution of Stack Effluents." *Tech. Pap. Bur. Min., Wash.*, No. 657 (1944).
59. Meetham, A. R., "Notes on Atmospheric Pollution and the Deposit Gauge." *Weather*, 3: 140–143 (1948).
60. —— and Monkhouse, A. C., "Atmospheric Pollution and

the Chemist." *Industr. Chem. chem. Mfr.*, 23: 443–448 (1947).

61. Middleton, W. E. K., *Metcorological Instruments*. Toronto, University of Toronto Press, 1943.
62. Neiburger, M., and Wurtele, M. G., "On the Nature and Size of Particles in Haze, Fog, and Stratus of the Los Angeles Region." *Chem. Rev.*, 44: 321–335 (1949).
63. Neuberger, H., "Air Pollution." *Miner. Ind.*, 18: 1–3 (1949).
64. —— "Air Pollution and Fog Properties" in *United States Technical Conference on Air Pollution*. New York, McGraw, 1951.
65. —— and Gutnick, M., *Experimental Study of the Effect of Air Pollution on the Persistence of Fog*. Proc. First Nat. Air Pollut. Sym., pp. 90–96, Stanford Res. Inst., Stanford, Cal., 1949.
66. Owens, J. S., "The Measurement of Atmospheric Pollution." *Quart. J. R. meteor. Soc.*, 44: 149–170 (1918).
67. Parr, A. E., "On the Probable Relationship between Vertical Stability and Lateral Mixing Processes." *J. Cons. int. Explor. Mer.*, 11: 308–313 (1936).
68. Pearson, J. L., Nonhebel, G., and Ulander, P. H. N., "The Removal of Smoke and Acid Constituents from Flue Gases by a Non-effluent Water Process." *J. Instn. elect. Engrs.*, 77: 1–48 (1935).
69. Richardson, L. F., and Proctor, D., "Diffusion over Distances Ranging from 3 km. to 86 km." *Mem. R. meteor. Soc.*, 1: 1–15 (1926).
70. Roberts, O. F. T., "The Theoretical Scattering of Smoke in a Turbulent Atmosphere." *Proc. roy. Soc.*, (A) 104: 640–654 (1923).
71. Schmidt, W., "Turbulente Ausbreitung eines Stromes erhitzter Luft." *Z. angew. Math. Mech.*, 21: 265–278, 351–363 (1941).
72. Sherlock, R. H., "Analyzing Winds for Frequency and Duration." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
73. —— and Stalker, E. A., "The Control of Gases in the Wake of Smokestacks." *Mech. Engr.*, 62: 455–458 (1940).
74. —— "A Study of Flow Phenomena in the Wake of Smokestacks." *Bull. Dept. Engng. Res. Univ. Mich.*, No. 29 (1941).
75. Sherwood, T. K., "The Geometry of Smoke Screens." *J. Meteor.*, 6: 416–419 (1949).
76. Smith, M. E., "The Forecasting of Micrometeorological Variables." *Meteor. Monogr.*, Vol. 1, No. 4 (1951).
77. Sproull, W. T., *Removal of Particulate Matter from Industrial Effluent*. Proc. First Nat. Air Pollut. Sym., pp. 80–85, Stanford Res. Inst., Stanford, Cal., 1949.
78. Steffens, C., "Measurement of Visibility by Photographic Photometry." *Industr. engng. Chem.*, 41: 2396–2399 (1949).
79. —— and Rubin, S., *Visibility and Air Pollution*. Proc. First Nat. Air Pollut. Sym., pp. 103–108, Stanford Res. Inst., Stanford, Cal., 1949.
80. Stern, A. C., "Atmospheric Pollution Through Odor." *Heat. & Ventilating*, 41: 53–57 (1944).
81. —— "Atmospheric Pollution Due to Gas." *Heat. & Ventilating*, 42: 64–66, 73–74, 117, 122 (1945).
82. Sutton, O. G., "A Theory of Eddy Diffusion in the Atmosphere." *Proc. roy. Soc.*, (A) 135: 143–165 (1932).
83. —— "The Problem of Diffusion in the Lower Atmosphere." *Quart. J. R. meteor. Soc.*, 73: 257–281 (1947).
84. —— "The Theoretical Distribution of Airborne Pollution from Factory Chimneys." *Quart. J. R. meteor. Soc.*, 73: 426–436 (1947).
85. ——"The Dispersion of Hot Gases in the Atmosphere." *J. Meteor.*, 7: 307–312 (1950).
86. Swain, R. E., "Smoke and Fume Investigations. A Historical Review." *Industr. engng. Chem.*, 41: 2384–2388 (1949).
87. Taylor, E. G. R., "Problems of Fog and Air Pollution." *Smokeless Air*, 12: 96–100 (1942).
88. Teverovsky, N., "Coagulation of Aerosol Particles in a Turbulent Atmosphere." *Izvestiia Akad. Nauk, SSSR, Ser. Geogr. i Geofiz.*, No. 1 (1948).
89. Thomas, M. D., "Automatic Apparatus for the Determination of Small Concentrations of Sulfur Dioxide in Air. III." *Industr. engng. Chem.*, Analyt. Ed., 4: 253–256 (1932).
90. —— and others, "Automatic Apparatus for Determination of Small Concentrations of Sulfur Dioxide in Air." *Industr. engng. Chem.*, Analyt. Ed., 15: 287–290 (1943).
91. Thomas, M. D., Ivie, J. O., and Fitt, T. C., "Automatic Apparatus for Determination of Small Concentrations of Sulfur Dioxide in Air." *Industr. engng. Chem.*, Analyt. Ed., 18: 383–387 (1946).
92. Thomas, M. D., Hill, G. R., and Abersold, J. N., "Dispersion of Gases from Tall Stacks." *Industr. engng. Chem.*, 41: 2409–2417 (1949).
93. *United States Technical Conference on Air Pollution*. New York, McGraw, 1951.
94. Vonnegut, B., *A Continuous Recording Condensation Nuclei Meter*. Proc. First Nat. Air Pollut. Sym., pp. 36–44, Stanford Res. Inst., Stanford, Cal. 1949.
95. Wait, G. R., "Some Experiments Relating to the Electrical Conductivity of the Lower Atmosphere." *J. Wash. Acad. Sci.*, 36: 321–343 (1946).
96. Washburn, H. W., and Austin, R. R., *Some Instrumentation Problems in the Analysis of the Atmosphere*. Proc. First Nat. Air Pollut. Sym., pp. 69–76, Stanford Res. Inst., Stanford, Cal., 1949.
97. Went, F. W., *The General Problem of Air Pollution and Plants*. Proc. First Nat. Air Pollut. Sym., pp. 148–149, Stanford Res. Inst., Stanford, Cal. 1949.
98. Western Oil and Gas Association, Committee on Smoke and Fumes, *The Smog Problem in Los Angeles County*. Second Interim Rep. by Stanford Res. Inst., pp. 1–64, 1949.
99. Wexler, H., and others, "Air Pollution in Donora, Pa.: Preliminary Report." *Publ. Hlth. Bull. Wash.*, No. 306 (1949).
100. Whipple, F. J. W., "The Influence of Urban Conditions on the Circulation of Electricity Through the Atmosphere." *Trans. Faraday Soc.*, 32: 1203–1209 (1936).
101. Whytlaw-Gray, R., "Disperse Systems in Gases." *Trans. Faraday Soc.*, 32: 1042–1047 (1936).
102. —— and Patterson, H. S., *Smoke: A Study of Aerial Disperse Systems*. London, E. Arnold & Co., 1932.
103. Wiegel, H., "Niederschlagsverhältnisse und Luftverunreinigung des Rheinisch—Westfälischen Industriegebietes und seiner Umgebung." *Veröff. meteor. Inst. Univ. Berlin*, Bd. 3, Nr. 3, 52 SS. (1938).
104. Willett, H. C., "Meteorology as a Factor in Air Pollution." *Industr. Med. Surg.*, 19: 116–120 (1950).
105. Wilsdon, B. H., and McConnell, F. J., "The Measurement of Atmospheric Sulfur Pollution by Means of Lead Peroxide." *J. Soc. chem. Ind., Trans. and Comm.*, 53: 385–388 (1934).
106. Zimmerman, P. W., *Impurities in the Air and Their Influence on Plant Life*. Proc. First Nat. Air Pollut. Sym., pp. 135–141, Stanford Res. Inst., Stanford, Cal., 1949.

CLOUDS, FOG, AND AIRCRAFT ICING

The Classification of Cloud Forms *by Wallace E. Howell* 1161

The Use of Clouds in Forecasting *by Charles F. Brooks* 1167

Fog *by Joseph J. George* 1179

Physical and Operational Aspects of Aircraft Icing *by Lewis A. Rodert* 1190

Meteorological Aspects of Aircraft Icing *by William Lewis* 1197

THE CLASSIFICATION OF CLOUD FORMS

By WALLACE E. HOWELL

Blue Hill Meteorological Observatory, Harvard University

Introduction

If clouds occurred in a naturally limited number of forms, like regular polyhedrons, the problem of classifying them would be simple. Such, however, is not the case. Clouds occur in an infinite variety of forms, and the classifications that have been devised to describe them bear the stamp of human and fallible notions about their appearance, origin, and structure. Order in clouds is the product of man's will, and man's will can change it.

In its present status, cloud classification comprises not only classifications now in use, but also notions and concepts abandoned in the past that are still valid and may be given new value in the future. In taking stock of our present position, therefore, valid notions from the past must not be overlooked. A review of the present status may properly take the form of an historical summary of selected developments.

Assessment of our present position has two aspects, logical and pragmatic. How completely and logically has nature been described? How suitable are the classifications for the uses to which they are put, and wherein lie their shortcomings? Thoughtful answers to these questions will point the way toward future progress.

The encoding of cloud forms for purposes of transmission or recording of observations must not be confused with classification. Encoding is a sequel to classification; once classes are established, a code may be devised by assigning to each class a number, or, if there are not enough numbers to go around, by grouping two or more classes under a single number. However, classification should never be artificially restricted to meet the demands of a numerical code. Natural phenomena, fingers and toes excepted, do not occur by tens, and the attempt to force them into such a mold will usually be wrong. In the present work the problems of encoding are not treated.

The Present Position

The classification of cloud forms has been developed from two main points of view. One, beginning in the early 19th century, concentrated on the appearance of clouds as seen from the ground; its latest stage is the international classification of cloud forms. The other viewpoint stressed the processes that form clouds. It began in the early 20th century, and is still in a rudimentary stage. No general agreement has been reached and no international codification attempted. In the following paragraphs an attempt will be made to point out the accomplishments from each viewpoint that appear to have value now and for the future.

A beginning has been made towards what may eventually become a classification of clouds based on their physical constitution, *i.e.*, drop-size spectrum, liquid-water content, form and size of ice crystals, etc. The factual basis for this approach is still far from complete, and few classifications based on it have yet been formulated.

The International Classification and Its Antecedents. The first classification that is still vital was that proposed by Luke Howard [11] in 1803. He recognized three simple modifications, which he named *cirrus*, *cumulus*, and *stratus*. The definitions he gave would pass as current today, except for minor details. He next defined two intermediate forms, *cirro-cumulus* and *cirro-stratus*, and two combined forms, *cumulo-stratus* and *cirro-cumulo-stratus* or *nimbus*. His intermediate forms survive today as the middle clouds and the cirrocumulus and cirrostratus, but his combined forms have been largely lost except for surviving traces of Howard's theory that rain, both of frontal and thunderstorm origin, is caused by the electrical interaction of two cloud layers. His definition of the nimbus therefore embraced both cumulonimbus and nimbostratus. It is worth noting that Howard did not call his system a classification nor claim that it embraced all clouds, but intended it only "to impose names on such of them as are worthy of notice."

Howard's work became the standard reference; but the authors of general meteorological texts, which began to appear in the middle 19th century, often paraphrased it carelessly and erroneously. It was through careless copying of Howard's definitions by various authors that the stratus gradually came to be described as "a cloud above the ground," whereas Howard clearly applied the term to fog lying on the ground as well as to low layer clouds. The only concrete contribution during that period was made by Kaemtz, who differentiated stratocumulus from stratus, recognizing only the cumulonimbus half of Howard's nimbus and defining the former broadly enough to include most rain clouds.

The initial success of the French meteorological bureau in the late 1850's stimulated the study of clouds as well as other branches of meteorology. It opened a period when many new ideas were advanced, a fertile if somewhat febrile period marked by originality in the invention of new classifications more than by perfection and consolidation of existing ones. Noteworthy because of its later influence was the system of Poëy [16] who regarded cirrus and stratus as the only fundamental cloud forms, the one being of ice and the other of water. He anticipated Bergeron [3] by ascribing the origin of rain to the coexistence of these two clouds which, when coexisting, he called "pallium." Clayton [7] suggested stratus and cumulus as the two fundamental forms, a

notion that has persisted strongly to the present day in many popular works on meteorology although it never was given any official recognition internationally.

The confusion resulting from a variety of irreconcilable classifications led, late in the nineteenth century, to steps towards international agreement. A classification into principal cloud types, devised by Abercromby and Hildebrandsson, illustrated in a cloud atlas by Hildebrandsson, Köppen and Neumayer in 1890, was recommended for international use by the International Meteorological Congress at Munich the following year. It became the basis for the first International Cloud Atlas [10]. The ten principal cloud types recognized were cirrus, cirro-stratus, cirro-cumulus, alto-cumulus, alto-stratus, strato-cumulus, nimbus, cumulus, cumulonimbus, and stratus, in that order.

Almost immediately after the appearance of this classification came proposals for a more detailed one following the same framework. Although successive editions of the International Atlas appeared in 1905 and 1910 without significant changes, official recognition of terms that were coming into widespread use could not be indefinitely postponed. In 1922 an international commission for the study of clouds was formed to rewrite the Atlas.

Two new guiding principles were adopted by this group. The first was acceptance of the formal framework of classification, first laid down by Linnaeus, which had become standard in many fields of science. This meant the division of *all forms* of cloud into families and their further subdivision into genera, subgenera, species, and varieties. The second principle was one suggested by Shaw, who noted that several of the forms were defined in terms of separate "unit clouds" and that several others were in terms of aggregations of these unit clouds.

The work of the commission resulted in the publication in 1929 of a revised International Cloud Atlas [1]. The new Atlas established four families on the basis of cloud height: high, middle, and low clouds, and clouds of vertical development. Within these four families the ten previously recognized forms were distributed as genera, stratus being moved from tenth place to seventh so as to place it in the correct family. Following Shaw's suggestion, three "forms" of clouds were recognized: isolated clouds; sheets of clouds divided into filaments, scales, or rounded masses; and more or less continuous sheets. These forms, however, were not given formal standing as families or genera. A number of subgenera, species, and varieties were defined on the basis of what was by then current usage acquired gradually during the preceding half century without any great changes being made. The latest edition of the Atlas [13], appearing in 1932, did not change any of the important concepts of its predecessor.

A different set of fundamental forms was suggested by Bergeron [4]: the cumuliform (*Cu, Sc vesp, Cb, Ac cast*), the wave-form (*St, Sc, Ac, Cc*), and the stratiform (*Cs, As, Ns*).[1]

1. An explanation of abbreviations is given in [13].

Genetical Classifications. The rapid advances made during the early years of the 20th century in the study of the upper air led to an increasing appreciation of the role of cloud forms as indicators of atmospheric processes. This gave rise to progress along the second principal road in the classification of clouds, namely what have been called *genetical classifications.* Because there has been no international codification of genetical classifications, meteorologists are less likely to be familiar with all the principal ones that have been proposed, and this article will therefore describe them in some detail.

One of the first of the genetical classifications was proposed by J. Bjerknes and Douglas in a memorandum that was later amplified by Douglas [9]. It divides clouds into four types on the basis of "the physical processes involved in the formation of clouds." These are:

A. Cloud systems due to slow upward motion over a large area, usually associated with continuous precipitation (*Ns, As, Cs,* and some forms of *Ci*). The main feature is the great horizontal and vertical extent of the cloud system.

B. The cumulus and cumulonimbus group (including *Ci* or broken *As* formed from anvils), due to smaller air masses rising through their environment. Of great interest in this group are the layers of stratocumulus or altocumulus which develop turreted tops and eventually may grow into cumulonimbus.

C. The clouds due to turbulent motion, which may be either irregular (*Fs*) or arranged in definite layers (*St, Sc, Ac, Cc*). These eddy sheet clouds result from unstable motion within the cloudy air. Genuine stratus clouds are due (according to Douglas) to cooling of the air at the earth's surface combined with air movement which sets up turbulence and raises the cloud off the ground.

D. Lenticular clouds and cloud patches of smooth appearance, indicating local ascent of a damp stratified layer.

To this classification Douglas appends the note that *Ci* clouds could not be completely classified from a physical point of view at that time.

At about the same time Stüve proposed a much more detailed subdivision of cloud forms in a genetical classification that he revised for a later publication in 1937 [18]. An outline of his classification follows:

A. Clouds formed outside the space they occupy.
   1. Fluid or solid (amorphous or crystalline): *virga.*
   2. Crystals: *Ci, Ci unc.*

B. Clouds proper to the space they occupy.
   1. Clouds of convection. Cumuliform clouds.
      a. Formed through thermal convection.
         (1) Having no connection with a surface of discontinuity; their bases determined by the condensation level of air rising from the ground and therefore horizontal; the tops perhaps governed by an inversion: *Cu.* With strong development resulting from marked instability: *Cb.*
         (2) Formed from *Sc* or *Ac* through lifting of

the layer until latent convective instability is released: *cast.*

b. Formed through dynamic convection. Always lying beneath a surface of subsidence, their bases parallel to it.

(1) Fluid. Distinguished according to height: *Sc*, *Ac.*

(2) Solid: *Cc.*

2. Clouds of advection.

a. Formed by active upglide over an inversion.

(1) Without precipitation: *As.*

(2) With virga: *As* with *virga.*

(3) With precipitation reaching the ground: *Ns.*

b. Formed through passive upglide over an inversion.

(1) Without precipitation: *mam.*

(2) With precipitation: *Cb.* (In its upper portion this becomes a cloud of thermal convection.)

c. Formed through upglide over a free upglide surface (*i.e.*, one not resting on the ground): *St mam.*

d. Through upglide beneath an upglide surface alone no clouds are formed, but rather this upglide favors the formation of forms under B, 1, *b*, and otherwise leads with the help of thermal and dynamic convection to the formation of *St* and *fog* beneath a surface of subsidence, if this bounds a homogeneous air mass that lies directly on the ground.

3. Clouds of lifting. Whenever a maximum of relative humidity is brought into being beneath an inversion by dynamic convection, lifting of the whole air mass can produce condensation there.

a. Water.

(1) Lifting above upthrusting *Cu: pil.*

(2) Lifting through advection at a lower level: *lent* (and also, according to the preponderance of eddy diffusion before and during the lifting, either more stratiform or more cumuliform).

b. Ice.

(1) Lifting above upthrusting *Cb: Ci pil.*

(2) Lifting of the upper layers through upglide of an air mass over an upglide surface: *Cs*, *Ci.*

4. Clouds of turbulence. They arise through eddy diffusion in a manner similar to the forms produced through dynamic convection. In this case, however, the eddy diffusion originates through friction with the ground.

a. Formation made possible through evaporation from precipitation falling into the zone of turbulence: *Fs.*

b. Formation made possible through a nearly adiabatic lapse rate: *Fc*, *Cu.*

C. Orographic clouds. Landforms and the conditions of ground friction they establish do not lead to new forms of cloud, but localize the regions of their formation.

1. Cloud formation because of mountains.

a. More or less homogeneous clouds that cling to mountains, mostly on the windward side.

b. Banner clouds, forming in the lee.

2. Coastal cloudiness resulting from changed friction.

3. Localization of cloud formation through the differing thermal properties of the earth's surface.

A genetical classification introduced by Petterssen [15] approaches the problem from the point of view of air-mass properties. His classification of four categories follows:

A. Internal clouds that characterize the stability conditions of the air masses to which they belong.

1. Clouds that form in unstable air masses: *Cu hum*, *Cu con*, *Cb cal*, *Cb inc*, *Cb cap*, *Cb mam*, and certain detached high clouds such as *Ci den* which originate from the anvils of *Cb.*

2. Clouds that form in stable air masses: *St* and *fog.*

B. Clouds that characterize the discontinuities within or between air masses.

3. Clouds that form in connection with quasi-horizontal inversions. These are most frequently related to category 2: *Sc* and certain types of *Ac.* A great variety of clouds belong to category 3 but show signs of instability and are therefore related to category 1: *Sc vesp*, *Sc cast*, *Sc cug*, *Cu und* and certain types of *Ac cast.*

4. Frontal clouds that are formed because of upglide movement along frontal surfaces. When the air is stable these are of a stratiform type: *As.* When it is completely unstable they may be of the cumulus type: *Cb arc.* When it is conditionally unstable, the lower part of the cloud system may be stratified, and *Cu* or *Cb* towers project from the upper part of the frontal cloud layer.

Although not generally called such, the international classification of states of the sky is actually a *genetical classification* which takes as its basis three main origins of clouds: in cyclones, or in particular extratropical cyclones; in regions of convection or thunderstorm activity; and in regions of turbulence. The cyclonic cloud systems are treated in greatest detail, being divided into emissary, front, central, lateral, rear, and connecting zones. In the thundery systems, pre-thundery, thundery, and rear zones are recognized. The details of the classification are published in standard references [13, 5], and need not be repeated here. It may be pointed out, however, that the codes for states of the sky, C_L, C_M, C_H, are designed specifically to describe the most significant aspects of nephsystems as they are manifest in the sky.

A genetical classification put forward by Kähler [14] corresponds quite closely to the principal divisions of Stüve's classification, adding but one new notion. This is the separate recognition of clouds formed through mixture of warm, moist air and cold air at an interface between them. The resulting cloud forms are described as being weak and of the *As* type.

Physical Classification. The only classification that has so far been advanced from the point of view of the

physical constitution of clouds is that proposed by Bergeron [2] on the basis of his investigation of colloidal instability of clouds. His classification is made in tabular form as follows:

Group	Cloud element	Characteristic precipitation	Cumuliform cloud	Stratiform cloud
I	Complete crystals	Ice needles	*Cc* with halo	*Cs neb*,
II	Complete crystals and skeletons	Dry powder snow	*Ci unc*	*Cs fil*, *Cc* without halo
III	Skeletons and fog droplets	Snowflakes and/or rime-graupel or frost-graupel	*Cb* (winter)	*As*
IV	Fog droplets	(Dry fog)	*Cu*	*St*, *Sc*, *Ac*
V	Drizzle droplets	Wet fog, drizzle	*Cu*	*St*, *Sc*, *Ac*
VI	Raindrops	Rain or hail	*Cb* (summer)	*Ns*

According to Bergeron, Groups I and IV (only complete crystals or only droplets) are colloidally stable, Groups II and V exhibit slight colloidal instability, and Groups III and VI are wholly unstable.

Assessment of Adequacies and Inadequacies

The international classification of clouds has behind it nearly a century and a half of development, and a half a century of experience has been gained with it in essentially its present form. Many great meteorologists have devoted their attention to perfecting it. It would seem to represent a high degree of perfection and refinement. However, there are serious criticisms that can be leveled against it. Intended to permit accurate classification of any observed cloud form, its genera are not mutually exclusive and many cloud forms fit none of the genera defined. Intended to bring about uniform usage, observers still fail by a wide margin to agree on the classification of common forms of cloud.

Two examples will serve to illustrate this unsatisfactory status. In 1942 the author showed a set of color pictures of clouds before a group of meteorologists all of whom had had at least three years of professional training in synoptic meteorology. The slides were selected to illustrate common states of the sky and no attempt was made to select either the most typical appearances of the cloud genera or the most ambiguous. Every person in the group was asked to write down, as each slide was shown, the genus of the principal cloud form illustrated. When the results were tabulated it was found that only 56 per cent of the group, on the average, agreed on the principal genus, 27 per cent preferring a second genus and the rest of the group scattering their selections among as many as six additional genera.

It may be argued that the group test showed only that the training in cloud recognition of the subjects was insufficient and nonuniform, and indeed this may have been the case. Nevertheless, if the classification is so ineffectual in the hands of meteorologists, regardless of its intrinsic quality it fails in an important part of its intended purpose. It must be so fitted to the demands that meteorologists will find it easy and natural to use accurately.

The second example is the result of a review the author made of more than thirty books appearing between 1938 and 1942, comprising both standard meteorological texts and training materials on meteorology intended for aviators. Nearly half of the books take no notice of the 1932 International Cloud Atlas, but retain the terms and definitions of the earlier editions. A sizeable number mention cirrus, stratus, cumulus, and nimbus as the four fundamental cloud forms—a notion that has no basis in the international classification or its antecedents. A few reduce the number of basic forms to two, cumuliform and stratiform, disregarding the occurrence of cirrus which conforms to neither. Another book explains that "nimbus" means "head," and that therefore a nimbostratus is a stratus with a head on it! Nor are these examples of divergences by any means all drawn from nonmeteorologist authors.

The day is past when meteorological terms were practically the private property of professional meteorologists. The airmen who use meteorological terms now outnumber the meteorologists, and their use of the word "stratus" to describe any low cloud that forms an extensive low ceiling, or "cumulus" to denote the most spectacular thunderhead, is an example of word usage that we cannot ignore.

At this point let us review the nature of a classification as such. A classification as conceived by Linnaeus corresponds to the geometric concept of a partition; that is, a whole that is divided into a finite number of mutually exclusive parts, the sum of all the parts constituting the whole. Application of this concept to the classification of clouds would require that every conceivable form of cloud be allocable to one and only one family; within that family to one and only one genus, species, etc. The problem of defining the classes is that of drawing the lines between them.

An attempt to subject the international classification to rigorous tests of logic meets with difficulty. For example, the definition of cirrus gives only three characteristics as belonging unequivocally to this genus: that they are detached clouds, have a delicate and fibrous appearance, and are without shading. No other genus is open to detached clouds of delicate and fibrous appearance. What is to be said, then, of a detached, delicate, fibrous cloud that is thick enough to show shading? Plate 6 of the International Atlas shows just such a cloud and the explanatory text even calls attention to the shading. Yet it is classified as cirrus. The example would be trivial if it were unique; it is the multiplicity of such ambiguities that makes the classification open to grave criticism. No matter how strictly the definitions are made mutually exclusive, the human element of observation necessarily blurs the boundary between them in actual use. No effort should be spared to eliminate indefiniteness from the definitions themselves.

Brooks [6], taking cognizance of the same problem, questions whether it is not too much to expect that the definitions of the genera could be made completely in-

clusive and exclusive without overloading them unduly. To some extent this is a matter of degree, as to how lengthy and detailed a definition may be without being so clumsy as to defeat its purpose. But it is also one of purpose, whether the definition is framed to describe a typical form or to delimit a class. A very suggestive result is obtained if one strips from the international definitions of the genera all characteristics that do not explicitly serve to distinguish one genus from another. The remainders are:

Cirrus: a detached high cloud that shows no shading.

Cirrocumulus: a layer (or groups) of cloud masses in regular arrangement, the elements of which show no shading.

Cirrostratus: a continuous sheet of high cloud not showing relief and not thick enough to blur the outlines of the sun or moon.

Altocumulus: a layer (or groups) of cloud masses in regular arrangement, or a continuous layer showing relief on its lower side, the smallest elements of which have a diameter less than ten solar diameters, and the thickest elements of which show shading.

Stratocumulus: a layer (or groups) of cloud masses in regular arrangement, or a continuous cloud sheet showing relief in its lower side, the smallest elements of which have a diameter of more than ten solar diameters.

Stratus: a uniform layer of low cloud, from which, if precipitation falls, it falls as drizzle.

Nimbostratus: a nearly uniform layer of low cloud, from which, if precipitation falls, it falls as continuous rain or snow.

Cumulus: a cloud of vertical development whose upper portion is not fibrous in appearance.

Cumulonimbus: a cloud of vertical development whose upper portion is fibrous in appearance.

The altered character of these definitions is at once apparent. They are less vividly descriptive than the international definitions, but emphasize more clearly the elements that distinguish one genus from another.

Genetical Classifications. The present situation of the genetical classifications somewhat resembles that of the classifications based on appearance shortly before the first efforts at international agreement. In the several classifications, major processes are universally recognized: the overturning of air by through-going thermal convection, the stirring that characterizes turbulent layers, and the widespread lifting that accompanies horizontal convergence of the lower air layers in the forward portion of a cyclone. The manner in which the processes are arranged and subdivided, the relative weight given to different manifestations of a process, the number of additional processes thought worthy of mention, and to some extent, the nomenclature of the resulting clouds are matters on which full agreement has not yet been reached. Some aspects do not come clearly to the fore in any of the classifications, notably the difference between forced turbulence and thermal circulation in sheets of air, the manifestations of shearing motion, and the part played by radiation. The descriptions of cloud forms resulting from the various processes are somewhat incomplete, especially with regard to the dissipative stages in the life of the clouds.

Physical Classifications. Bergeron's classification of clouds according to their constitution, whether composed of complete crystals, skeletal crystals, or water droplets, covers only a single aspect of cloud constitution, namely colloidal instability. Before a comprehensive physical classification can begin to take shape, the collection and assimilation of a more adequate factual basis is necessary. A start in this direction has been made by Nakaya[2] with regard to the laws that govern the crystalline structure of ice crystals and snow, and by Diem [8] and others with regard to the drop sizes characteristic of the various cloud forms. Information about the drop-size distribution and its origin [12] should also find application here.

A View into the Future

It is of course easier to criticize than to improve. Nevertheless, it is possible to foresee and outline certain steps that will contribute toward progress in the field of cloud classification.

Classification According to Form. One of the most pressing needs is an approach to a more rigorous logical treatment of the international definitions, making each category of the classification (family, genus, species, etc.) as nearly as possible a geometrical partition. The problem presents serious difficulties if the definitions are not to become cumbersome. It can be simplified, however, if the definitions, particularly those of the genera, are stripped of all provisional statements (*i.e.*, those including "sometimes" or "usually," etc.) and all statements that do not contribute to rigorous distinctions between classes. A further simplification may be achieved by abandoning the arbitrary restriction of the number of genera to ten and defining additional genera. For example, if detached lenticular clouds were given status as a genus, the definition of altocumulus (or other genera) would not have to be stretched to include these clouds.

The effect of eliminating nondefining elements from the definitions of the genera would be to decrease the emphasis now placed on the genus category and increase the importance of the species category. Such a change conforms with the increased emphasis placed on species in the formulation of the codes for states of the sky.

Toward Uniform Practice. As long as the materials and practices used in the training of meteorologists, meteorological observers, and other users of meteorological information remain as divergent as they are today, it will be very difficult to attain uniformity in the use of the international classification. The publication of an authoritative atlas and its distribution to libraries and similar repositories is not adequate as a means of achieving even a reasonable degree of uniformity.

The situation would be helped by the republication of an official abridgement of the International Atlas in a form suitable for wide distribution, so that it would find its way more generally into the libraries of weather-

2. Consult "The Formation of Ice Crystals" by U. Nakaya, pp. 207–220 in this Compendium.

conscious people. With the collaboration of the directors of national meteorological services, it might be possible for such an atlas to be distributed to every weather station in place of the several different publications now so distributed.

Another pressing need is for adequate, uniform training materials and outlines of study. At present, cloud classification is taught not from the International Atlas, which is too bulky and expensive for use in the classroom or for purchase by individual students, but from textbooks that usually accompany brief extracts from the international definitions with introductory and explanatory material that may differ substantially from the Atlas. Training material and courses should be designed to meet the individual needs of introductory, intermediate, and advanced stages in the training of meteorologists and meteorological observers, and appropriate special training material should be provided for the use of student aviators and other students of applied meteorology in its many branches. In order to spread uniform training as widely as possible, the publications should be realistically designed for inexpensive production. Film strips and if possible time-lapse motion pictures should be prepared as auxiliary teaching aids and given the widest possible distribution.

Genetical Classifications. Although the international classification of states of the sky (nephsystems, as distinguished from the C_L, C_M, C_H code) has not gained widespread use in synoptic meteorology, the need for an internationally recognized genetical classification is growing, rather than diminishing. The time is near when our body of information about the relation of cloud forms to the atmospheric processes that form them will be adequate for the purpose and a definitive genetical classification will be achieved.

In the steps toward a comprehensive genetical classification, it will be necessary to broaden the basis of nephanalysis and make the connection between the structure of nephsystems and cyclones more specific. The notion of an idealized cyclone that moves steadily across the country without changing its structure has given way to the concept of a cyclone that passes through a definite life cycle of wave development and occlusion. The development of the nephsystem should be related to that of the cyclone and the connection between them should be clarified by an analysis of the atmospheric processes that are common to both.

A Unifying Principle. The form and appearance of clouds, which is the basis for the international classification, is the product of the genetical processes that formed the clouds, and is influenced by their physical constitution. This forms a compelling basis for interrelation between the different kinds of classification so far discussed. Since the international definitions have intentionally avoided mention of genetical processes as much as possible and refer only to aspects of physical constitution that are visibly apparent, recognition of the interrelations have mostly taken the form of references in the genetical and physical classifications to corresponding international cloud types.

There is nevertheless an inviting possibility that the different classifications may benefit mutually through the development of closer correspondences. There are already many instances of almost perfect correspondence between a cloud form as internationally defined and the genetical class into which it falls. For example, cirrocumulus is associated with only a single subdivision of Stüve's classification, and it is the only form that belongs in that subdivision. The establishment of more correspondences of this sort in the course of revising the international classification and framing an authoritative genetical classification should materially assist in the improvement of both classifications. It provides a powerful principle for eliminating the omissions and ambiguities now troubling us and for broadening the field of general agreement in all aspects of cloud classification.

REFERENCES

1. *Atlas international provisoire des nuages.* Paris, Office National Météorologique, 1929.
2. Bergeron, T., "Über die dreidimensional verknüpfende Wetteranalyse, I." *Geofys. Publ.*, Vol. 5, No. 6 (1928). (See pp. 29 ff.)
3. —— *On the Physics of Clouds.* Memo, Meteor. Assoc., Intern. Union for Geodesy and Geophysics, Paris, 1933.
4. —— *Vorträge über Wolken und praktische Kartenanalyse.* Translated from the German ms. by S. P. Chromov. Moscow, Central Hydrometeorological Service, 1934.
5. Berry, F. A., Jr., Bollay, E. and Beers, N. R., ed., *Handbook of Meteorology.* New York, McGraw, 1945. (See pp. 893-901)
6. Brooks, C. F., "International Cloud-Nomenclature and Coding." *Trans. Amer. geophys. Un.*, Part III: 494-502 (1944).
7. Clayton, H. H., "Diurnal Cloud and Wind Periods at Blue Hill Observatory during 1887." *Amer. meteor. J.*, 5: 321-332 (1888).
8. Diem, M., "Messungen der Grösse von Wolkenelementen II." *Meteor. Rdsch.*, 2: 261-273 (1949).
9. Douglas, C. K. M., "The Physical Processes of Cloud Formation." *Quart. J. R. meteor. Soc.*, 60: 333-341 (1934).
10. Hildebrandsson, H., Riggenbach, A., et Teisserenc de Bort, L., *Atlas international des nuages.* Paris, Comité Météorologique International, 1896.
11. Howard, L., "On the Modifications of Clouds, and on the Principles of Their Production, Suspension, and Destruction." *Phil. Mag.*, 16: 97-107, 344-357 (1803); 17: 5-11 (1803).
12. Howell, W. E., Wexler, R., and Braun, S., "A Theory for the Drop Size Distribution in Clouds," Part One (II) in *Contributions to the Theory of the Constitution of Clouds*, pp. 6-18. Mount Washington Observatory Res. Rep., Final Report, year ending Oct. 20, 1949.
13. *International Atlas of Clouds and States of the Sky.* Paris, Office National Météorologique, 1932.
14. Kähler, K., *Wolken und Gewitter.* Leipzig, J. A. Barth, 1940. (See pp. 97 ff.)
15. Petterssen, S., *Weather Analysis and Forecasting.* New York, McGraw, 1940. (See pp. 35-36)
16. Poëy, A., *Nouvelle classification des nuages suivie d'instructions pour servir à l'observation des nuages et courants atmosphériques.* Paris, Dépot des Cartes et Plans de la Marine, No. 513, 1873.
17. Schereschewsky, P. L., et Wehrlé, P., "Les systèmes nuageux." *Mém. off. nat. météor., Paris*, No. 1 (1923).
18. Stüve, G., "Thermodynamik der Atmosphäre," *Handbuch der Geophysik*, Bd. IX, Lief. 2. Berlin, Gebr. Borntträger, 1937. (See pp. 312-314)

THE USE OF CLOUDS IN FORECASTING

By CHARLES F. BROOKS

Blue Hill Meteorological Observatory, Harvard University

Introduction

A storm may be described as precipitation surrounded by clouds and usually accompanied by wind. The essence of forecasting is in telling when precipitation will occur and when it will not. Since rain or snow does not occur without advance notice in the form of clouds, we conclude that clouds are a prime element in forecasting.

Weather proverbs from long ago [52, 55] indicate attempts at spot forecasting from cloud observations. The professional forecaster, while less dependent upon scanning the local sky, cannot neglect the synoptic reports of cloud systems and their motions and development. Cloud observations supplement the more direct and quantitative aerological data from balloons, airplanes, and mountain stations, which, moreover, are not always available.

The very presence of a cloud indicates condensation in the atmosphere and hence the occurrence of some cooling process. Cooling in the free air is most commonly caused by ascent and expansion of air; therefore, a cloud usually indicates ascending air. If the cloud is cumuliform, the ascent is localized; if it is stratiform, the ascent is general, probably from broad lifting, such as convergence on a large scale or the upglide of one air layer over another. The sharpness, density, and form also indicate lapse rates, vertical currents, wind shear, and turbulence. The progressive motion provides the wind velocity at or immediately below the cloud height. Better observations and greater attention to interpretation of clouds and their sequence are therefore urged in the interest of improved local and general forecasting.

For comprehensive discussions of aerological interpretations of clouds, the reader may refer to books by Stüve [93], Süring [95], and Kähler [56], to Brooks' systematic article [13], to Möller's review of the literature from 1939–1946 [71], and to relatively recent detailed treatments of special phases such as the studies of *Ci*[1] by Ludlam [62], the studies of *Sc* and *St* by Raethjen [78], Schwerdtfeger [89], and Neiburger [72], of convective phenomena by Bleeker and Andre [10], and of convective clouds in the tropics by Craddock [28–30] and Deppermann [35].

Cloud Observing

Cloud observing presents many problems: separating the levels, observing the forms, extents, and positions, and determining the heights and motions. Summaries of methods of determining cloud heights and motions have been published by Middleton [68], Süring [96, 97], and Eredia [43], and a brief evaluation of the accuracy of different methods of determining cloud heights, by the International Meteorological Organization [54]. Discussions of cloud observing have recently been prepared by Brooks [12], emphasizing the range-finder and dew-point methods for determining cloud heights, and by Fergusson [44], and Conover, Fergusson, and Brooks [26], emphasizing the nephoscope method of obtaining cloud motions. It appears that a coincidence-type of range finder with a base of at least 1½ m is by far the quickest means of getting cloud heights within an accuracy of 5 per cent when there are sharp, well-lighted clouds within 15 km in any part of the sky other than the zenith (where the ceilometer is better). However, the cloud-range meter, of radar type, described by Moles [70], may have some advantage over the range finder within its limit (7000 yd), since sharpness in cloud outline and illumination are not required. The dew-point method is good within 8 per cent, Clayton found [24], when care is exercised to apply it only to clouds forming by the convection of fair-sized parcels of air likely to have the same temperature and dew point as at the observer's station. Large deviations result under sea-breeze or morning-inversion conditions. Heights by the single-theodolite, pibal method were found by Rossi [82] to differ from double-theodolite heights by 5 per cent under stable conditions and by 16 per cent under labile ones; Berg [3] found them to differ from those by airplane by 4 per cent for *Ac* and *Sc*, 8 per cent for *Cu*, and 13–20 per cent for other clouds.

The quickest way of getting cloud motions is with a window-sill nephoscope consisting of a plane, black, horizontal mirror of eight to ten inches diameter, without markings on the mirror other than a depression at the center, and a peephole eyepiece through which the observer can watch the motion of the image of a cloud and follow it with a small marker. When followed for a standard period (or easy fraction or multiple), the direction and relative speed are determined with a single placement of a ruler [12, 26, 44].

Clouds in Short-Term Forecasting at a Single Station

Application of the Diurnal Cycle to Observed Clouds. Early on a calm, clear morning, the occurrence of low clouds moving rapidly from any direction, but particularly from northwest or north, is a good indication of a day with strong winds from the direction from which the clouds are moving. Diurnal convection will soon bring this wind down to the ground.

Early-forming, thermal-convection *Cu* or *Sc* on a winter morning give promise of a generally cloudy day with little rise in temperature, for thermal convection will increase as the sun becomes effective, and the cloud masses will spread under the inversion responsible for the flattening already observed.

1. The international abbreviations of cloud genera will be used throughout

Morning *St* from the south is usually a prognostic of a marked rise in temperature, for it is formed by the mixture of warm, moist air at a low height with colder, surface air, and the mixing layer will soon work through to the ground. This is true especially if there is sufficient insolation to warm the surface layer to a temperature high enough to engage it in turbulent or convective interchange with the warm layer above. Insolation thus favors daytime arrivals of warm fronts at the surface. Unless the advection is raising the dew point rapidly, the mixing through a deeper layer and the rise in temperature tend to reduce the cloudiness by mid- or late morning, changing it from *St* to *Sc* or even *Cu*.

Altocumulus clouds observed in the evening are more likely to last all night, and to increase, growing upward and downward, when the surface wind is south than when the wind is northwesterly. In the first case it is usually the southerly wind, with its transport of heat and moisture, that is responsible for their formation, while in the second, such clouds may be merely the leftovers of flattened tops of convectional clouds formed during the daytime convection from the surface. Radiational cooling from the top of either type, however, favors growth and continuance.

The occurrence of sharp-topped *Cu* or *Sc* at sunset, when the wind is westerly or northwesterly, is a prognostic of a marked fall in temperature, for the sharpness, due to strong convection, indicates the steep lapse rate which is characteristic of a fresh cold air mass. At night radiational cooling as well as advection strengthens cold air masses and, indeed, favors nocturnal arrivals of cold fronts.

Interpretation of Observed Trends in Cloud Transformations. An increase in cloudiness will flatten the diurnal variation of temperature.

A wind-shift line at the earth's surface is usually marked by some forced ascent of air; thus the appearance and approach of cloud forms due to forced ascent indicate coming change, in anywhere from a few minutes to a few hours, depending on the distance and speed. This change may often be accompanied by rain. If there is a well-defined cloud base, the height of which is estimated, successive measurements of the angular altitude of the base as it approaches will give a reasonably accurate estimate of the time of arrival of the wind shift.

On a dull day we expect clearing if it becomes lighter, and vice versa. It is chiefly a change in light which attracts attention. A pall of forest-fire smoke which darkens the sky often brings out the umbrellas. Indeed, a solar eclipse has brought in washing off the line!

A meteorologist can intelligently make up any number of qualitative rules for short-term forecasting at a single station [*cf.* 14; 45; 47, p. 213; 58; 101]. The foregoing remarks must suffice as samples.

The expectancy of *any* rain following the first appearance of various cloud forms was worked out by Clayton [23], apparently in more detail than by anyone else. Sweetland [98] added *St nimbiformis* (=*Ns*) and Palmer [75], halos.

Clayton's analysis of cloud sequences by layers before 135 rainstorms [23] showed that the pre-rain sequence of *Ci* or *Cs* to *As* to *Nb* or *Cb* occurred in 84 per cent of the cases. The average expectancy of rain at Blue Hill, including cases when rain was falling at the zero hour, was 43 per cent in twenty-four hours; but when, as in the cases in Table I, precipitation was not already in progress, the average expectancy was 36 per cent. Only in the case of *Cs* with halos, *As*, and *Nb*(=*Ns*) is the expectancy substantially above average.

Table I. Expectancy of Rain at Blue Hill, Milton, Mass., Following Various Cloud Forms

Genus	Cases	Per cent of cases in which rain follows within 24 hr	Most frequent interval to precipitation† (hr)
Ci	159	33	26
Cc	137	36	22
Cs	160	44	13
Ac	84	45	8
As	142	68	6
*Ns**	70	73	4‡
Halos	569	<57	16‡

* *St nimbiformis.*

† Cases when precipitation began at an unknown hour during the night have been omitted; intervals are therefore somewhat weighted toward daytime conditions. Since rain occurs more readily at night, the intervals would probably have been somewhat less without these exclusions.

‡ Average interval.

Halos Before Rain. The Zuñi Indians say, regarding halos, "When the sun is in his house, it will rain soon." Several tabulations of the occurrence of rain after halo-producing *Cs* are summarized in Table II.

In most of the studies, solar and lunar halos are separated, but the differences are mostly small—frequency of precipitation is 1 per cent greater after lunar halos at Wauseon and 5 per cent greater at State College, but 9 per cent less at Blue Hill and 10 per cent less in London.

Differences between summer and winter are given by Mikesell [57], Neuberger [73], Palmer [75], and Russell [83], the precipitation being decidedly less in summer: Wauseon, 54 per cent vs. 62 per cent, and State College, 42 per cent vs. 78 per cent at twenty-four hours; and Blue Hill, 61 per cent vs. 70 per cent, at thirty-six hours. In London, the frequency in December is as high as 83 per cent (vs. 68 per cent for the year) at twenty-four hours.

Mikesell [57] found that the prognostic value of halos at Wauseon varied markedly with the pressure and pressure tendency (see Table III). The marked concentration of halos with high but falling pressure is in line with the typical position of *Cs* on the back slope of a high.

A few other points from studies of halos and *Cs* follow. Martin [65, 66] reports that at Fort Worth a halo with an east wind is followed by precipitation within thirty-six hours in 87 per cent of the cases, and within forty-eight hours in 95 per cent of the cases. At Columbus, when the wind at the time of the halo or soon after is from the southwest, the precipitation prospect in forty-eight hours is 88 per cent, according to

Martin [65]. Whittier [101, p. 4] urges that more reliance be placed on halos as prognostics of precipitation when they end owing to the thickening of the *Cs* into *As*. Reeder [79] cites the brilliant halo of Sept. 27, 1906, at Columbia, Mo., as the first indication of a tropical cyclone moving northward in the Gulf of Mexico. Dole [39] has cited colored sunsets at a coastal station as a valuable clue to the existence of a tropical storm. According to Neuberger [73, 74], the prognostic value of noncircular halos is greater than that of circular ones, being 85 vs. 77 per cent in Pennsylvania and 73 vs. 68 per cent in northwest Germany; bright halos are better than moderate or weak ones: 72 vs. 66 or 61 per cent in Pennsylvania but the reverse in Germany, 65 vs. 68 slightly higher than Palmer's 69 per cent for *Cs* carrying solar halos. The small difference is probably not significant. We cannot say the halo has no value of its own, however, for the presence of a halo identifies *Cs* as a thin ice cloud and is one of the chief criteria for differentiating *Cs* from the thicker *As*.

Double Cirrus. Süring [95, pp. 24–25], following Mylius' lead that bad weather, especially precipitation, is preceded a short time by multilayered clouds, found that double-layered *Ci* (usually the feathery type) and *vertebratus* are followed by warm-front precipitation, mostly of thunderstorm or squall characteristics, in 90 per cent of the 37 cases observed. These are characteristic of secondaries, particularly on the right front

Table II. Frequency of Any Precipitation Within a Given Number of Hours after Halo

Place	Compiler	No. of halos	No. of yrs.	Interval after halo						Freq. of precip: general average	
				12 hr (%)	18 hr (%)	24 hr (%)	36 hr (%)	48 hr (%)	Average (hr)	24 hr (%)	48 hr (%)
Fort Worth, Tex.	Martin [66]	168	6	—	26	36	48	59	24	—	—
Columbia, Mo.	Reeder [79]	40	2	—	—	24	—	49	—	—	—
Wauseon, Ohio	Mikesell [57]	2918	40	—	—	58	—	—	—	—	—
Columbus, Ohio	Martin [65]	185	12	—	54	65	75	86	26	—	—
State College, Pa.	Neuberger [73]	497	6	37	50	58	69	77	15	—	50†
Blue Hill, Milton, Mass.	Palmer [75]	569	10	—	—	<57	67	—	16	50‡	70‡
York, N. Y.	Stewart [91]	317	7	—	—	64	—	83	20	—	—
London, England	Russell [83]	977	7	—	—	68	—	—	—	—	—
NW Germany	Neuberger [74]	316	—	58*	—	68	—	—	8	44†	—

* Frequency of halos followed by precipitation on the same day.
† Nonhalo days.
‡ Eight-year average.

or 69 per cent, and short-lived ($\leq$15 min) are better than long-lived ($\geq$120 min) halos: 80 vs. 60 per cent in Germany, by the end of the following day. These are to be compared with the chance of precipitation within forty-eight hours when no halo is observed, which is 50 per cent at State College. Brooks [17] found precipitation within thirty-six hours after each of seven complex halos, and after a 46-degree halo Reeder [79] said that precipitation [always] occurred in from twenty-four to thirty-six hours in Missouri, while Neuberger found a 78 per cent probability in Germany.

Table III. Halos, Pressure, and Precipitation

Pressure and tendency	Cases	Per cent of halos followed by precipitation within 24 hr
Below normal and falling	334	83
Near the lowest point	205	66
High but falling	893	64
About normal	572	58
Low but rising	199	53
At about highest point	495	42
Above normal and rising	220	37

In evaluating the halo as a prognostic, in comparison with *Cs* without halo, it is of interest to note that Clayton's tabulation of all daytime *Cs*, whether halo-producing or not, shows a rain probability of 73 per cent (interpolated) within thirty-six hours, which is of a general cyclone. Sometimes these clouds are seen after the end of the warm-front precipitation. Double-layered *Ci* occur in 16 per cent of all cases of *Ci*.

Undulated Sheet Clouds. Sweetland [98, p. 39] found that, at Blue Hill, undulated sheet clouds at all levels are more likely to be forerunners of warmer weather and rain than the more indefinite cloud types (Table IV). Sharp contrasts between over- and under-running

Table IV. Undulated Clouds and Precipitation at Blue Hill [98]

	Ci, Cs	*Cc*	*Ac, As*	*Sc*	*St*
Cases	42	16	24	36	6
Precipitation within 24 hr (per cent)	50	66	67	53	67

winds usually result in sharp cloud markings and striking wave formations and produce a maximum tendency toward forced ascent. Martin [67] has presented records of a "mackerel sky" at Columbus, Ohio, which may be compared with Sweetland's undulated *Ac* (Table V).

Mammatus Clouds. According to Sweetland [98], these are followed by some precipitation within twenty-four hours in 51 per cent of 51 cases. Angot [2] remarks that *Cu mam* mean rain without wind in cold air masses below a warm front surface in the lowlands of Lancashire, but indicate a gale in the exposed Orkneys. Whittier

[101, p. 4] says they are an almost sure precursor of rain.

Altocumulus and Rain. In addition to the observations on mackerel sky, other studies of *Ac* have been made. After *Ac* at the morning observation at Podersdam in northwest Bohemia, Schindler [86] found that rain followed on the same day (14 hours) in 42 per cent of the 700 cases and by the end of the following day in 68 per cent.

Table V. Mackerel Sky and Precipitation

	Cases	Cases with precipitation within				
		12 hr number	24 hr		48 hr	
			number	per cent	number	per cent
Mackerel sky (Martin [67])	17	10	15	(88)	17	(100)
Undulated *Ac* (Sweetland [98])	16	—	10	(62)	—	—

Altocumulus Castellatus as a Prognostic of Rain and Thunderstorms. Ley [60] seems to have been the first (1882) to note that morning *castellatus* is commonly followed by afternoon showers, and afternoon *castellatus* by night thunderstorms when, as he said, radiation from *Ac* tops favors convection. A tabulation of subsequent investigations is given in Table VI.

Table VI. Altocumulus Castellatus and Thunderstorm

Region	*Acc*	Thunderstorm on same day (per cent)	Thunderstorm on same or following day (per cent)
Lake Constance (Peppler [76])	Strong	Nearly 100	—
England (Douglas [41])	Strong	—	69*
Bohemia (Schindler [86])	In morning	—	67
N. of Alps (de Quervain [37])	46 cases	74*	87*
Blue Hill Observatory (Sweetland [98])	12 cases (Apr.–Sept.) 1894–1896	66†	—
Blue Hill Observatory (Brooks [19])	51 cases (May–Sept.) 1946–1949	24†	—

* At station or within 100 mi.
† At station or nearby, no rain in some cases.

In the rainier climates of western and central Europe, higher percentages of storms are noted. The greater frequency of *Acc* observations now than fifty years ago seems to indicate that many *Ac*, formerly not designated *castellatus*, are now thrown into that class, diluting the apparent prognostic value that is logically attached to this indicator of a steep lapse rate. In the more recent Blue Hill study, the thunderstorm expectancy for *Acc* days (24 per cent) was nevertheless greater than the expectancy for all days of the period involved (14 per cent). Schinze and Siegel [87, Figs. 67*a* and *b*] have indicated by cross-section diagrams and charts of temperature and humidity vs. height how *Acc* and *floc* are formed and precede *Cb*.

Altocumulus Lenticularis—A Sign of Wet Weather. The appearance of *Ac lent* usually means an increase in humidity in middle levels commonly owing to the advection of moist air in a wind of increasing speed. Such clouds appear first over mountains where the obstructing masses locally force the wind to ascend. Thus it is that new caps on mountains or lenticulars in their lee are early signs of coming wet weather. Rink [80] describes a striking case of the "Moazagotl Wetterwolke," a standing foehn cloud in a horizontal whirl in the lee of the Riesengebirge, that indicated rain inside twenty-four hours.

De Quervain [37] found that *pileus* was accompanied on the same day or followed on the next by *Cb* in 15 out of 21 cases.

Cirrocumulus. Various indications by *Cc* are cited by Angot [2], who says it means fine weather in Britain (and, he could have added, New England [23]), but the opposite in southern Europe, especially Italy. In the tropics, he says, it has no significance. Schindler [86] found *Cc* more a fair- than a foul-weather cloud in Bohemia, with rain on the same day 33 per cent of the time and by the end of the next day 59 per cent (both 9 per cent less than for *Ac* there). Clayton's figure for twenty-four hours was 36 per cent. Bruckmayer and Seiler [20] corroborate Angot's findings.

Directions of Motion, Point of Convergence. Ley [60, pp. 120, 125, 126] early systematized the prognostics of directions of cloud motion, presenting them in qualitative terms in two groups—for *Ci* and for *Cs*. For the latter, he used only *Cs rad* and the direction of the convergence, or V point on the horizon, which he found was an average of 15 degrees to the right of the direction from which the clouds move. Clayton [23] found that 50 per cent of the motions are within 5 degrees of the V point. Rules are given for each of the eight directions. Thus, "Cirrus from N.W., when not tending to form Cirro-filum, is an indication of temporary fine weather, especially in summer"; and "A V point N.W. is an unfavourable symptom, and when it occurs with or just after an increase of barometric pressure, is an indication of a sudden decrease of the same with rain and wind."

Henry, Bowie, Cox, and Frankenfield [49, p. 287] found four helpful rules: (1) *Ci* from the southwest in the lower Mississippi Valley and Texas indicate that precipitation is likely to occur in that region or to the northwest 24–48 hours after their first appearance; (2) *Ci* from the southwest over Arkansas and Oklahoma, changing to *Cs* overcast, indicate that rain will probably soon extend into the middle Mississippi Valley; (3) *Ci* from the northwest in the southeast quadrant of a high seem to indicate that the high will increase in strength and move slowly; and (4) when *Ci* stripes lie northwest to southeast [probably indicating usual prestorm motion from the northwest], the chance of rain is greater than when their direction is northeast to southwest [the usual orientation at the rear of a storm].

Brooks and Harwood [18] showed that heavy snowstorms at Blue Hill are generally preceded by *Ci* from either north or south of west, but not from straight west.

Guilbert [48], in France, found that *Ci* come from the centers of depressions, that their speed indicates the strength but not the speed of the depression, and that *Ci* observations must be used in conjunction with the weather map. Peppler [76], in southern Germany, however, found that the relation between direction of motion of *Ci* and subsequent weather is not very close, but best when taken in conjunction with the pressure tendency.

Deppermann [34], at Manila, warns that striped *Ci* usually arise from the so-called *false Ci* tops of *Cb*, strung out into lines by a strong upper-air stream.

Table VII. Direction of Densest Clouds and of Cloud Motion, and Frequency of Precipitation within 24 hr

		Cs	*As*	*Ac*	*Sc*
(*A*) Direction in which clouds are densest	Direction	SW or N	W	W	S
	Cases	56	16	80	30
	Frequency of precipitation (per cent)	68	81	66	80
(*B*) Direction from which clouds are moving	Direction	NW	SW	SW	S
	Cases	45	16	52	11
	Frequency of precipitation (per cent)	62	69	75	91
Best combinations of (*A*) and (*B*)	Direction where densest	SW	W	SW or W	S
	Direction of motion	SW or W	SW or W	W	S, SW, or W
	Cases	30	13	24	23
	Frequency of precipitation (per cent)	67	92	88	87

Parallel *Ci* give an appearance of convergence at the horizon. He says, "in quite a few cases where the writer thought he had convergent *Ci*, careful observation showed that there were generally two sets of parallel *Ci* of different altitudes meeting near the horizon."

Direction of Motion and Direction Where Densest. Clayton [23] noted that direction of motion and direction of maximum density when considered jointly have better prognostic value than either alone (Table VII).

The Motions of Cirrus as Prognostics of the Movements of Cyclones. The speed of *Ci* is usually a good index of the changeableness of the weather. Since the wind at the *Ci* level is usually dominated by the gradient of temperature in the troposphere, and since this gradient is a large element in the strength of the zonal circulation, there is, necessarily, a relation between the speed and direction of the *Ci* and the movement of cyclones and anticyclones. The *Ci* usually move from the west and so do the pressure systems, and the faster the average movement of the *Ci*, the faster the pressure systems move. When the temperature gradient of the troposphere is small, however, cyclones and anticyclones may dominate the pressure gradients up to great heights, including the *Ci* levels. Under such conditions, the *Ci* may move from easterly directions and vary considerably in short spaces of time or distances [38].

Pick and Bowering [77] related 162 nephoscope observations of *Ci* at Sealand and Holyhead (Britain) to storm tracks and tabulated the difference in degrees between *Ci* motion and the direction of advance of the depression center. More than half fell between 0 and 60 degrees difference; a few exceeded 90 degrees.

Clayton [23] and Pick and Bowering [77] have made statistical studies of *Ci* motions showing a fair, but not sharp, relationship to the progress of cyclones. Clayton analyzed two years (1887 and 1888) of hourly observations of high clouds at Blue Hill, 1821 in number, in relation to the speed of movement of cyclones, using mean monthly speeds of each (see Table VIII). When he grouped the mean monthly *Ci* speeds by classes of 10 m sec^{-1} and with them the storm movements, he found a fairly regular increase of the latter coordinated with that of the former, with ratios of the *Ci* movement to cyclone movement rising from 2.4 in the 15–25 m sec^{-1} *Ci* speed class to 4.5 in the 65–75 m sec^{-1} class. It seems probable, however, that Clayton's ratios are too high, because he compared average monthly *Ci* motion at Blue Hill Observatory with average monthly storm movement for the entire United States. Most of the storms, however, pass through the northeastern section of the country.

Pick and Bowering, 1929 [77], using individual cases (162 cloud observations), found no regular relationship, though their tables (summarized in Table VIII) indicate

some connection between the speed of *Ci* and the speed of cyclones (British Isles).

Cirrus moving at high speed, even if associated with a fast-moving storm, do not necessarily mean prompt arrival of a storm center. Usually the faster the *Ci*,

TABLE VIII. CIRRUS SPEED AND STORM SPEED

(*Adapted from Clayton* [23, p. 463])

Cirrus speed		Storm speed (mph)	Ratio *Ci*/storm
Range (mph)	Average (mph)		
34– 56	49	20	2.4
56– 79	70	25	2.8
79–101	90	29	3.1
101–123	106	32	3.4
146–168	152	34	4.5

(*Adapted from Pick and Bowering* [77])

Number of cases	*Ci* speed (mph)	Storm speed		
		Range (mph)	Per cent of total cases	Per cent probability
20	<26	<25	85	35
		<45	100	81
52	26–50	25–45	52	46
		<45	96	81
74	51–100	>25	82	65
		>45	34	19
16	>100	>25	75	65
		>45	25	19

the greater the ratio of their speed to that of the storm, and therefore, the farther they will have advanced ahead of the storm center.

Another of Clayton's tabulations [23] relates *Ci* motions to temperature changes (Table IX). Observations

TABLE IX. CIRRUS MOTIONS AND TEMPERATURE CHANGES, BLUE HILL

Direction from which *Ci* moved	Temperature change	All year 410 cases (per cent)	Winter hf yr 205 cases (per cent)	Fast *Ci* 71 cases (per cent)
NW	*Rise* in 12 hr	60	64	76
W	*Rise* in 12 hr	58	69	74
SW	*Fall* in 12 hr	57	66	63
NW	*Rise* in 24 hr	43	48	35
W	*Rise* in 24 hr	34	34	32
NW	*Fall* in 24 hr	48	46	59
W	*Fall* in 24 hr	57	61	61
SW	*Fall* in 24 hr	70	83	75

of *Ci*, totaling 410, all made at 8 A.M., were used and frequency of rises and falls in the subsequent twelve and twenty-four hours noted. Cirrus from the west and northwest were usually followed by a rise in twelve hours and then a fall. The rise could be partly discounted as normal diurnal variation, 8 P.M. being somewhat warmer than 8 A.M. In other cases, falls predominated.

If indications of this type were strengthened with more cases and related to other features, they might be of more value.

Brooks and Harwood [18] noted that fast *Ci* occurred before twenty-four out of thirty-two heavy snowstorms (75 per cent) but before only fourteen of twenty-five very heavy ones (56 per cent). While a fast-moving storm favors precipitation in the form of snow, because of the lack of time for advection of warm air, it also passes so soon that the precipitation is more likely to be moderately heavy than very heavy.

The relationship of *Ci* velocities to weather changes is summarized in Table X (after Clayton [23]).

TABLE X. COMPARISON OF CIRRUS VELOCITIES AND WEATHER CHANGES

Ci velocity range (mph)	Average *Ci* velocity (mph)	Mean daily change in pressure (in.)	Mean daily change in temperature (F deg)	Rain variability* (per cent)
34–56	49	0.11	3.8	19
56–79	70	0.12	4.2	37
79–101	90	0.18	5.3	39
101–123	106	0.19	6.4	31
146–168	152	0.23	7.8	51

* Rain every other day would be 100% variability.

Showers. Cloud observations are particularly useful in forecasting the occurrence and locations of showers. Showers developing from pre-warmfrontal instability at middle levels are usually merely punctuations to general upglide precipitation. Their occurrence, however, can sometimes be anticipated from *Acc* or *Ac densus mammatus*, which reveal the presence of strong convection in middle levels. Longer in advance, *Ci floccus* and *Ci densus*, and especially *Ci nothus*, will indicate the approach of showery weather. If these are seen in the front zone of an open-wave cyclone, it is fair to surmise that they are related to pre-warmfrontal convection rather than simply to cold-frontal overturnings.

Cloud Motions and Shapes as Indicators of Wind, Wind Shear, Vertical Currents, and Turbulence. The progressive motion of a cloud, turbulence on its periphery, and deviations of tall clouds from the vertical indicate actual and relative winds aloft. When cumuliform clouds lean or *Ci* trails depart from the vertical, it is obvious that the wind at one height in the cloud is not the same as that at another height. In a rapidly growing cloud, the amount of deviation from the vertical is the resultant of the rate of fall of the cloud particles and the rate of change of wind velocity with height, both of which may vary. Mrs. Malkus [63, 64] has recently described the process for *Cu*, and Ludlam [62], for *Ci*. Ludlam has presented evidence of a more rapid fall of crystals in *Ci* when they grow in passing through moist layers. When crystals begin to melt they will fall faster. In other words, the up-and-down shape of a cloud represents simple or complex streamlines of the rising or falling air columns the motion of which is rendered visible by the cloud material [13; 62, pp. 48–49]. Strong shear, adverse to thunderstorm development, may be recognized in advance on days otherwise favorable for thunderstorms and used accordingly in forecasting, as Brooks pointed out in 1922 [15, p. 283],

and as recently discussed quantitatively by Byers and Battan [22, p. 173; 99].

The development of *Cu congestus* in the morning is to be used with caution in forecasting afternoon showers; because a strong inversion may stop them short of sufficient vertical development. Or, dry air aloft, usually revealed by fraying tops, may be tapped by the convection, thereby raising the cloud base and lowering the temperature and consequent vertical reach of the clouds. Similarly, as noted above, *Acc* frequently do not indicate a shower.

Altocumulus Direction and Speed and Thunderstorm Movement. In a period when afternoon thunderstorms are occurring or are expected, the direction of motion and the apparent speed of the *Ac* should be noted. If, for example, the motion is from the west, any large cloud heaps or arched tops of thunderstorms in that direction should be watched closely. If the *Ac* have been moving fast, shelter should be sought soon, but if they have been going slowly, a distant storm may not arrive for two hours. Furthermore, if a growing thunderstorm is seen approaching apparently from the northwest, even if the *Ac* have been moving rapidly from the west, the storm is not likely to break as soon as might be expected, for the nearest portion of the storm is not coming toward the observer. Under such conditions, however, the squall will come from the west-northwest or northwest, and is likely to prevail for an appreciable interval before rain begins.

Wind-Shift Line Marked by Clouds. When this condition is noted at either a squall line or a cold front, showers are likely. A wind-shift line at the earth's surface is usually marked by some forced ascent of air; the appearance and approach of cloud forms due to forced ascent thus indicate coming change, often rain, within a few minutes to a few hours, depending on the distance to which the clouds can be seen and the rate of approach. If there is a well-defined cloud base, the height of which is estimated, successive measurements of the angular altitude of the base as it approaches will give a reasonably accurate estimate of the time of arrival of the wind shift.

The approach of a strong wind-shift line may be detected an hour or more in advance by observing the thin white arch of the *Ci* border to the anvil top of the approaching *Cb*. Prefrontal weather is so hazy that only this whitish arch will reveal the presence of a moderately distant *Cb*, for the shadowed air under the heavy anvil will be invisible behind the sunlit hazy blue air near the observer. Thus this part of the sky will appear to be clear and will resemble the blue sky above the *Ci* arch.

When a cold front has passed, the lifting effect of the invading cold air mass continues to make clouds form in middle and upper levels. Though the bank of clouds marking the rainy area associated with the front continues to recede in the east or southeast, new clouds, in bands parallel to the front, continue to form. These new bands will usually not form as rapidly as the general movement of the front carries the whole cloud system away. Indeed, in the daytime, solar heating may throw the western or northwestern edge more rapidly eastward or southeastward than the rate of progress of the front. In the late afternoon and evening, however, the cooling of the moist layers by radiation may, in the case of slowly moving fronts particularly, make the cloud edge retrograde, giving an apparent threat of renewed precipitation which, of course, is quite possible if the front is not far away and if it has some waves in it.

A daytime phenomenon of the clearing-off zone, which Dr. Aili Nurminen has told the author not infrequently makes airport forecasts of improving ceilings fail, is the formation of an abundance of very low *fractocumulus* as soon as the middle and upper clouds permit moderate insolation to reach the wet ground. Fortunately, these soon break, for they so reduce the insolation that convection is weakened. Though more clouds will then form owing to renewed insolation, their bases will be higher.

Difficulties in Forecasting Fronts in Tropics. In the tropics, where fronts are usually slow-moving and often stationary for long periods, the forecasting of frontal passages, which are the chief elements of change in those latitudes, is difficult, as Deppermann [34] very reluctantly admits. Fronts at Manila come mainly from the north, and the upper-air changes due to the overriding of the northers by the trade will occur north of the front and not near Manila. The rare returns of fronts from the south may be heralded by changes in the trade wind aloft. The equatorial front has little overrunning, for the temperature differences between the trades and the southwest monsoon are slight. Even the pre-typhoon sequence of clouds is not perfectly regular, and almost all the cloud forms found in typhoons can also appear on days of strong northers or of many thunderstorms.

Synoptic Types of Sky: How Synoptic Cloud Data May Aid Weather Forecasting

We in the United States pay less attention to states of the sky and cloud sequences than do meteorologists of western Europe, because, except on the Pacific Coast, we have a fairly adequate network of stations, and so do not have to depend upon cloud forms and relationships to give us an idea of weather conditions to the west. Nevertheless, if we do not use the cloud portion of the weather picture, we are depriving ourselves of this considerable advantage in analyzing the weather situation and in following current trends. Schereschewsky [84] emphasized this strongly, as did Bigelow [7] and Ley [60] before him.

The forecaster's situation is still happier, however, when the usual synoptic surface and upper-air data are supplemented with accurate and frequent cloud observations. Of the nineteen elements plotted on the U. S. Daily Weather Maps, seven relate to cloud conditions, and three of these are devoted to thirty separate indications of clouds and their trends at various levels.

The thirty indications are divided into three groups of ten, C_H, C_M, and C_L, each roughly representative of conditions at high, middle, and low levels. The sequence of numbers in each group is generally that from better to worse weather as a storm approaches, ending with the

breakup of the storm clouds as the center or front passes. The various types of sky indicated by the combination of the current C_H, C_M, and C_L have been related to the different portions of cyclones, anticyclones, and frontal systems and to their intensification or weakening. Thus the synoptic maps of types of sky will, by themselves, give a fairly complete picture of the positions and intensities of the cyclones, anticyclones, and fronts of a weather map. Moreover, the diurnal sequence of cloud forms characterizes the type of air mass so well that other features of air mass weather can often be forecast.

An Historical Summary of Cloud Synoptics. Ley, in England, was the first (1872) to show *Ci* motion on a synoptic chart [61], and the first (1878) to make a composite chart of cloud distribution in a typical cyclone [60], which included forms and motions and showed the open sector clear to the center on the south and southwest, with the storm as a whole moving northeastward. Hildebrandsson, in Sweden, was the first (1883) to make a systematic study of the distribution of meteorological elements, including the kinds of clouds and their motions, about barometric minima and maxima [50]. In the United States, Davis [32], in 1894, published a generalized, dynamic cloud cross section of a tropical cyclone and a generalized map of *Ci* and other clouds and rain, winds and isobars of a well-developed low; in 1896 Clayton [23] published a monumental study of the distribution of cloud forms and their motions in cyclones and anticyclones at Blue Hill as compared with Europe, including tables, cross sections, and maps of averages and frequencies; Sweetland [98] followed in 1897 with the cyclonic and anticyclonic distribution of special cloud forms, and in 1900 Bigelow [7] published a detailed 747-page study of the cloud observations at stations throughout the country. The works of Ley, Hildebrandsson, Clayton, and Bigelow were outstanding contributions to our knowledge of the characteristics and circulations of cyclones and anticyclones. These (except Bigelow's) and other researches have been brilliantly analyzed and collated by Hildebrandsson and Teisserenc de Bort [51]. Brooks [11] interpreted, with a cross-section diagram, the cloud details observed at Washington during the passage of a strong cyclone in 1919. During World War I, the French and Norwegians were handicapped by a lack of weather data from the west, and had to base their weather forecasts largely on synoptic interpretations of the cloud systems they observed coming in from the Atlantic. Indeed, this led to the development of air-mass and frontal analysis in Norway.

Bjerknes, in 1919, first related the cloud pattern to frontal systems, with diagrammatic maps and cross sections [8] and, in 1922 with Solberg, added occlusions [9]. Though Guilbert [48] had laid the groundwork in 1922, Schereschewsky and Wehrlé [85] developed the mapping of cloud systems in France in 1923. In developing the dynamics of air masses and fronts, Bergeron in 1928 [4] found clouds helpful. He and Wehrlé, as the most active members of the International Commission for the Study of Clouds, from 1926 to 1932 worked cloud synoptics into the relatively simple pattern presented verbally and diagrammatically in the *International Atlas of Clouds and States of the Sky* [53]. This pattern was amplified by Dedebant and Viaut in 1936 [33] and more clearly depicted by means of cloud cross sections along five lines through an occluded cyclone, and a map of two cloud systems in western and central Europe on February 21, 1935. Viaut [100] has recently improved the idealized pattern and combined the five cross sections into four. Other sections of single storms have been made, giving the details of growing winter cyclones. Photographs from an airplane were taken by Conover [25] to illustrate a cold front, by Douglas [42] to show a cold front and squall line, and diagrams were drawn by Simon [90] to show fronts in Egypt.

Alpert [1] and Berkofsky [6] report the use of low- and middle-cloud sky-cover charts over the easternmost tropical portion of the Pacific Ocean during World War II. These were based on information obtained from patrol flights. In some cases, isolines were drawn for every 0.2 of cloud cover. Such charts proved to be "exceedingly helpful in locating the intertropical convergence zone." Though not strictly synoptic, this was only a slight drawback, for the zones were quasi-stationary.

In 1948, Deppermann [34], after a study of *Ci* stripes in the Philippine Islands, wrote, "Since upper-air currents are themselves subject to considerable changes of path as they proceed, it is evident that an intelligent use of striped *Ci* to indicate the seat of frontal action or a typhoon center demands a good network of upper-air wind and velocity data."

Cloud Cross Sections. In 1926 and 1928, respectively, Stüve [92] and Kopp [59] used Lindenberg data (non-instrumental and range-finder observations of clouds in detail, and records from kite, balloon, and airplane meteorographs) as bases for very detailed representations of temperature inversions, of circulation, cloud, and precipitation cross sections of a warm front, of primary and secondary cold fronts (Stüve), and all the way from anticyclone center to anticyclone center across an occluded front (Kopp). Stüve [94] improved the details of his warm-front section in 1937 and added a cross section through a cold air tongue. In 1943 Schinze and Siegel [87, Figs. 61, 62, 64, 65, 75, 76] published detailed cross sections of warm, cold, and occluded fronts, in which the composition of the clouds (ice, supercooled water, ice and water mixed) and type of precipitation (five types) are shown.

Schneider-Carius [88] has just published a critique of cloud systems, stating that "with the vast experience gained from meteorological flight observations Stüve's cloud system, as a sequel to that of Bjerknes, is considered as antiquated, [while] retaining their importance are the cloud systems of Schinze-Siegel, Schwerdtfeger, and Wehrlé." He finds a "ground layer" the types of which have definite relations to the various states of the low-cloud types of sky of the *Atlas of Clouds* and apply in the tropics as well as in middle latitudes.

Detailed studies of clouds in a particular cyclone are

an aid in interpreting cloud sequences. A cross section showing the clouds and their movements as a storm passes a single station [11] indicates trends but is not synoptic. Conover and Wollaston [27] have recently prepared a detailed study of the cloud systems of a winter cyclone based on a large amount of data from all reporting stations, both surface and upper air, which was reduced to maps of the eastern United States at 3-hour intervals throughout several days. These give a true picture of the cloud development from the origin through the various stages of growth of a cyclone into a full-fledged snowstorm.

Cloud Indications of New Developments. Cloud observations, synoptically mapped, are particularly valuable for indicating the intensification of a low before this trend is apparent in the isobars or surface winds. Often middle clouds afford the first sign of the formation of secondaries, as Miller [69] and George [46] have indicated. Miller points to the sequence of C_M3, C_M5, or C_M7, then C_M2 (*As* becoming *Ns*) in the earliest stages of the development of middle or south Atlantic coastal storms. This cloud development is easily recognizable in Type A storms (new lows, off the coast, usually) but in Type B (secondaries) the new development will be superposed on the lateral sky of the primary, but will be recognizable either by being separated from clouds of equal development nearer the center of the primary or by the apparent progress of this cloud development at a greater rate than the actual advance of clouds into the region.

Cloud and Weather Types. Deppermann [36] has classified Manila weather types more or less directly according to cloud types. His major divisions are Pure Trade, Trade, Northers, Convective Trade, Mild SW Monsoon, Frontal SW Monsoon or SW Monsoon Sector (with typhoon quite distant), and Typhoon types, each with one to seven subdivisions, making twenty-seven altogether. These he has related to an elaborate classification according to frontal types of eighteen main divisions, each with one to seven subdivisions. The map gives the frontal type for the day and the reference table then presents the cloud and other weather features to be expected.

In 1894, Ley [60, see pp. 202–204] proposed something of this sort: that chart books be prepared, showing the distribution of weather, and probabilities for each of a number of synoptic situations. These chart books, he proposed, should be distributed to users of the forecasts and be consulted when the synoptic type was identified by telegram from the forecaster.

Besides the cloud photographs of the various states of the sky in different parts of a cyclone published in the *Atlas of Clouds* [53], an excellent set of such pictures has been published by Miss Douglas [40, see Figs. 142–162, pp. 108–116], and a set of drawings by Brooks [16].

Cloud Indications of How an Approaching Low Will Pass a Station. As the clouds of an approaching storm gather, how can the observer apply the knowledge gained by statistical studies and the idealized or actual charts and cross sections of the distribution of cloud forms mentioned above? The problem is to forecast the track the center will take to one side or the other, the nearness of approach of the center, and the general intensity of the storm. As the clouds thicken steadily from *Ci* to *Cs* and from *Cs* to *As*, an observer naturally expects further development to *Ns* with its rain or snow. There are usually counterindications if such will not be the case, or if the precipitation will come late, amount to little and end soon, or will come in two brief periods separated by several hours of mild, more or less sunny weather. If the northern horizon remains clear until *As* overspreads most of the sky, the storm will probably be passing to the south without bringing precipitation. If the northern horizon is slow in clouding up but becomes covered by the time the cloud sheet is principally *As*, there will probably be some precipitation but not much. In both cases, the temperature will stay low.

If the cloud sheet after increasing to *As* breaks into *As* and *Ac* or *Ci den* and the sky above is seen to have lost most of its *Cs*, the storm is either weakening or passing on the north. In the latter case, the east or northeast wind will veer to southeast and south, unless a secondary is developing to the southwest or south, in which event a new cloud system will appear and go rapidly through the north-of-the-center sequence, already described.

Through the southern margin of a storm, the *Ci* appear first in the west or northwest. They increase hesitatingly. The wind blows from the southeast and does not become strong. The *Ci* thicken to broken *Cs* and these to broken *As*, with local patches of *Ac*. There may be a little rain, and then the wind will turn to south as the warm front passes. The warm, muggy, sometimes showery weather that follows will last for some time before the cold front sweeps over from the northwest or even north-northwest, perhaps with a thunderstorm, but, except for the squall, with less wind afterward. If the low has already occluded, there will be but one brief rain, if any.

The approach and passage of a squall line or cold front is the most spectacular phenomenon of a trip through a low south of the center when the action is strong. A typical sequence is as follows. In the northwest, there appears a long bank of dense, towering clouds, with *Ci*, *Cs*, and *As* in front. The bank soon stretches from west to north, then from west-southwest to north-northeast, and rises higher in the sky. The wind is still increasing. Thunder may now be audible. The high clouds cover most of the sky. The wind begins to weaken as a long, low arch of gray cloud appears across the northwestern sky, which is darkening and growling ominously. The arch rises and lengthens at an alarming rate, as the wind dies to a calm. Now, stretching from the northeastern to the southwestern horizons, the storm collar, with ragged pendants, reaches the zenith as the thundersquall strikes suddenly from the northwest. A downpour follows, with lightning and sharp thunder, and the squall diminishes. After half an hour or more, the sky lightens in the northwest, and the rain soon ceases. As the massive cloud bank retreats in the southeast a cool, dry northwesterly wind begins

to blow. (For another account of the cloud sequence through a storm, see Rossby [81].)

After the low has passed north of the observer, if the wind continues to shift to north or even northeast, and if the high clouds do not clear away, and especially if middle clouds continue, a wave is probably forming on that portion of the front to the southwest, now quasi-stationary and developing into a warm front, and precipitation may be expected again soon, particularly if the clouds are moving rapidly.

Though we cannot agree with Schneider-Carius [88] that it is naive to think one can forecast from clouds alone, it is desirable to emphasize, as did a recent reviewer of Douglas' *Cloud-Reading for Pilots* [40], that an attitude of caution should be adopted towards forecasts deduced solely from cloud structures, since the actual sequences of weather are extremely variable. Knowledge of cloud characteristics of air masses and fronts on the part of the user of forecasts enhances the value of the official forecasts. Brunt [21] wrote: "The greatest difficulty which meets the official forecaster is probably that of timing. . . . The recipient of the forecast can learn, by careful study of the clouds, to check for himself whether the timing of a forecast is correct or not, and in many cases can make use of the forecast when the timing is wrong. . . ."

Desirable Lines of Development

Though cloud systems are already used in professional forecasting, closer attention to clouds should improve both local and general forecasting. For better interpretation of cloud data we need more detailed cloud charts, such as those already mentioned [1, 6, 11], and many more detailed studies of the sequences of clouds in relation to the development and movement of pressure systems and their fronts.

For better observations we need more nephoscopes and more ceilometers and range finders. Nephoscopes should be available at all regularly reporting synoptic stations, to facilitate accurate reports of cloud motions. This would provide a reliable areal continuum for upper-air conditions between the rather openly spaced balloon stations. Ceilometers and range finders, as well as nephoscopes, should be installed at all aerological stations, to supplement the six-hourly pibal and twelve-hourly radiosonde or rawinsonde observations, and to provide a continuous picture of aerological changes.

Wood [102] has suggested that, in the future, automatic radio weather stations might be equipped with television to provide the distant forecaster with a view of the types of sky over his synoptic area. For the time being, we shall have to be content with photoelectric registration of the intensity, degree, and rapidity of variation of light from a small spot of sky overhead to give a rough representation of cloud type (Falconer [43*a*]).

By means of photographs taken from a V-2 rocket, Bergstrahl [5] has shown the distribution of *Cu* in distinct areas, and Crowson [31] has obtained a composite picture of an area of over 500,000 square miles in the southwestern United States and northern Mexico. This comprehensive view made it possible to correlate clouds with atmospheric structure on a large scale. The use of guided missiles to gather atmospheric data has tremendous potentialities, he believes. If television is used instead of conventional photography, complete coverage can be obtained without the problem of film recovery.

Although rocket photography technique is well adapted to desert areas where clouds are thin, it could not do justice to the many layers of damper climates. Perhaps the answer here is television both from above and below!

REFERENCES

1. Alpert, L., "Weather over the Tropical Eastern Pacific Ocean, 7 and 8 March, 1943." *Bull. Amer. meteor. Soc.*, 27: 384–398 (1946).
2. Angot, A., *Traité élémentaire de météorologie*, 2e éd. Paris, Gauthier-Villars, 1907.
3. Berg, H., "Ergebnisse und Kritik von Wolkenmessungen europäischer Wetterflugstellen." *Beitr. Phys. frei. Atmos.*, 21: 75–91 (1934). (Summary in *Meteor. Z.*, 50: 452–459 (1933).)
4. Bergeron, T., "Über die dreidimensional verknüpfende Wetteranalyse." *Geofys. Publ.*, Vol. 5, No. 6, 111 pp. (1928).
5. Bergstrahl, T. A., "Photography from the V-2 at Altitudes Ranging up to 160 km" in *Upper Atmosphere Research*. Rep. No. 11, Naval Res. Lab. Rep. R-3171, 1947. (See pp. 119–130)
6. Berkofsky, L., "Synoptic Cloud Charts." *Bull. Amer. meteor. Soc.*, 28: 105 (1947).
7. Bigelow, F. H., *Report on the International Cloud Observations May 1, 1896 to July 1, 1897*. U. S. Weather Bureau, Rep. of Chief, 1898–99, Vol. 2. Washington, D. C., U. S. Govt. Printing Office, 1900.
8. Bjerknes, J., "On the Structure of Moving Cyclones." *Geofys. Publ.*, Vol. 1, No. 2, 8 pp. (1919). (See pp. 4–6)
9. —— and Solberg, H., "Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation." *Geofys. Publ.*, Vol. 3, No. 1, 18 pp. (1922).
10. Bleeker, W., and Andre, M. J., *Convective Phenomena in the Atmosphere*. Tech. Rep. to Office Nav. Res., Contr. N6-ori-20, Task Order No. 2, Proj. NR 082 003, 24 pp. Dept. Meteor., Univ. Chicago, 1949. (Also *J. Meteor.*, 7: 195–209 (1950).)
11. Brooks, C. F., "A Cloud Cross-Section of a Winter Cyclone." *Mon. Wea. Rev. Wash.*, 48: 26–28 (1920).
12. —— *Cloud Observing*. Third Quart. Rep., (Harvard-Blue Hill) U. S. Weather Bureau Contr. Cwb-8120, 1950. (See pp. 24–45)
13. —— "Clouds in Aerology (I)." *Bull. Amer. meteor. Soc.*, 22: 335–345 (1941).
14. —— "Local Forecasting from Clouds" in Gregg, W. R., *Aeronautical Meteorology*, 1st ed. New York, Ronald, 1925. (See pp. 102–105)
15. —— "The Local, or Heat, Thunderstorm." *Mon. Wea. Rev. Wash.*, 50: 281–284 (1922).
16. —— "Waves, Wind, and Weather" in *Science from Shipboard*, pp. 11–52. Washington, D. C., Science Service, 1943. (See pp. 31–35)
17. —— "A Quartet of Complex Halos and the Weather Which Followed Them." *Bull. Amer. meteor. Soc.*, 16: 305–306 (1935).
18. —— and Harwood, E. M., Jr., "Cloud-Observations in

Short-Term Forecasting of Snow-Storms." *Trans. Amer. geophys. Un.*, 16 (Pt. I): 110–114 (1935).

19. Brooks, Eleanor S., *Table of Ac cast and floc and Thunderstorms at Blue Hill, Milton, Mass.* Ms., 1950.
20. Bruckmayer, F., und Seiler, R., "Cirrocumuli als Schlechtwettervorzeichen." *Meteor. Z.*, 61: 247–249 (1944).
21. Brunt, D., "Foreword" to Douglas, Ann C., *Cloud Reading for Pilots*. London, J. Murray, 1943. (See pp. viii-ix)
22. Byers, H. R., and Battan, L. J., "Some Effects of Vertical Wind Shear on Thunderstorm Structure." *Bull. Amer. meteor. Soc.*, 30: 168–175 (1949).
23. Clayton, H. H., "Discussion of the Cloud Observations Made at the Blue Hill Observatory." *Ann. Harv. Coll. Obs.*, Vol. 30, Pt. 2, pp. 364–365, 457–479 (1896).
24. —— "Measurements of Cloud Heights, Velocities, and Directions." *Ann. Harv. Coll. Obs.*, Vol. 42, Pt. 2, App. C, pp. 207–209 (1900).
25. Conover, J. H., "Observations and Photographs of a Cold Front Made from an Airplane." *Bull. Amer. meteor. Soc.*, 29: 313–318 (1948).
26. —— Fergusson, S. P., and Brooks, C. F., *The Blue Hill Observatory Nephoscope, 1948 Model*. Supp. to Third Quart. Rep. (Harvard-Blue Hill) U. S. Weather Bureau Contr. Cwb-8120, 1950.
27. Conover, J. H., and Wollaston, S. H., "Cloud Systems of a Winter Cyclone." *J. Meteor.*, 6: 249–260 (1949).
28. Craddock, J. M., "The Development of Cumulus Cloud." *Quart. J. R. meteor. Soc.*, 75: 147–153 (1949).
29. —— "The Effect of Water Droplets on the Buoyancy of a Rising Cumulus Cloud." *Meteor. Mag.*, 78: 351–354 (1949).
30. —— "A Note on Cumulonimbus Photographed near Singapore on March 5, 1947." *Meteor. Mag.*, 78: 290–295 (1949).
31. Crowson, D. L., "Cloud Observation from Rockets." *Bull. Amer. meteor. Soc.*, 30: 17–22 (1949).
32. Davis, W. M., *Elementary Meteorology*. Boston, Ginn, 1894. (See Fig. 60, p. 204; Fig. 72, p. 228)
33. Dedebant, G., et Viaut, A., *Manuel de météorologie du pilote*. Paris, Blondel la Rougery, 1936. (See Figs. 21, 22, 36–39, 85)
34. Deppermann, C. E., "Cirrus Stripes and Typhoons." *Bull. Amer. meteor. Soc.*, 29: 166–174 (1948).
35. —— "Relative Humidity Gradient and the Form of Cloud Bases." *Bull. Amer. meteor. Soc.*, 26: 267–270 (1945).
36. —— *The Weather and Clouds of Manila*. Manila, Bureau of Printing, 1937.
37. de Quervain, A., "Beiträge zur Wolkenkunde." *Meteor. Z.*, 25: 433–453 (1908).
38. Dight, F. H., "The Significance of Nephoscope Observations." *Meteor. Mag.*, 65: 280–284 (1931).
39. Dole, R. M., "The Fire-Colored Sunset as a Valuable Clue to the Existence of a Tropical Storm." *Mon. Wea. Rev. Wash.*, 49: 191 (1921).
40. Douglas, A. C., *Cloud-Reading for Pilots*. London, J. Murray, 1943.
41. Douglas, C. K. M., "Alto-cumulus Castellatus Clouds and Thunderstorms." *Meteor. Mag.*, 66: 106–109, 211–212 (1931).
42. Douglas, R. H., "O'er Which They Flew." *Weather*, 4: 298–300 (1949).
43. Eredia, F., *Gli Strumenti di Meteorologia ed Acrologia*. Rome, Bardi, 1936. (See pp. 219–231)

43*a*. Falconer, R. E., *A Method for Obtaining a Continuous Record of the Type of Clouds in the Sky during the Day*. Gen. Elec. Res. Lab. Occ. Rep. No. 8, 7 pp., Project Cirrus, 1949.

44. Fergusson, S. P., "Nephoscopes (I)." *Bull. Amer. meteor. Soc.*, 32: 259–266 (1951).
45. Fuller, M. L., "The Use of Clouds in Local Forecasting." *Mon. Wea. Rev. Wash.*, 47: 473–474 (1919).
46. George, J. J., *Weather Forecasting at N. Y., Associated with the Formation of Atlantic Secondaries*. Atlanta, Eastern Air Lines, Inc., 1948. (See p. 26)
47. Grant, H. D., *Cloud and Weather Atlas*. New York, Coward, 1944.
48. Guilbert, G., "Sur l'application des cirrus à la prévision du temps." *C. R. Acad. Sci., Paris*, 170: 1398–1399 (1920).
49. Henry, A. J., and others, *Weather Forecasting in the United States*, 370 pp. U. S. Weather Bureau No. 583, Washington, D. C., 1916.
50. Hildebrandsson, H. H., "Sur la distribution des éléments météorologiques autour des minima et des maxima barométriques." *Nova Acta Soc. Sci. upsal.*, (3) Vol. 12, No. 6, 31 pp. (1885).
51. —— et Teisserenc de Bort, L., *Les bases de la météorologie dynamique*. Paris, Gauthier-Villars, 1907. (See Vol. 1, p. 129, Pl. 21; Vol. 2, pp. 1–144)
52. Humphreys, W. J., *Weather Proverbs and Paradoxes*, 2nd ed. Baltimore, Williams & Wilkins, 1934. (See pp. 49–53)
53. International Meteorological Committee, *International Atlas of Clouds and States of the Sky*. Paris, Off. Nat. Météor., 1932. (Title changed in 1939 reprint from ". . . *States of the Sky*" to ". . . *Types of Skies*.")
54. International Meteorological Organization: Commission for Instruments and Methods of Observation, 1st Sess., 1947, *Abridged Final Report*. O. M. I. Publ. No. 68, pp. 70–73. Lausanne, Switzerland, 1949.
55. Inwards, R., *Weather Lore*, 4th ed. London, R. Meteor. Soc., 1950. (See pp. 78, 85–87, 89, 98, 224 for halos and coronas; and pp. 120, 123–149, 224–229 for clouds)
56. Kähler, K., *Wolken und Gewitter*. Leipzig, Barth, 1940.
57. Kirk, J. M. [for Mikesell, T.], "Halos and Precipitation at Wauseon, Ohio." *Mon. Wea. Rev. Wash.*, 42: 616 (1914).
58. Knoch, K., und Blüthgen, J., "Klimatologie und Meteorologie (1929–1938)." *Geogr. Jb.*, Jg. 59 (1948). (See p. 40)
59. Kopp, W., "Über Wolkenmessungen in Zusammenhang mit der Wetterlage." *Arb. preuss. aero. Obs.*, 15: 271–291 (1926).
60. Ley, W. C., *Cloudland: A Study on the Structure and Characters of Clouds*. London, Stanford, 1894.
61. —— *The Laws of the Winds Prevailing in Western Europe, Pt. I*. London, Stanford, 1872. (Abstract in *Quart. J. R. meteor. Soc.*, 1: 109–110 (1872).)
62. Ludlam, F. H., "The Forms of Ice Clouds." *Quart. J. R. meteor. Soc.*, 74: 39–56 (1948).
63. Malkus, Joanne S., "Effects of Wind Shear on Some Aspects of Convection." *Trans. Amer. geophys. Un.*, 30: 19–25 (1949).
64. —— Bunker, A. F., and McCasland, K., *Observational Studies of Convection*. Woods Hole Ocean. Instn., Ref. No. 49–51, 1949.
65. Martin, H. H., "Further Study of Halos in Relation to Weather." *Mon. Wea. Rev. Wash.*, 46: 119–120 (1918).
66. —— "Halos at Fort Worth, Tex., and Their Relation to the Occurrence of Subsequent Precipitation." *Mon. Wea. Rev. Wash.*, 44: 67 (1916).

67. —— "Mackerel Sky as a Prognostic of Precipitation." *Mon. Wea. Rev. Wash.*, 48: 156 (1920).
68. Middleton, W. E. K., *Meteorological Instruments*, 2nd ed. Toronto, University of Toronto, 1947.
69. Miller, J. E., "Cyclogenesis in the Atlantic Coastal Region of the United States." *J. Meteor.*, 3: 31–44 (1946).
70. Moles, F. J., "The Cloud Range Meter." *Gen. elect. Rev.*, 49: 46–48 (1946).
71. Möller, F., "Makroskopische Wolkenphysik," *Meteorologie und Physik der Atmosphäre*, R. Mügge, sen. ed., *Naturforschung und Medizin in Deutschland 1939–1946* (*FIAT Rev.*). Wiesbaden, Dieterich Verlag, 1948. (See Vol. 17, pp. 109–113)
72. Neiburger, M., "Temperature Changes during Formation and Dissipation of West Coast Stratus." *J. Meteor.*, 1: 29–41 (1944).
73. Neuberger, H., "Forecasting Significance of Halos in Proverb and Statistics." *Bull. Amer. meteor. Soc.*, 22: 105–108 (1941).
74. —— "Halobeobachtungen in Norddeutschland." *Z. angew. Meteor.*, 54: 131–135, 160–169 (1937).
75. Palmer, A. H., "Halos and Their Relation to the Weather." *Mon. Wea. Rev. Wash.*, 42: 446–451 (1914).
76. Peppler, W., "Zur Aerologie des Castellatus." *Beitr. Phys. frei. Atmos.*, 13: 45–53 (1926). See also "Über Temperatur und Feuchtigkeit in der freien Atmosphäre." *Ibid.*, 21: 121–128 (1934).
77. Pick, W. H., and Bowering, D. F., "Cirrus Movement and the Advance of Depressions." *Quart. J. R. meteor. Soc.*, 55: 71–72 (1929).
78. Raethjen, P., "Strömungsvorgänge und Wolkenformen." *ErfahrBer. dtsch. Flugwetterd.*, 2. Sonderband (1932). (See pp. 147–157)
79. Reeder, G., "Observations of Halos at Columbia, Mo." *Mon. Wea. Rev. Wash.*, 35: 212 (1907).
80. Rink, J., "'Moazagotl'-Wetterwolke." *Meteor. Z.*, 54: 190–192 (1937).
81. Rossby, C.-G., "Amateur Forecasting from Cloud Formations" in *Climate and Man: Yearbook of Agriculture*. Washington, D. C., U. S. Govt. Printing Office, 1941. (See pp. 656–661)
82. Rossi, V., "Untersuchungen über die Wolken nach meteorologischen und aerologischen Beobachtungen in Utti 1932–1939." *Mitt. meteor. Zent-Anst. Helsinki*, No. 32, 46 pp. (1948).
83. Russell, S., "Note on Halo Frequency and the Succeeding Occurrence of Precipitation in London 1918–1924." *Quart. J. R. meteor. Soc.*, 52: 118–127 (1926).
84. Schereschewsky, P. L., "Clouds and States of the Sky" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See pp. 882–926)
85. —— et Wehrlé, P., "Les systèmes nuageux." *Mém. off. nat. météor., Paris*, No. 1 (1923).
86. Schindler, G., "Der Altocumulus und seine Bedeutung als 'Wettervorzeichen.'" *Ann. Hydrogr., Hamburg*, 67: 194–196 (1939).
87. Schinze, G., und Siegel, R., "Die luftmassenmässige Arbeitsweise." *Wiss. Abh. D. R. Reich. Wetterd.*, Sonderband. Leipzig, W. Keller, 1943. (Text, 99 pp.; figures, 167 pp. (bound separately).)
88. Schneider-Carius, K., "Die Bedeutung des Schichtenbaues der Troposphäre für die Aufstellung von Wolkensystemen." *Arch. Meteor. Geophys. Biokl.*, (A) 2: 97–118 (1950).
89. Schwerdtfeger, W., "Meteorologische Erfahrungen bei Wettererkundungsflügen über See." *Forsch. ErfahrBer.*, Reihe B, Nr. 7 (1942). (Summarized in "Meteorologie und Physik der Atmosphäre," R. Mügge, sen. ed., *Naturforschung und Medizin in Deutschland, 1939–1946* (*FIAT Rev.*). Wiesbaden, Dieterich Verlag, 1948. See pp. 181–182.)
90. Simon, A., "Contribution à l'étude des nuages." *Bull. Inst. égypt.*, 26: 149–162 (1944).
91. Stewart, M. N., "Halo Observations at York, N. Y." *Mon. Wea. Rev. Wash.*, 43: 444–445 (1915).
92. Stüve, G., "Wolken und Gleitflächen." *Arb. preuss. aero. Obs.*, 15: 214–224 (1926).
93. —— "Thermodynamik der Atmosphäre" in *Handbuch der Geophysik*, Abschn. IV, Bd. IX. Berlin, Gebr. Borntträger, 1937. (See pp. 254–314)
94. —— "Die atmosphärischen Zirkulationen." *Ibid.*, Abschn. VI, SS. 434–591. (See Figs. 258 and 259, p. 493)
95. Süring, R., "Photogrammetrische Wolkenforschung in Potsdam in den Jahren 1900 bis 1920." *Veröff. preuss. meteor. Inst.*, Bd. VII, Nr. 3 (1922).
96. —— *Die Wolken*, 2. Aufl. Leipzig, Akad. Verlagsges., 1941.
97. —— "Wolkenmessung" in Kleinschmidt, E., *Handbuch der meteorologischen Instrumente*. Berlin, Springer, 1935.
98. Sweetland, A. E., "A Study of Special Cloud Forms." *Ann. Harv. Coll. Obs.*, Vol. 42, Pt. I, pp. 28–40 (1897).
99. U. S. Weather Bureau, *The Thunderstorm*, 287 pp. Washington, D. C., U. S. Govt. Printing Office, 1949. (See pp. 101–108)
100. Viaut, A., *La météorologie du navigant*. Paris, Blondel la Rougery, 1949. (See Figs. 93, 96, 97; Pl. I)
101. Whittier, B. B., *Clouds as an Aid to the Forecaster*. Ms. in Blue Hill Library, 10 pp., January 1922.
102. Wood, L. E., "Automatic Weather Stations." *J. Meteor.*, 3: 115–121 (1946).

FOG

By JOSEPH J. GEORGE

Eastern Air Lines, Atlanta, Ga.

Introduction

Fog. The chapter entitled "Fog" in Byers' *General Meteorology* [2] begins with the simple sentence, "It is unwise to attempt an exact definition of fog." It is difficult to find a more adequate beginning.

When a cloud consisting of minute water droplets or ice crystals envelops the observer and restricts his horizontal visibility to 1000 m or less, the international definition of *fog* has been satisfied. Under similar conditions, but with visibility greater than 1000 m the condition is described by the word *mist*, although in the United States this term is popularly applied to the condition meteorologists know as *drizzle*.

Few other meteorological phenomena depend so much upon the location of the observer. A motorist may on one occasion encounter dense fog in the valleys of a hilly terrain and perfectly clear air as he tops the ridges. On another occasion when there is a very low stratus cloud, he may be in clear air only in the valleys and in fog on the ridges. Petterssen [20] points out these inconsistencies clearly and George [6] goes as far as to recommend classing together all types of low clouds forming under conditions which produce fog. The latter course has much to recommend it from the restricted viewpoint of aviation forecasting, but it is completely inadequate for many ordinary uses and impossible for observing practice. The conclusion seems inescapable that with all their clumsiness the present definitions are the most practical solution.

Smoke and Haze. Any discussion of fog must include mention of the related phenomena of smoke and haze. Smoke is considered to be atmospheric pollution caused by the products of imperfect combustion, usually finely divided carbon. Haze consists of very tiny dust particles. Against a dark background such as mountains, haze presents a blue appearance while mist or fog has a gray tone. It will be pointed out later that under many conditions it is impossible to tell when dry haze becomes moist haze, due to the action of hygroscopic nuclei, and when moist haze becomes fog, except for arbitrary visibility limits. The process is a continuous one.

During stagnant anticyclonic conditions, when stabilizing influences are extremely marked, industrial pollution is added to the atmosphere in a normal amount but is confined to the lowest layers of the air rather than being dissipated throughout a deep layer and thus at times produces a high concentration of foreign particles in the air. The combination of fog and smoke under such conditions may remain day and night in a greater or lesser degree. This condition is popularly called *smog*.

Physics of Fog Formation

Since fog is a cloud which happens to form at or near the surface of the earth, the physics of its formation should not be very different from that of a true cloud. Accordingly it is not the purpose of this section to deal with cloud physics except for those features of it which are unique to the formation of fog, or which may differ from their counterparts in cloud above the ground.

The Role of Condensation Nuclei. Through the works of Wilson, Aitken, Wigand, and others, it has been known for many years that condensation under atmospheric conditions requires the presence of some sort of nuclei if it is to take place at any reasonable humidity. At first it was concluded that ordinary dust furnished these nuclei, but it is now believed that only special types of hygroscopic nuclei are really effective. Among these are (1) salt particles which apparently have become airborne from sea spray and have been carried long distances by the wind, and (2) hygroscopic products of combustion such as sulfur trioxide and possibly nitric oxide. According to a description by Landsberg [15], condensation does not take place in the absence of ordinary nuclei until about fourfold supersaturation is attained, and it is believed that this condensation takes place with ions as nuclei. Spontaneous condensation apparently begins at about eightfold supersaturation and takes place as droplets "so fine they resemble a cloud or fog."

In the presence of the hygroscopic nuclei mentioned above, condensation begins at humidities well below saturation. During the formation of a fog, the relative humidity is observed to increase gradually, and correspondingly the amount of suspended liquid water increases as more and more nuclei attract condensation or as already existing droplets grow larger. It is known that many fogs, especially those in industrial areas, form and exist with humidities well below 100 per cent and humidities as low as 80 per cent have been reported. Accordingly the process of fog formation is a continuous one, in which the time required for formation is governed by the rate of increase of relative humidity and the quantity and type of nuclei present. Figure 1, adapted from Neiburger and Wurtele [16], illustrates this point with data gathered at the Los Angeles Airport under conditions in which sea salt was presumed to supply the nuclei. The authors point out that the visibility becomes approximately constant below a humidity of about 67.5 per cent. At this point they consider that the drops become crystalline and the obstruction to vision becomes essentially a dry haze. It is well known that there are nearly always sufficient numbers of active nuclei present to form fog when the humidity reaches saturation. Supersaturation in the production of fog is therefore generally considered to be limited to the order of one per cent or less when it exists at all.

Willett [24] points out the fact that fogs produced at unusually low humidities are particularly dense and

dry, and consist of an unusually large number of very fine particles. These fogs are slow to be dissipated and very disagreeable to people living in regions where they occur. Of course a substantial part of such a fog must be solid particles of combustion.

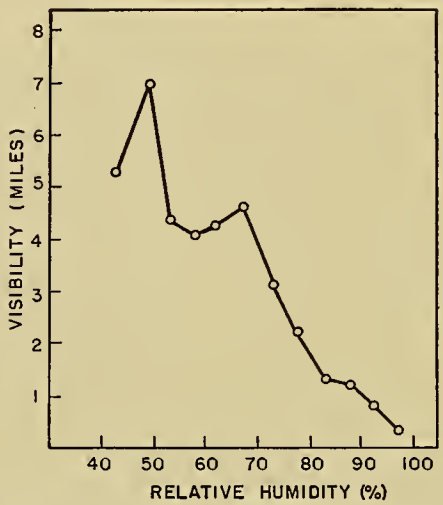


FIG. 1.—The relation between visibility and relative humidity at Los Angeles Airport. (*Adapted from Neiburger and Wurtele* [16].)

Drop-Size Distribution and Liquid-Water Content. Among the most important physical characteristics of fog are the size distribution and water content of the particles comprising it. Although various indirect methods which gave some idea of the drop-size distribution had been known for a considerable time, it remained for Houghton and Radford [13] to perfect a direct method for examining these droplets microscopically. They found that the maxima of the volume distribution curves obtained from forty sets of data taken in sixteen fogs (all of an advection nature, since the location was near the Atlantic Ocean) ranged from 12 to 90 μ. An average of about 30 per cent of the liquid water contained in the fog fell within a 10-μ band centered about the peak of the curve. Individual drop diameters were measured between the extremes of 2 and 130 μ. It should be particularly noted that these figures refer to volume of water and not to frequency of drop diameter. For example, in a given fog the arithmetical mean of drop diameters observed was about 12 μ while the volume mean diameter amounted to 40 μ.

More recently, Heverly [10] has found droplet diameters (in fogs probably purely radiational in character) markedly less than the figures given by Houghton and Radford. In one particular case he cites a range of only 1.5 to 15 μ for drop diameters while in practically every case they examined Houghton and Radford found drops ranging as high as 60 or 80 μ. Heverly [10] observes,

> It appears that coalescence in drop measurements has been overlooked or, at least, underestimated by a number of experimenters. In one instance, when an especially high density of the droplet field was obtained, a great deal of coalescence was microscopically observed. Coalescence took place instantly whenever droplets of water in the vaseline came into contact with one another. Unless the droplets are separated by a distance equivalent to several drop diameters, coalescence appears to be a serious factor and is, probably, the reason for some of the high medians of fog-droplet sizes reported in the literature.

Neiburger and Wurtele found that the most frequent drop diameter for Los Angeles stratus was 14 μ, a value which seems to be in agreement with Houghton's results. Furthermore, these investigators found drops up to 75 or 80 μ, particularly near the base of the stratus, which would presumably correspond with the surface measurements in a fog at ground level. This small but quite possibly important difference in drop-size measurements could of course be due to a difference in techniques as was suggested by Heverly, but it seems more likely that there is a fundamental difference in the characteristics of the fogs examined. Houghton's measurements were made almost entirely in pure advection fog uncontaminated by industrial pollution, while Heverly's apparently were made in the Pennsylvania mountains where industrial pollution is strong and where fogs are almost entirely radiational in character.

The liquid-water content of fogs is one phase of the physical aspects of the matter which appears to be open to little question. Radford's diagram [13, p. 29], reproduced here as Fig. 2, relates the measurements of several independent investigators at various locations to the horizontal visibility, both on logarithmic scales.

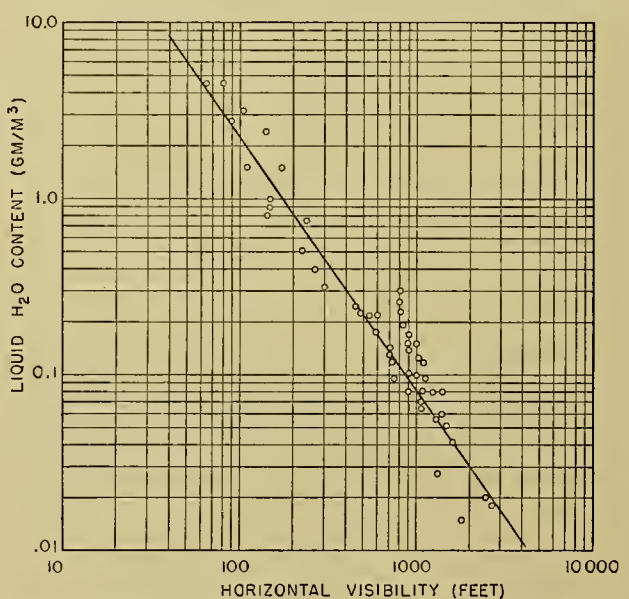


FIG. 2.—Relation between liquid-water content of fogs and visibility. (*After Radford* [13].)

The fit of the curve is surprisingly good for this type of data. It is of interest to note from this curve that visibility will be reduced to less than 1000 m, thus creating fog, by as small an amount of liquid water as 0.02 g m^{-3}.

It seems likely that visibility in industrial areas would be considerably less for corresponding liquid-water contents than is shown by this graph. Willett [24] remarks,

The smaller number of particles is characteristic of the wet and less dense type of fog which consists of relatively large particles. The larger number is the characteristic of the very dense and dry type of low level industrial fog which consists of a great number of very small particles. The wet fog with 10^2 particles per cc may contain more condensed water per cc than the dry fog with 10^4 particles, but for the same amount of liquid water per cc the dry fog of numerous particles is much more effective in reducing the visibility.

Fog over Snow Cover. The formation of fog over snow introduces considerations which are not present in the absence of snow cover. The principal effect is that the snow cover acts as a fog-inhibitor. Petterssen [19] cites climatological evidence to the effect that fogs are rare over snow-covered areas in winter, just when radiational conditions are nearly perfect. His evidence, however, indicates that for temperatures below −40C fogs increase again and are frequent during such low temperature conditions when they are ordinarily composed entirely of ice crystals. Byers [2] comments on the frequency of fog particularly at very low temperatures and quotes from Frost's *Climatological Review of the the Alaska-Yukon Plateau*, "At Fairbanks dense fog always accompanies temperatures of −45F or lower . . . ."

However, important recent evidence is presented by V. J. and M. B. Oliver [18] to the effect that these low-temperature fogs form only around settlements and are almost nonexistent in regions only a few miles removed from them. Pilots accustomed to flying in the arctic winter confirm this, and it seems that many of the accepted statistics concerning arctic fogs at low temperatures are impeached by this evidence. The Olivers offer the hypothesis that smoke particles from the towns act as sublimation nuclei and are the necessary element for the formation of fog under these conditions. They suggest that the smoke contains sublimation nuclei not present in sufficient numbers otherwise. They also point out, however, that the sources of smoke are also sources of moisture which may be of primary importance.

The reasons for the low incidence of fog over snow cover are completely treated by Petterssen [19]. Briefly, for temperatures above freezing he considers that the temperature of the snow surface remains near 0C and that the usual condition of an increase of specific humidity with elevation causes an eddy flux of moisture toward the surface, with the result that condensation takes place on the snow. This removes moisture from the air and acts as an inhibitor of fog but does not entirely prevent it. The higher the temperature above freezing, the greater the dissipating effect. (Although it is not mentioned by Petterssen, it is apparent that this process would have no such dissipating effect upon low stratus clouds.) At a temperature of exactly freezing for both air and snow surface no dissipating influences exist, since the saturation vapor pressure is the same over water and over ice at this temperature. For temperatures below freezing, Petterssen ascribes the dissipating process to the difference between the saturation vapor pressure over water and over ice.

The Olivers [18] mention another dissipating influence on supercooled water fogs.

A mixture of water fog and ice fog occasionally forms at temperatures between zero and freezing, when clearing skies permit rapid cooling by radiation. This condition lasts for only a few hours, during which period the water fog rapidly changes to ice fog or rime and the ice fog thins out and disappears even though temperatures continue to fall. This general weather situation also produces rapid accumulation of hoar frost on trees, wires, and other exposed surfaces. Apparently, the snow surface, trees, wires, etc., are much better radiators than air, and so become much colder than the air, thereby permitting rapid frost deposits, rapid drying of the air, and lowering of the dew point.

It is well known that fogs contain supercooled water down to temperatures of −20C to −40C. The recent work of Schaefer [22] indicates that at −39C fog particles of 10- to 15-μ diameter will freeze without outside stimulus. Köhler [14], some time ago, found that large drops became ice at higher temperatures than small ones, and Heverly [11] has recently produced quantitative results which seem to show that fog droplets with diameters of 50 μ would begin to freeze spontaneously at about −28C or −30C. Even if it is argued that fog droplets are much smaller than this at these temperatures and consequently would have a spontaneous freezing point nearer to Schaefer's −39C, there is still another effect which would lead to a mixture of ice and water particles below −28C. For example, if saturated air at a temperature of −28C is cooled further, as by radiation or upslope movement, the further condensation resulting will [11] "produce relatively fewer and smaller water droplets and more ice crystals." Accordingly we may conclude that nearly all fogs in air at −28C will contain some ice crystals and that the proportion of ice crystals to supercooled droplets will increase rapidly for lower temperatures.

Bergeron [1] has shown that a mixture of supercooled water and ice crystals is "colloidally unstable." Due to the difference in saturation vapor pressure over water and over ice, the vapor pressure over the water droplets would be appreciably greater than that over the ice particles, and a flux would be set up causing evaporation of water particles and an increase in size of the ice particles. There is a tendency for mixed water and ice fogs to change over entirely to ice fogs by this process. Figure 3 is particularly interesting in the light of the above discussion. It is from the Olivers' paper [18] on Alaskan fogs. The almost complete absence of fog at temperatures above −33C and the very rapid increase at lower temperatures seems to add credence to the hypotheses discussed above. All authorities agree that for temperatures below −39C fogs are composed entirely of ice crystals which are in equilibrium with a snow surface and therefore have no tendency to dissipate due to the surface condition.

Physical Processes. The classical concept of the radiative and heat balance of fog situations is about like this: After nightfall, on calm clear nights, the radiational cooling of surface objects is much more rapid than that of the air. Consequently the surface layers of

the air are cooled by conduction, and if this process is carried far enough, dew forms. If there is an increase of specific humidity with altitude, an eddy flux of vapor

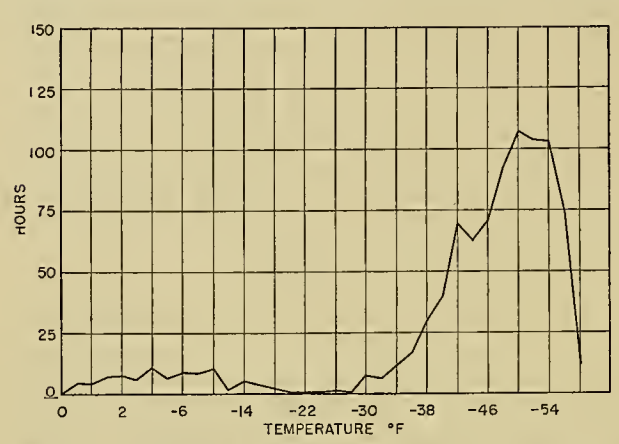


FIG. 3.—Number of hours of fog for various temperatures at Fairbanks, Alaska 1942–1947 inclusive. (*After V. J. and M. B. Oliver* [18].)

downward is established to replace that which has condensed. At the same time, if there is any amount of turbulence at all, heat is conducted toward the surface layers along with the moisture, and some of the cooled air is forced aloft and slightly cooled still further by adiabatic expansion. This process, if continued, is supposed to cool the entire surface layer of air to saturation and thus fog forms. If there is no turbulent interchange of air or if the moisture content decreases vertically, then there will only be a deposition of dew. Obviously the presence of a cloud cover will offer counter radiation to that of the earth and permit so little cooling of the surface layers of air that the possibility of fog is practically eliminated. Likewise if the movement of air is sufficiently rapid for the surface roughness of the area concerned there will be established a turbulent layer deep enough so that more cooling will be required than is available. The intermediate condition between fog formation at the ground and no fog at all (given fog conditions with varying wind velocity) is that in which a stratus cloud deck forms.

Recently, Emmons and Montgomery [5] have pointed out that there is another factor which has not been thoroughly considered. In their words,

> What has not previously been considered in explanations of fog formation is that air with dew point higher than the temperature of the cold surface necessarily loses water vapor by eddy diffusion and real diffusion toward the surface and condensation on it regardless of the nature of the surface. Under the conditions existing when fog forms, eddy diffusion is equally effective in transporting both heat and water vapor. In the laminar layer next to the cold surface, the losses depend on conduction and diffusion, which are effective to nearly the same degree; actually the ratio of thermometric conductivity to diffusivity of water vapor in air is only 0.84. Therefore temperature and dew point both decrease, and although their difference decreases, they cannot become equal, *i.e.*, saturation cannot result, by this process alone.
>
> It must be concluded that fog can form next to a cold surface only when there is further cooling by radiation directly from the air or when saturation is not required.
>
> Likewise, in the opposite case of air in contact with warmer water (whether a water surface or falling precipitation) . . . air receives heat as well as water vapor, so that saturation does not necessarily result. However, in this case three circumstances help bring about saturation and fog; (a) Because diffusion is slightly more effective than conduction, the air gains relatively more water vapor than heat. (b) There may be some heat loss by radiation directly from the air (*e.g.*, when warm raindrops fall through air over cold ground). (c) If the temperature contrast is very large (as in the case of sea smoke), mixing of unmodified air with air modified by the water may produce greater water concentration than is required for saturation at the temperature of the mixture.

Strangely enough, radiation, the most important single element in the formation of fogs, requires less discussion than almost any other aspect of the subject. The reason for this seeming anomaly is not that radiation is of little importance, for without it the only fog formations would be sea fogs, fogs occurring on mountains due to upslope winds (really clouds), and perhaps a few isolated cases of fogs over snow cover. The explanation is, of course, that the difference between various radiative conditions usually amounts to only a very few degrees, and the other factors in the formation of fog make this difference so small that it may usually be neglected.

Suggestions for Research. Deficiencies in our knowledge of the physical processes of fog formation are present at all stages. Beginning with the nuclei, it is apparent that our knowledge of what constitutes them and how they operate is rudimentary. For example, the questions raised by Neiburger and Wurtele [16] concerning the role played by industrial impurities in lowering the relative humidity in fogs appears to cast doubt on at least one accepted idea. The submicroscopic size of these particles makes it difficult to suggest any direct avenue of attack, but perhaps it might be possible to devise laboratory experiments which would answer some pertinent questions concerning the effects of various atmospheric contaminants.

The matter of drop-size distribution appears to be fairly well answered, but the question raised by Heverly should be cleared up, perhaps by a series of observations taken at a number of locations such as mountain valleys and flat areas at some distance from water surfaces of any kind. Furthermore, it might be very instructive to have a series of such measurements made during the formation and dissipation of pure radiation fogs of shallow nature. A series of vertical measurements not only of drop size but of humidity and liquid-water content would also help to reveal the nature of fogs, particularly if the fog area were shallow enough to obtain a complete cross section from dense fog at the base to clear air at the top of the observations.

The question of the number of fog droplets per unit volume has never been accurately answered. Estimates as low as 1 drop cm^{-3} and as high as 1000 drops cm^{-3} have been made. Of course it is true that droplet count varies with the amount and kind of nuclei and with different kinds of fog; however, exact knowledge is

entirely lacking. Probably entirely new techniques will be required to resolve this question satisfactorily.

The formation of fogs over snow cover at very low temperatures probably can be adequately described from available knowledge. However, it would be desirable to have confirming data as to whether ice crystal fogs form over regions remote from human settlements and whether these settlements cause the formation of fog by the addition of otherwise absent sublimation nuclei or in some other way.

The vertical gradient of water vapor in the lowest thousand feet of the atmosphere is something which almost all authorities treat as of considerable importance in the formation of fog. Yet quantitative knowledge of the subject is limited. Some pertinent questions concerning vertical moisture gradients are (1) Exactly what kind of gradient is required for the formation of various densities of fog? (2) How high above the surface do moisture inversions usually extend during fog formations? (3) What rate of flux attends various stability conditions, wind flow, etc.?

It should not be too hard to answer some of these questions at least in part by using existing observation towers with perhaps simple additions to equipment, although the location of these towers with respect to sufficient quantities of fog appears doubtful. It would seem that such observational data might easily be a key not only to broader concepts of the physical process of fog formation but to better methods by which to forecast it.

Synoptic Aspects of Fog

Practically all treatises on fog begin with the author's idea of a classification of fog, usually according to causes. Most such classifications are eminently logical and serve the purpose for which their author intends them adequately and well, and they should be considered according to the user's needs. Willett's table [25], slightly modified by Byers [2, p. 509], is given below and serves as an adequate model, despite the fact that the original version appeared as long ago as 1928.

A. Air-mass fogs.
   1. Advection types.
      a. Types due to the transport of warm air over a cold surface.
         (1) Land- and sea-breeze fog.
         (2) Sea fog.
         (3) Tropical-air fog.
      b. Types due to the transport of cold air over a warm surface.
         (1) Steam fogs (arctic "sea smoke").
   2. Radiation types.
      a. Ground fog.
      b. High-inversion fog.
   3. Advection-radiation fog (radiation over land in damp sea air).
   4. Upslope fog (adiabatic-expansion fog).

B. Frontal fogs.
   1. Prefrontal (warm-front) fog.
   2. Postfrontal (cold-front) fog.
   3. Front-passage fog.

In addition to the classes listed by Willett, the following types appear in Petterssen's scheme [19]: (1) isobaric fog, (2) isallobaric fog, and (3) fog formed by horizontal mixing. All of them (except the last one, under special conditions) are designated by the author as rather unimportant.

If a classification such as these is to be used for forecasting, then it is not only desirable but necessary to eliminate all of the causes listed which are not of direct forecasting value. If this simplification is not made, the forecaster must continually concern himself with rules for the forecasting of various kinds of fog whose causes overlap and which cannot be anticipated by any clear-cut methods. Furthermore, in the case of many types of fog the frequency of formation is so low that any attempt to forecast their occurrence results in continual errors on the side of forecasting a phenomenon which does not occur.

If all of the nonessential types of fog (for practical purposes) are eliminated or modified, only the following classes remain:

A. Air-mass fogs.
   1. Advection types.
      a. Sea fog.
   2. Radiation types.
      a. Restricted heating fog.
      b. Air-drainage, including marsh fog.
   3. Advection-radiation fogs.

B. Frontal fogs.
   1. Pre-warmfrontal fog.
   2. Mixing-radiation fog.

This list is restricted to only six categories of fog, compared with the eleven given in the Willett-Byers classification; fourteen, if the extra types described by Petterssen are added. If the reduction (which is usually made in forecast offices whether or not it is officially recognized) is justified, it simplifies the forecasting problem. Accordingly, a brief consideration of the categories modified or eliminated seems desirable.

Land- and Sea-Breeze Fog. This type of fog is conceded by most authorities to be essentially the same as sea fog. Furthermore its occurrence is distinctly unusual except perhaps in restricted areas of certain sea coasts. Its influence is seldom felt more than two or three miles inland. For these reasons it does not seem that fog of this type merits a separate forecasting division except possibly in very special instances which are better handled as exceptions.

Steam Fogs. Although these fogs, which occur over open water in very cold temperatures, are extremely interesting, they are always shallow and usually do not restrict the visibility to very low values. Furthermore, they are confined mostly to the edges of open water. Byers ascribes them entirely to a difference in vapor pressure between the open water and the air above it of the order of 5 mb. Again, the basic requirement of forecast need is not likely to be present for this type of fog and it is therefore omitted.

Upslope Fogs. Almost every textbook on meteorology

lists this as an important classification of fog. One of the requisites for the formation of this type of fog has been listed by Petterssen as conditional stability of the air mass. If it is not stable, then vertical mixing as the air is lifted will add drier air from above, thus lowering the humidity of the surface air and destroying the chances of fog formation. But it is just in these stable cases that the air flow in the surface layers is most nearly laminar. In this type of flow the mixing of the different layers is at a minimum and consequently, in a typical case, the air arriving at some higher elevation upslope will be far more likely to consist of air which was originally well above the surface, consequently much drier at a point upwind (and downslope) and far from being in prime condition for fog to form. If we take the case in which the wind velocity causes a homogeneous turbulence layer to form, we still must find some method which will cause this layer to gain an abnormal amount of moisture, and strong upslope areas do not usually contain sources of moisture at the surface.

At least in the United States, there is no source of air-mass properties which produces a stable, moist air mass of the proper qualities to form fog by being blown up a geographic slope. Either the air mass is too dry, though stable, or it is moist and unstable. An example of the first type is Canadian *Cp* air, and an example of the latter is *mT* air from the Gulf of Mexico. The region in the United States most noted for upslope fog is the Cheyenne area. Yet, even in 1938, Clapp [3] wrote,

> The importance of humidification and even saturation of the surface air by falling precipitation is emphasized by the fact that only 16% of all fogs studied occurred with a twenty-four hourly precipitation preceding the fog, both at Cheyenne and in surrounding territory, of less than .01 inch. In fact, the majority of fogs occur with precipitation of some form falling before and during the fog.

Clapp also mentions the importance of cloud cover, and it seems reasonable to suspect this factor of being the source of much of the 16 per cent mentioned which occurred without precipitation.

Despite the hallowed spot which the so-called upslope fogs have attained in the minds of most meteorologists, it seems clear that as a primary cause of fog this factor may be completely eliminated from consideration. This is not meant to minimize the importance of upslope cooling in the production of low visibility in pre-warm-frontal precipitation areas, for this is frequently an important factor.

The only place at which this category should receive serious consideration is along the higher portions of a mountain range or even of an isolated mountain. Here, clouds form purely by upslope motion and certainly the resultant fogs, to the comparatively few who experience them, must be classified as pure upslope fog.

Postfrontal (Cold-Front) Fog. As Byers points out, this type of fog is rare and occurs almost entirely when the cold front has become nearly, or quite, stationary, and resembles a warm-front condition in most respects. There are exceptions caused by geography, as might be expected, and they may be important to individual forecasters, but there is little profit to be gained from a detailed consideration of this fog classification other than to recognize the fact of its existence.

Advection-Radiation Fog. One of the biggest sources of confusion to forecasters has always been the breadth with which the term "radiation fog" has been treated. For many years it was the custom of meteorologists to term a fog "radiation" if it formed at night and consequently required the air to lose heat by radiation. George [6] has pointed out the fallacy of this practice. Consider most locations within a hundred or so miles of a temperate latitude seacoast. When conditions are such that air blows from the sea to the land the formation of fog over wide areas of the coastal plains is common and, aside from the immediate seacoast, it forms only at night. When the air flow is from land to sea, under clear skies, no fog forms even under perfect radiative conditions. Which factor is more important, the advection of moist air from the sea, or the loss of heat by radiation at night? Both of these factors are necessary for the formation of fog.

It was for this very common type of fog that the classification *advection-radiation* was established. It is defined as fog occurring in any air mass which has been over an extensive water surface during the daylight hours preceding the night of its formation. Obviously, this definition includes air which was originally colder than the water surface as well as air which was warmer. In the first case, the process of fog formation involves the rapid addition of moisture to the air by its passage over the water and requires strong nocturnal cooling over land to stabilize the lower layers before fog can form. In the latter instance of warm air passing over cold water, evaporation is slow or nonexistent, but the combination of negligible heating by solar radiation while over the water, together with the cooling and stabilizing effect of the cold water surface on the lower layers of air, prepares these layers for fog by requiring little or no nocturnal cooling over land.

It is apparent that, following a strict classification according to causes, two different processes are represented here. For the purpose of scientific investigation it would be necessary to recognize two types, but for forecasting purposes they may conveniently be grouped together in one category and called advection-radiation fog.

Radiation Fog. The question next arises as to exactly what types of fog should be called "radiation." From the same source as the preceding discussion it seems reasonable to adopt the convention that pure radiation fog is that which forms in air that has been over land during the daylight hours preceding the night of its formation. Furthermore, there are two main conditions favorable for radiation fog: (1) The air has been under a cloud cover with or without precipitation falling through it during the day previous to its formation. (2) Pools of air cooled to an excessive degree have collected in valleys, causing air-drainage fogs. Although they are not quite the same, it appears reasonable also to place in this category those fogs which form in low marshy land and along river valleys on calm, clear

nights. This definition of radiation fog carries with it the implication that such fog does not form unless one of the two conditions mentioned above is present.

Some misunderstanding of these principles has appeared; for instance, O'Connor [17] specifies a nonconvective cloud cover, and states that fogs formed as a result of this type of restricted heating are usually post-coldfrontal. In reality, a large proportion of cloud covers which result in radiation fogs are strictly convective in origin, and although practically all fogs in clear air behind cold fronts are formed from this cause, they are very few indeed when compared with the total number of radiation fogs caused in this way.

For forecasting purposes there is little point in distinguishing between ground fog and stratus as two different types of fog, for in nearly all cases the forecasting problem is identical. Whether ground fog or stratus results is entirely dependent upon (1) the terrain, (2) the characteristics of the air mass, and (3) whether the wind velocity exceeds a certain value. Accordingly, in the forecasting classification "ground fog" and "high inversion fog" are replaced by "restricted heating" and "air drainage fog."

Forecasting Methods

The forecasting of fog is, above all other considerations, a local problem. There are many instances in which airport sites which have been moved only a few miles required definite revision of methods for forecasting fog. This is especially true of locations at which an advective element is present. Each season usually requires a new technique, and the duration of the season must be determined from local data.

Advection Fog (Sea Fog). As the name implies, fog of this type forms over the open water as a result of the advection of warm moist air over colder waters. This process obviously stabilizes the air and if continued long enough cools the air to its dew point and fog is formed. The forecasting of these occurrences for sea areas can best be done by combining a computation of future trajectories [2, 7] at various points with a careful plot of mean sea-surface temperatures.

Radiation (Restricted Heating) Fog. There is only one type of fog which can be forecast by using a single tool for most localities, and even this type must be hedged with restrictions. The tool is George's restricted heating radiation fog graph. This method has been tested at various locations in the eastern United States and found to be valid at all locations tested which (1) were not within about 50 miles of a seacoast or other large body of water (including some large rivers) or (2) presented no severe air drainage problems. Figure 4 shows the general form of the radiation graph from Roche's study [21] of Carolina stations. It is not entirely clear why a locality near a seacoast does not experience fog caused by restricted heating in the same degree as inland stations, but it seems likely that land-breeze effects at night added to the normal divergence caused by decrease in friction as the air moves from land to sea may very well thin the moist layer enough to overcome all but the most ideal situations.

The presence of precipitation as a primary element in the formation of fogs has been recognized for some time. Counts [4] cites it as a basic precept in forecasting

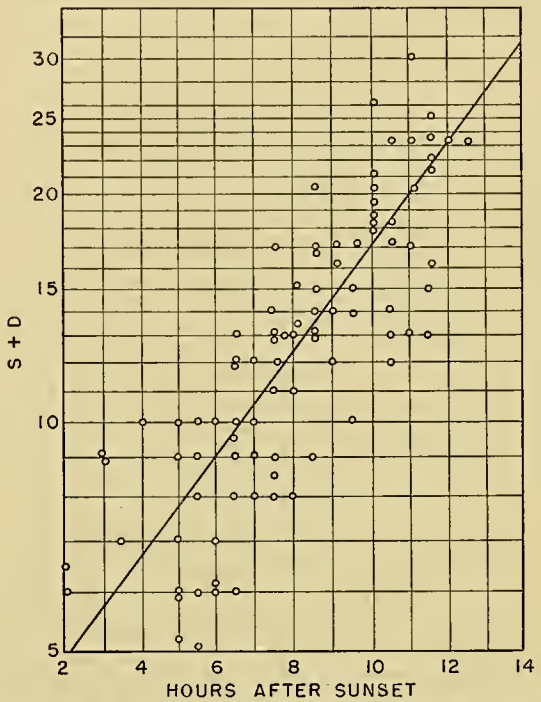


FIG. 4.—Radiation fog graph devised by George and Bradley. Data are by Roche, for Carolina stations. S + D (ordinates) are the sum of the number of sunshine hours during the day plus the number of degrees between temperature and dew point at sunset. The units are mixed but retained in this inaccurate form for practical convenience. Abscissas give number of hours after sunset that fog should form.

fog in California's San Joaquin Valley, and earlier in this article it was pointed out that Clapp found it an almost essential factor for the formation of fog at Cheyenne, Wyoming. Precipitation occurring during the day aids in the formation of fog in two ways: (1) it adds moisture to the lower air layers, and (2) the attending clouds act to prevent the normal diurnal rise in temperature.

At one location (Atlanta, Ga.) a study [8] has been made of the difference in fogs and stratus clouds caused by daytime precipitation conditions and those formed when only cloudiness (without precipitation) was present. Figure 5 illustrates this difference. For geostrophic winds of 15 mph, the resulting stratus clouds in rain situations have an average minimum ceiling of about 200 ft (Fig. 5*a*) and not infrequently they may lower to the surface as dense fog. For the same wind velocity, Fig. 5*b* shows the minimum ceiling in "no rain" situations to be about 700 ft. For geostrophic winds below 12 mph the "no rain" graph indicates that the fog forms and remains a true fog, that is, one at ground level. According to these data, the effect of rain is to add enough moisture to the lower layers to offset to some extent the dissipating effect of higher wind velocities.

Radiation (Air-Drainage) Fog. At various locations, unusual geographic conditions exist which cause the

formation of fog under circumstances that otherwise would not produce fog. The air-drainage fogs which cause the valleys of the Allegheny chain to have one of the highest incidences of fog in the United States [23]

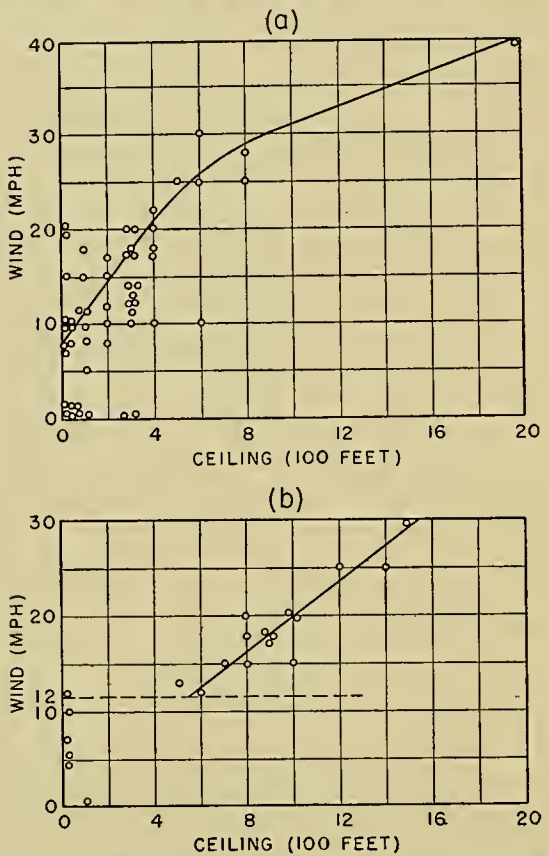


FIG. 5.—(*a*) Forecasting graph for restricted heating fog with rain in the trajectory (for Atlanta). Ordinates are geostrophic wind velocities scaled from the noon chart prior to the night of fog formation. Abscissas give the minimum ceiling attained the following morning. (*b*) Same as (*a*) but no rain in the trajectory.

are caused by geographic conditions. Another example is Charleston, W. Va., which is in a river valley opening to the Ohio Valley toward the northwest. Fog results when anticyclonic conditions prevail in such a manner as to prevent flow of air down the river valley and consequently make air drainage the paramount feature. In many cases, no restriction of heating is concerned; it may be, and frequently is, clear all day at the station and to windward. Methods of forecasting involve the use of the gradient wind direction and the moisture content.

River valleys, especially low-lying marshy ground immediately adjacent to the river, are frequently subject to fogs on calm nights despite a lack of cloud cover during the day or any special air-drainage characteristics. Such areas require special treatment for forecasting, but so far as is known no location of this sort has received much study.

Advection-Radiation Fog. Stated simply, advection-radiation fog is a fog which requires both factors before it may form. The description which follows is, with a few modifications, that given by George [6].

The area over which advection-radiation fogs may form is immediately and clearly defined as that portion of the coastal plains which the lower air layers may reach in one night's travel from the water. A practical limit to gradient winds which allow fog to form is about 30 mph but almost never will this value exist both for the full distance inland and the entire time consumed in traveling this distance; the average transport of air will also be much less because of surface friction. In practice, about 200–250 miles during winter and 150 miles in summer were found to be the inland limits. This allows a maximum average transport of air amounting to 14–18 mph.

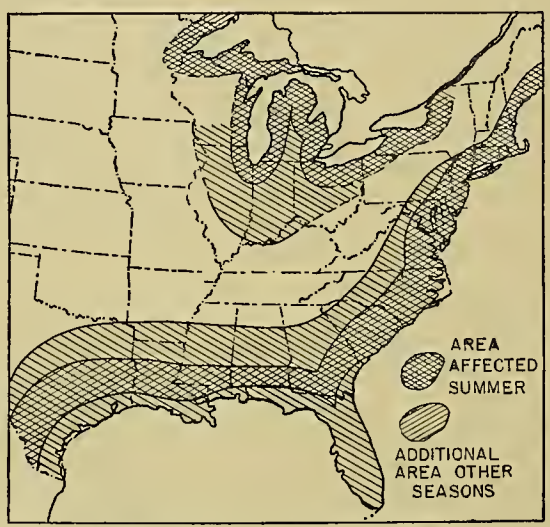


FIG. 6.—The distribution of advection-radiation fog in the eastern United States, showing how the areas affected increase during the portion of the year which has longer nights.

In Fig. 6, the double hatching indicates the area along the Great Lakes and the eastern coast of the United States which is subject to advection-radiation fog in summer and the single hatching indicates the additional area affected during the remainder of the year. During summer, there is a strip 30–50 miles wide along the Gulf and Florida coasts in which no fog forms. The reason this exists is probably because gradient winds from sea to land, sufficient to overcome the tendency for land-breeze formation, allow insufficient time for stabilization caused by radiational cooling to take place. If the land breeze does develop, the accompanying subsidence and other effects are sufficient to prevent the formation of fog. During the colder months, and even in summer farther north, the colder shore waters precool the air, and fog may form even along the coast with reasonable wind velocities. During winter, fall, and spring, the area which is subject to such fog is considerably extended, but it should be noted that this area has nothing to do with the frequency of formation, and, for instance, north of latitude 37°, advection fogs are rare in winter.

The accurate forecasting of advection-radiation fogs is difficult and varies markedly from place to place. Figure 7 illustrates how the gradient wind, no doubt controlled by irregular land and water surfaces, varies

during fog conditions at New Orleans Airport [12]. Other localities require more complicated treatment than this. For instance, Figs. 8*a*, *b*, and *c* show the

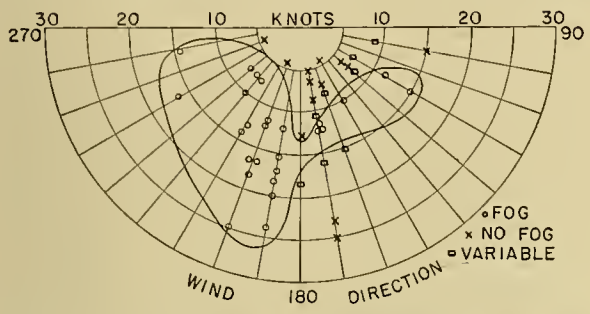


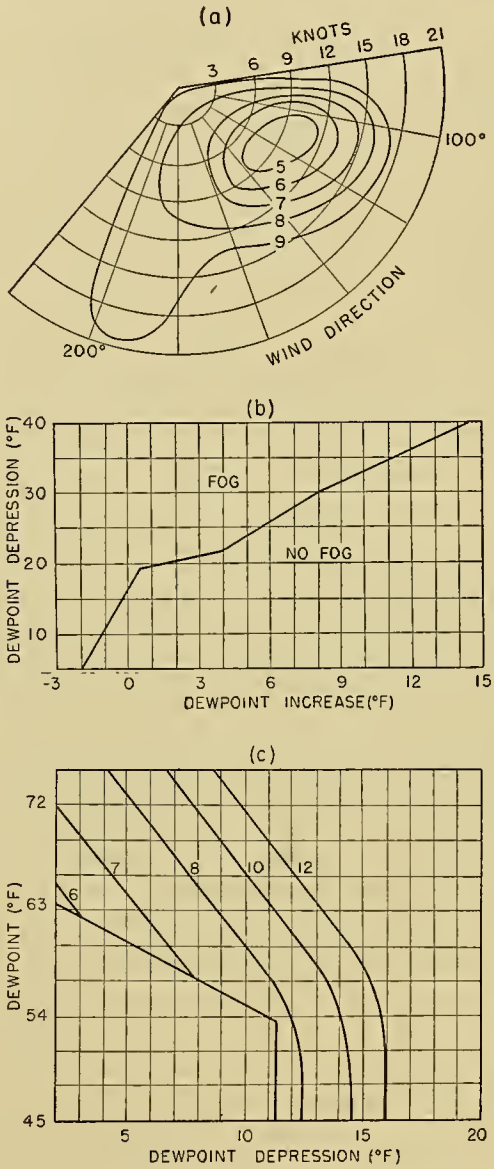
FIG. 7.—A typical method for forecasting advection-radiation fog for New Orleans Airport during winter (*after Hilworth* [12]). The data are gradient winds taken from the afternoon upper-wind sounding.

steps required with present methods for forecasting this type of fog at Charleston, S. C., during the fall months. Other and still different methods are used for the remaining seasons.

Pre-Warmfrontal Fog. The forecasting methods for this class of fog usually given in textbooks are rather general in nature and sometimes they are omitted altogether. It is difficult to find much literature on the subject which would be of real assistance to an investigator wishing to learn how to forecast these occurrences. Perhaps the biggest difficulty is the vaguely recognized fact that local geography again plays a dominant role. For example, pre-warmfrontal fogs occur very rarely at Cleveland, probably due either to downslope compressive heating in the cold air or to the frictional divergent effects caused by the proximity of Lake Erie, or to both. At this locality, the clouds consistently remain well above the ground in pre-warmfrontal situations, while at Atlanta, Washington, D. C., and Chicago, fogs formed in this way are common.

Some rather ingenious methods have been utilized for forecasting the formation of low stratus in warmfront situations. An example is Roche's graph [21] for Charlotte, N. C., shown in Fig. 9. In nearly all such methods, however, the authors have used synoptic material which was readily available in a practical sense. Not much effort has been made to apply quantitatively any relationship between temperature of falling precipitation and the surface air layers, which Petterssen has pointed out to be very important to the problem. Also, consideration of the stability of the cold air has remained largely a qualitative problem.

Mixing-Radiation Fog. This is not at all a usual type, but it is important to recognize the significant role played by dissolving fronts at many locations. Particularly in late spring and summer there are many areas which have little or no fog not associated with a weak frontal condition. Whether the formation of this kind of fog involves actual mixing of two dissimilar air masses is open to considerable question, but nocturnal radiation is certainly a requirement. Furthermore, there are many occasions when fog does form in a narrow band along a dissolving front even when no daytime clouds attend it. Forecasting methods are very general for this unusual class and depend largely upon climatology and the individual experience of the forecaster.



FIGS. 8*a*, *b*, *c*.—These graphs indicate a representative solution for advection-radiation fog a little more complicated than most. They are for winter cases at Charleston, S. C. (*after George*). Figure 8*a* takes care of the gradient wind requirement as well as giving a preliminary estimate of time of formation from the numbered contour lines (in hours after sunset). Figure 8*b* eliminates certain instances which would otherwise produce an erroneous forecast for fog. Ordinates are dew-point depressions at maximum temperature and abscissas are obtained by subtracting the temperature of the dew point at maximum temperature from that at sunset. Figure 8*c* provides the humidity element and another estimate of timing.

Timing the Clearing of Fog or Stratus. Several methods have been devised to aid in estimation of the time required to dissipate fog or stratus. Probably the simplest and most direct method is the relationship between thickness of the stratus and time required to clear, given by Wood [26] and reproduced here as Fig. 10.

This method, of course, has reference only to stratus clouds and requires nothing except a knowledge of the cloud thickness, which is usually available to the fore-

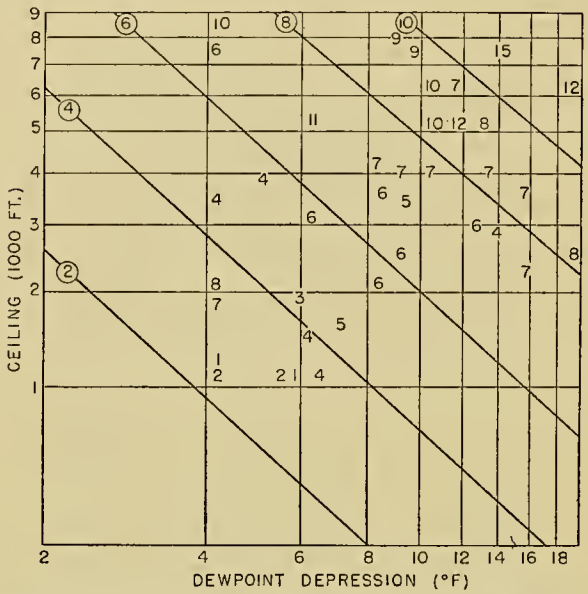


Fig. 9.—A method of forecasting pre-warmfrontal fog or low stratus (*after Roche* [21]). The graph is entered with the dew-point depression and the ceiling at time of beginning precipitation. Sloping lines give time in hours before the stratus forms below 600 ft, as designated by the circled numerals.

caster. Krick has devised a method based on the use of (1) a recent sounding in the vicinity for which the information is desired, and (2) normal surface-heating curves during stratus conditions. The requirement of having both the sounding and the heating curves prevents very widespread adoption of what would otherwise be a reasonably precise method.

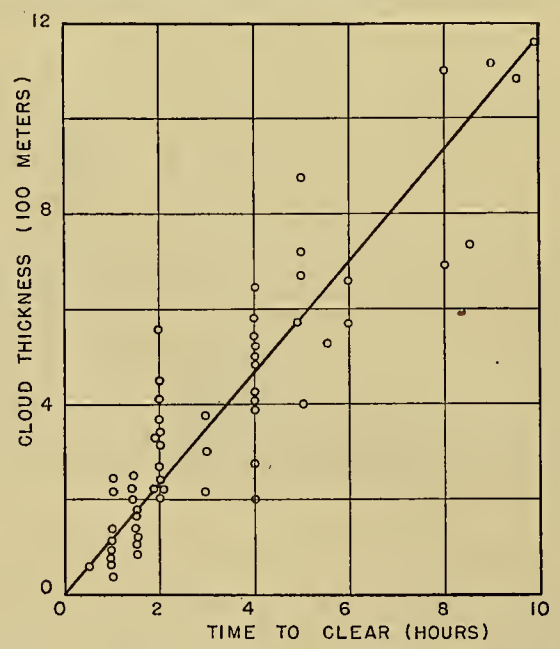


Fig. 10.—The relation between time required for stratus clouds to clear and their thickness (*after Wood* [26]) with data from several coastal stations.

Still another method (Fig. 11) has been used as an approximation for forecasting the dissipation of fog [9]. This is an indirect method of taking into account the

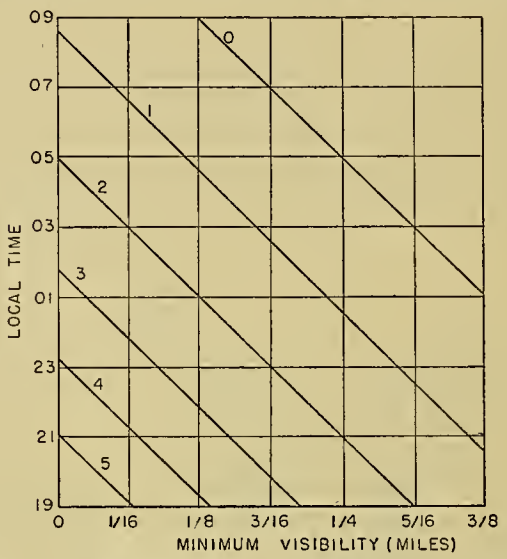


Fig. 11.—Another illustration of method for predicting the clearing of fog. Sloping lines are the number of hours after sunrise required for clearing. (*After George.*)

amount of liquid water which must be evaporated (see Radford's visibility scale, Fig. 2), and the total amount of foggy air measured by the length of time the fog has been present. This method has the advantage of being extremely practical and dealing with easily obtained data, but possesses the disadvantage (which may be inherent in all methods) of requiring the clearing forecast to be made at a time when complete information about the fog is available, or may be anticipated. In none of these methods is any account taken of the nuclei characteristics; it generally is presumed that industrial pollution is not present to a marked degree.

Evaluation and Recommendations. In perspective, it seems clear that our knowledge of the synoptic conditions under which fog forms is far from complete. It is almost as clear that the framework of classification of fog by causes is laid on a solid foundation of knowledge. This is fortunate indeed, for classifications must be accurate before the various causes can be isolated and objectively examined. Fog is so intimately connected with local geography that its forecasting must remain on a local basis. Much progress has been made in the past decade toward the establishment of local synoptic studies in fog, but much more remains to be done and refinements of existing studies are easily possible and highly desirable.

In addition to continued work with local studies, which seems to offer the greatest immediate return for effort expended at present, there are several courses of investigation which should be pursued if any really new techniques in dealing with this phenomenon are to be developed. Of primary interest is the desirability of studying in great detail the vertical structure of the lowest layers of the atmosphere to determine the distribution of temperature and humidity and their

changes during the formation of fog. It is true that good work has been done along these lines at the University of California at Los Angeles, but it has dealt with a very specialized type of fog—the Southern California stratus. Elsewhere, the almost universal rule is that no use whatever is made of these elements either in understanding the causes of, or in forecasting, fog. It is highly significant that radiosondes are almost never used by workers in this field, yet this approach must offer rich possibilities. Perhaps, however, such use would require a method of obtaining more detailed information in the lowest 100 mb of the atmosphere.

Winds in the lowest few thousand feet have been used extensively in synoptic fog research and are of obvious importance. Yet our knowledge of the diurnal changes in wind direction and velocity and, what is really the same problem, the relation of vertical wind sheer to stability considerations, is still far from a practical solution.

A field that might be productive to some extent is that of surface characteristics. The different radiative properties of various ground covers and the day-to-day variations in lake and ocean temperatures, while certainly not neglected, have never received the attention which they perhaps merit.

It is particularly true of fog, as it is generally of most phases of meteorology, that no great discovery is about to solve all the problems, or even cause any sensational advance. Continued work on the physical aspects of the problem will be necessary for advancement and, at the same time, a great deal of so-called "pick and shovel" work in applied synoptic meteorology is absolutely essential if real progress is to be made.

REFERENCES

1. Bergeron, T., "On the Physics of Cloud and Precipitation." *P. V. Météor. Un. géod. géophys. int.*, Lisbonne, 1933, II, pp. 156–178, Paris (1935).
2. Byers, H. R., *General Meteorology*. New York, McGraw, 1944.
3. Clapp, P. F., "The Causes and Forecasting of Fogs at Cheyenne, Wyoming." *Bull. Amer. meteor. Soc.*, 19: 66–73 (1938).
4. Counts, R. C., Jr., "Winter Fogs in the Great Valley of California." *Mon. Wea. Rev. Wash.*, 62: 407–410 (1934).
5. Emmons, G., and Montgomery, R. B., "Note on the Physics of Fog Formation." *J. Meteor.*, 4: 206 (1947).
6. George, J. J., "On the Technique of Forecasting Low Ceilings and Fog." *J. aero. Sci.*, 8: 236–241 (1941).
7. —— "Fog, Its Causes and Forecasting with Special Reference to Eastern and Southern United States." *Bull. Amer. meteor. Soc.*, 21: 135–148, 261–269, 285–291 (1940).
8. —— *Further Studies in Air Mass Fog at Atlanta, Part II, Radiation Fog*. Atlanta, Eastern Air Lines Meteor. Dept., 1948.
9. —— *A Brief Forecasting Study for Charleston, S. C.* Atlanta, Eastern Air Lines Meteor. Dept., 1947.
10. Heverly, J. R., "Similitude of Artificial and Natural Fogs." *Trans. Amer. geophys. Un.*, 30: 15–18 (1949).
11. —— "Supercooling and Crystallization." *Trans. Amer. geophys. Un.*, 30: 205–210 (1949).
12. Hilworth, J. T., *A Short Study of Air Mass Fog at Moisant Airport, New Orleans*. Atlanta, Eastern Air Lines Meteor. Dept., 1948.
13. Houghton, H. G., and Radford, W. H., "On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 6, No. 4 (1938).
14. Köhler, H., "Zur Thermodynamik der Kondensation an hygroskopischen Kernen nnd Bemerkungen über das Zusammenfliessen der Tropfen." *Medd. meteor.-hydr. Anst. Stockh.*, Vol. 3, No. 8 (1926).
15. Landsberg, H. "Atmospheric Condensation Nuclei." *Beitr. Geophys.*, (Supp.) 3: 155–252 (1938).
16. Neiburger, M., and Wurtele, M. G., "On the Nature and Size of Particles in Haze, Fog, and Stratus of the Los Angeles Region." *Chem. Rev.*, 44: 321–335 (1949).
17. O'Connor, J. F., "Fog and Fog Forecasting" in *Handbook of Meteorology*, F. A. Berry, Jr., E. Bollay, and N. R. Beers, ed. New York, McGraw, 1945. (See p. 733)
18. Oliver, V. J., and Oliver, M. B., "Ice Fogs in the Interior of Alaska." *Bull. Amer. meteor. Soc.*, 30: 23–26 (1949).
19. Petterssen, S., "Some Aspects of Formation and Dissipation of Fog." *Geofys. Publ.*, Vol. 12, No. 10 (1939). (See pp. 20–21)
20. —— *Weather Analysis and Forecasting*. New York, McGraw, 1940.
21. Roche, R. D., *Weather Forecasting in the Piedmont Province*. Atlanta, Eastern Air Lines Meteor. Dept., 1947.
22. Schaefer, V. J., "The Production of Clouds Containing Supercooled Water Droplets or Ice Crystals under Laboratory Conditions." *Bull. Amer. meteor. Soc.*, 29: 175–182 (1948).
23. Stone, R. G., "Fog in the United States and Adjacent Regions." *Geogr. Rev.*, 26: 111–134 (1936).
24. Willett, H. C., *Descriptive Meteorology*. New York, Academic Press, 1944.
25. —— "Fog and Haze." *Mon. Wea. Rev. Wash.*, 56: 435–467 (1928).
26. Wood, F. B., "The Formation and Dissipation of Stratus Clouds beneath Turbulence Inversions." *Bull. Amer. meteor. Soc.*, 19: 97–103 (1938).

PHYSICAL AND OPERATIONAL ASPECTS OF AIRCRAFT ICING

By LEWIS A. RODERT

National Advisory Committee for Aeronautics

Introduction

When aircraft are exposed to certain meteorological conditions, ice forms on exposed areas in a manner that impairs the efficiency of the components and reduces the usefulness of the craft. Because most aircraft are vulnerable to such ice formations, one of the limiting conditions for all operations is that ice protection must be provided or flight must be restricted to non-icing weather conditions.

Icing conditions vary widely in intensity. Fortunately, the conditions that create a serious hazard are not frequently encountered except in limited geographical locations. The problem thus presented by the formation of ice on aircraft has been met by extensive research and development of means whereby ice protection may be afforded and by meteorological investigations conducted to improve the accuracy of weather forecasting and to aid pilots in avoiding local regions of serious icing conditions. The results of research on ice-prevention methods are extensively reported in aeronautical engineering literature to which reference will be made in the text which follows. Results of the meteorological research have also been reported. They are analyzed and summarized elsewhere in this Compendium.[1]

An analysis of the physical phenomena of ice formations on vulnerable components of the airplane under various atmospheric conditions is presented in this paper. This analysis deals with the atmospheric states and aerodynamic and thermodynamic phenomena in or near functioning airplane components during commonly experienced icing occurrences. The analysis is limited to subsonic flight conditions at altitudes below the tropopause. A description of the effects of ice on the airplane components is also presented with limited quantitative evidence and reference sources. Data relating to the impairment of military equipment by icing are omitted for security reasons. Significant research and development work yet to be accomplished on the airplane-icing problem are enumerated in the closing discussion.

Although the collection of snow in ducts or other frontal openings and the impact of hail cause the impairment of certain components, and even structural damage, these problems are not discussed in this analysis.

The preparation of this report has been made possible by the results of extensive ice-prevention and meteorological researches conducted by the National Advisory Committee for Aeronautics, the U. S. Weather Bureau, other governmental agencies, private industrial and university laboratories, and the laboratories of several foreign governments.

Analysis of Aircraft Ice Formation

The formation of ice on an airplane requires that water come in contact with the surface of the vulnerable airplane component when the surrounding temperature is below 32F and under conditions that provide for the dissipation of the heat of fusion liberated during the phase change to ice. The subsequent analysis concerns the presence or formation of liquid water in the atmosphere near the airplane, the impingement of water droplets on the airplane surface, and the droplet environmental temperature and thermal dissipation required in the phase change to ice. Although most occurrences of icing are explainable on the basis of these factors, several unique types of icing involving other considerations are treated as individual cases.

Availability of Liquid Water. The natural existence in the atmosphere of cloud droplets at temperatures below 32F is responsible for most of the external ice accretions on aircraft. Observations in the atmosphere [11] show that water will remain in the liquid phase in clouds over a wide range of temperatures below the normal freezing point.

Liquid-water droplets existing at temperatures below 32F are commonly called supercooled droplets, the degrees of supercooling being the difference between the melting point of ice and the ambient temperature. It is tacitly assumed that the droplet temperature is substantially the same as the surrounding air temperature. It is probable that liquid cloud droplets can occur at all temperatures found in the normal atmosphere below the tropopause. Observations have also shown [11] that cloud droplets effective in producing icing are usually no greater than 50 μ in diameter; in fact, the typical cloud droplets during icing conditions are from 8 μ to 20 μ in diameter.

Water droplets at temperatures below 32F may also be generated by normally operating airplane components. Consider clear air expanding rapidly along an imaginary stream tube from a condition of near water-vapor saturation. Such a condition may be found in the air stream that enters an engine carburetor when the throttle is partly closed. The geometry and flow pattern for this condition are illustrated in Fig. 1.

As nearly saturated air moves along the stream with decreasing pressure, the temperature decreases according to the adiabatic process, and the total temperature at region 1 (Fig. 1) on the downstream side of the throttle is substantially the same as that of the ambient region 0 until the temperature of water-vapor saturation is reached.

1. Consult "Meteorological Aspects of Aircraft Icing" by W. Lewis, pp. 1197–1203.

Where the flow degenerates to a turbulent condition, such as at regions 2 (Fig. 1), the temperature of the gas particles is substantially equal to that of the fluid

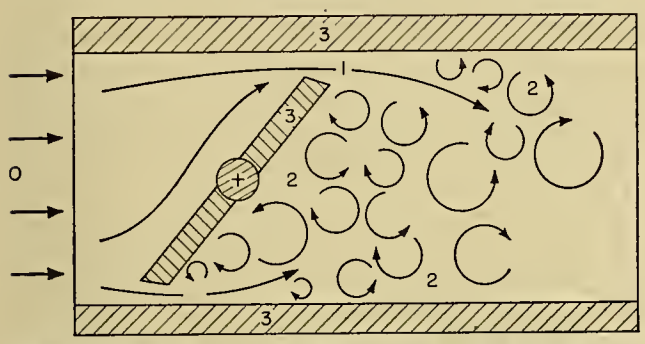


FIG. 1.—Flow through a partially closed carburetor; 0—ambient region, 1—laminar flow at high velocity and low pressure, 2—separated turbulent flow, 3—solid boundary.

in the moving free stream at region 1. The relations between pressure, temperature, and the liquid water precipitated because of expansion along the stream tube are given in a practical and usable form in the pseudo-adiabatic chart as developed by the U. S. Weather Bureau. Because of the velocity of the stream in region 1 and the time interval required for a droplet to grow to a significant size, visible condensation may not occur near the throttle in the moving stream but may be seen at some distance downstream. Droplets may reach sufficient size to cause an icing problem, however, in the separated turbulent regions 2 because of the comparatively stagnant movement through these regions and the greater time interval consequently available for droplet growth. Under certain conditions, vorticity may develop in the standing regions. This will be conducive to the precipitation of more water because of the reduced stream temperature within a vortex. It can be assumed that the temperature of the droplets so formed conforms closely to the stream temperature and that in most cases, including the example of the engine carburetor, the contents of the air-borne droplet do not undergo the phase change to ice until mechanically disturbed as when the droplet strikes the walls of the carburetor.

The expanding flow of clear saturated atmospheric air over an airfoil such as illustrated in Fig. 2 will result

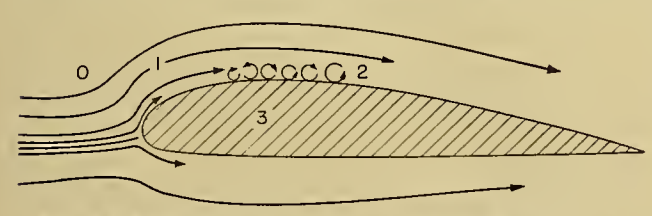


FIG. 2.—Flow over an airfoil; 0—ambient region, 1—low-pressure region, 2—turbulent flow, 3—solid boundary.

in a supersaturated condition in region 2 for the same reasons as those given in the discussion of the carburetor; however, the comparatively high velocity in the turbulent regions over exterior aerodynamic forms prevents droplet growth to a significant size. In wing and propeller tip vortices, visible condensation does form, however, because of the length of time a gas particle remains in the vortex and because of the low temperature which probably exists at the vortex core.

It is therefore evident that supercooled water droplets from which ice may form are found in nature and may also be generated by the aerodynamic and thermodynamic processes of stream flow over the airplane components. Ice is also formed from cloud droplets at temperatures above 32F, although under a combination of operating and weather conditions not often encountered. Snowflakes also are intercepted by airplane surfaces to a very limited extent, and ice formations are thereby accumulated. These special cases of icing and the precipitation of moisture on airplane surfaces from the gaseous state and the subsequent formation of frost are briefly discussed below.

Water Impingement. As an airfoil moves through a cloud, some of the droplets in the volume of space swept through by the airfoil are intercepted by the moving body. The analysis and explanation of the droplet impingement on an airfoil are made for the hypothetical condition in which air passes over a stationary airfoil. In such a flow pattern (Fig. 2), gas particles of the atmosphere follow streamlines past the airfoil. Those passing near the forward stagnation region of the body have small radii of curvature. The inertial effect tending to keep the droplet moving straight in such a curving flow field will provide the force necessary to overcome the viscous forces which tend to hold the droplet motion to that of the air streamlines. The result is that droplets will cross streamlines in the direction of the solid boundary of the airfoil and some will actually strike the surface in the vicinity of the leading edge. In the preceding description of cloud droplets it was noted that the average droplet size is very small, that is, of the order of 8 μ to 20 μ. Furthermore, the velocity of the droplets across the streamlines is small. It follows that the Reynolds number of the droplet motion in the air will be small and that the droplet path can therefore be analyzed by the equations of Stokes' flow, from which it follows that the resistance to motion is proportional to the velocity across the streamlines.

The development of the theory of cloud-droplet interception by airfoils is given in [7, 10], from which quantitative impingement rates and area of catch for particular conditions may be calculated. A method of determining the rate of droplet impingement on an airplane windshield is given in [6]. A discussion of the ingestion of water droplets into the carburetor of an engine is discussed in [9]; however, no attempt has been made to analyze the efficiency or the location at which ingested water comes in contact with the walls or throttle plate of a carburetor.

The area over a component on which the droplets impinge is determined by the liquid-water content of the atmosphere, the range of droplet sizes in the cloud, the velocity, and the shape and size of the airplane component. As the radius of curvature of the entering

surface increases, the percentage of the droplets caught in the swept-out volume decreases and the percentage of the total body area wetted decreases. Increases in the radius of the leading edge also have the effect of limiting the impingement area to the region near the forward stagnation pressure point.

Conditions for Ice Formation. When supercooled droplets impinge upon the surface of an airplane component, the formation of ice follows at a rate determined by the temperature gradient in the immediate environment and the thermal conduction through the gaseous and solid boundaries. It is apparent that the average environmental temperature must be below the melting point of ice. If the droplet freezes quickly upon impact, the shape of the formation will be determined by the droplet impingement pattern; however, if freezing is delayed somewhat after impact, part of the droplet will flow along the solid surface and freeze at a point farther in the rear. The time required for the droplet to freeze therefore determines the shape and the location of the ice, both of which are significant factors in determining the effect of ice on airplane performance.

The temperature of the air particle on the surface of an airplane differs from that of the ambient region because of the velocity of the stream; this departure is a function of the second power of the velocity in addition to other factors. Factors in the boundary layer that determine the temperature of the air particle on the surface include the conversion of the kinetic energy of the gas particle to thermal energy, the conversion of mechanical energy or work to thermal energy in the viscous boundary layer, and the conduction of heat through the gaseous boundary layer. The combined action of these factors determines the recovery temperature of an air particle on the surface of a body in an air stream; however, it should be noted that the heat involved in the phase changes of the water droplet on the surface and conduction through the solid boundary have thus far not been considered.

When the supercooled water droplet strikes the solid boundary, the mechanical disturbance to the droplet causes the phase change to ice. If it is assumed that there is no exchange of heat with the environment and that the droplet temperature rises to 32F during the freezing process, the percentage of the droplet undergoing the phase change will be proportional to the ratio between the degrees of supercooling of the droplet when it struck the surface and the latent heat of fusion.

After impingement, evaporation from the droplet starts. The rate of evaporation is determined by the difference between the vapor pressures at the droplet-air interface and in the ambient region. The evaporation of water from the droplet after impact absorbs heat. (It is to be noted that the temperature and vapor pressure at the droplet-air interface are interrelated and that calculations for one must start with an assumed value for the other from which increasingly accurate evaluations may follow by successive approximations.)

The flow conditions that may exist along the surface of an airplane component such as a wing include the stagnation region at the leading edge, a laminar region also near the leading edge, a turbulent region following the laminar one, and in some cases, a turbulent separated-flow region. It may be concluded from the foregoing discussion that the temperature on the air-water interface of a droplet is affected by frictional heating, by conduction through the boundary layer, and by heat absorbed in evaporation. These effects differ in magnitude and in relation to each other in the different flow regions. It therefore follows that the temperature of the air-water interface differs for various points on the surface of a component along a line parallel to the stream flow, because the only other factor arising from the moving stream—the conversion of kinetic energy to thermal energy—remains constant at all points on the body. It may be further concluded, therefore, that the rapidity with which the water is converted to ice varies for various locations on the surface of an airplane. Inasmuch as all the factors involved in these variations are functions of the velocity, increasing the velocity increases the variations. It is further apparent that if different points on a component move at different velocities (*e.g.*, the points along the radius of a propeller or helicopter blade), still another factor exists that causes surface-temperature variations and therefore variations of freezing conditions.

The discussion has been presented without a consideration of the flow of heat through the solid boundary. Obviously, if heat is applied or taken away through the solid boundary, the conditions for icing will be affected. When the solid boundary is a thermal conductor, which it usually is, variations in the departure from ambient temperature on the air-water interface or on a component surface produce thermal gradients in the solid boundary layer and a flow of heat from the hottest region to the coldest. Thus heat flows from the tip of a metal propeller blade to the hub region. Likewise, because the recovery temperature on the leading edge of an airfoil at the stagnation point is higher than in the laminar region, heat will flow rearward in a metal skin, and icing will be more rapid at the leading edge than would be explainable by the conditions of that local region.

Although of minor importance, the heat from the sun and other sources of radiant thermal energy and the kinetic energy of the impinging water droplet when converted to thermal energy also affect the conditions for icing on the surface of an airplane.

Because not all of the droplet will freeze immediately in most encounters with icing, the liquid water will be subject to abrasion by the air stream, which may blow away in the boundary layer a part of the droplet remaining in the liquid phase or move some of the liquid along the airfoil surface in the direction of the moving stream. The smoothness of the surface will affect this factor. Also, part of a droplet, particularly if it is a large droplet, may bounce away from the solid boundary (in part or entirely) upon impact and re-enter the boundary air. It is possible that droplets contained in the laminar streamlines of the boundary layer and near the leading edge which have zero or negative

velocity components in the direction of the solid boundary may be brought back into contact with the wing surface on the areas that are adjacent to the region of turbulent separated flow (region 2 in Fig. 2). Experimental evidence on the path and history of individual droplets after the first impingement has occurred, however, has not been obtained and discussion of this part of the subject must remain speculative.

The conditions for ice formation are therefore made more favorable for rapid freezing by the following factors:

1. The degree of supercooling with which the droplet strikes the component surface.
2. The conduction of heat through the boundary-layer air.
3. The evaporation of water from the wetted component surface.
4. Conduction of heat away from the vulnerable region through the solid boundary.

On the other hand, conditions for ice formation are made less favorable for freezing by the following factors:

1. The conversion of kinetic energy of the gas particle to thermal energy on the airplane surface.
2. The generation of frictional heat in the boundary-layer air.
3. Conduction of heat to the vulnerable region through the solid boundary.
4. Minor factors such as the conversion of kinetic energy of the water droplet to thermal energy upon impact, and radiant thermal energy from the sun and surroundings.

It is apparent that the velocity and ambient-air temperature are important parameters in most of these factors.

In addition to the aforementioned factors, the rate at which ice forms on an airplane is determined by the following parameters:

1. Liquid-water content of the icing cloud.
2. Droplet size.
3. Shape of airplane component.
4. Size of airplane component.
5. Relative humidity of the atmosphere.

Additional considerations upon which the formation of ice on aircraft depends are:

1. Loss of liquid water from component surface due to abrasion of air stream.
2. Component-surface smoothness.

The analysis of the flow of heat from the surfaces of an airplane wing, with consideration of most of the factors just discussed, has been reported in [4, 13, 17].

Special Cases of Ice Formation. When an airplane moves from a very high altitude at which the temperature of the ambient air is very low to a lower altitude at which the temperature may be above the freezing point of water, the structure of the airplane and therefore the surfaces of the vulnerable components may be below 32F. If at the lower altitude the airplane passes through liquid-water clouds, ice may form over all the wetted areas. Although such an encounter may have annoying consequences for a short time, as soon as the structure gains heat from the atmosphere the temperature will rise and the ice will be shed. The case is of interest in that the solid body is for a short time a heat sink for the heat of fusion.

If the surface temperature of the vulnerable components on a rapidly descending aircraft falls below the saturation temperature of the boundary air, condensation will occur. For this reason fog or frost may form on an airplane windshield even in the absence of visible moisture in the atmosphere. The greater the specific heat and mass of the component, the longer the loss of vision will persist.

A serious operational problem is presented by the formation of frost which is caused by the loss of heat by radiation from the airplane surfaces to the surrounding environment. In this case both the heat of condensation and the heat of fusion are dissipated from the vulnerable area. The formation of frost occurs most commonly at night with the airplane at rest on the ground.

The addition of a nonmiscible volatile liquid, such as aviation gasoline, to a region of droplet-air interface (*e.g.*, in a carburetor) will have the effect of increasing the heat lost from the droplet environment owing to evaporation of the volatile material. The addition of the volatile liquid in this instance causes an increase in the amount of water frozen. If a volatile liquid is added, however, which is miscible and also acts as a freezing point depressant when added to water, the ice formation will be reduced or prevented under some operating conditions.

Effects of Ice on Operation of Airplane Components

The effect of the formation of ice on the function of an airplane component is determined by the magnitude of the formation and the vulnerability of the component to icing.

Wings and Control Surfaces. The formation of ice on an airfoil, whether the airfoil is used as a wing or as the fixed stabilizing component of a control surface, reduces the maximum lift coefficient and increases the drag coefficient. The shape of the ice formation, as determined by the rate of freezing of the droplet and the area of impingement on the airfoil leading edge, has an important effect in the deleterious results of the formation. Wind-tunnel research has shown that a protuberance of a given height on the surface of a wing produces the maximum loss in aerodynamic efficiency when the protuberance is located near the region of minimum local pressure. For many airfoils the minimum local pressure region may be at about the 5 per cent chord point [5]. A protuberance located in the vicinity of the stagnation pressure point has comparatively little effect on the acrodynamic characteristics. It follows that ice forming on the forward extremity of the airfoil will not impair flight as much as will formations located 5 per cent from the leading edge and on the upper surface of a wing. Icing conditions resulting from large droplets that strike over a wide area on the leading edge therefore cause a more serious effect than do small droplets. Ice formed under

droplet environmental conditions that do not result in sudden phase change but allow a part of the liquid to move slightly rearward on a wing is a more serious problem than ice that freezes where the droplet strikes near the leading edge.

The formation of ice on the wing of an airplane has been observed to increase the drag of the airplane by 35 per cent [14] during flight in natural icing conditions. The formation of ice on the control surfaces of an airplane not only increases the drag but also may adversely affect the control of the craft [15]. The nature of the problem of icing on airfoils is such that protection equipment is considered to be necessary on craft expected to operate in inclement weather.

The formation of frost on the wings causes a serious loss of lift and an increase in drag. It is considered to be extremely dangerous to undertake a take off with an airplane on which even a small frost formation exists.

Propeller Blades. The formation of ice on propeller blades, the blade being an airfoil, has an aerodynamic effect similar to that previously described for wings and control surfaces. Furthermore, during operation without icing protection, the ice formation remains on some blades but is shed by others. The rotation of blades, some with ice, others without ice, creates asymmetry in both the centrifugal and aerodynamic forces on the propeller and, therefore, will produce serious propeller vibrations. A further problem arising from ice formations on propellers is due to fragments of the ice leaving the blade during flight and striking other airplane components with damaging effect. Flight observations made during natural icing conditions [14] show that the efficiency of a propeller may be reduced by as much as 19 per cent because of the formation of ice on the blades; however, most encounters with icing do not reduce the efficiency by more than 10 per cent. Ice protection for airplane propellers is provided for aircraft that are to be operated in inclement weather.

Engine Induction System. The formation of ice in the air-induction system of aircraft engines impairs the engine operation by reducing the air flow to the engine and, in certain instances, alters the operation of the fuel-metering system. Ice forms on the inner walls of the air-inlet duct, screen, guide vanes, impact pressure tubes (part of the fuel-metering system), throttle plate, and all protuberances exposed to the air stream; in fact, this part of the icing problem presents the most serious hazard to the airplane. During cruising operations the manifold pressure may be gradually reduced because of the closing of the air inlet to the engine, or the engine may stop suddenly when changes in the fuel metering result in noncombustible charges being drawn into the engine cylinder. When the engine power is reduced for a glide to a landing, the formation of ice in the throttle area may be accelerated and prevent full power should it be needed for further cruising or climbing. If the glide path undershoots the landing area, therefore, an accident is almost unavoidable, because the power for extending the glide will not be available and a landing must be made in an unprepared area outside the airport. Ice protection is provided for the induction system in all practical aircraft. The extent to which special devices are needed to protect the induction system is largely determined by (1) the internal aerodynamic smoothness of the induction system, (2) the extent to which the entrance excludes water droplets and still admits air, (3) the manner in which fuel is delivered to the cylinder (in a mixture with air or liquid injection), and (4) the air- and fuel-metering arrangements [3, 9]. By attention to these fundamentals some recently developed designs have demonstrated increasing invulnerability to icing.

Windshield. The formation of ice on an airplane windshield or other exposed transparent area impairs the vision through the area within a few seconds after icing conditions are encountered [6, 12]. Protection for the essential areas is provided on aircraft that are to be operated in inclement weather. The location of the windshield in the forward areas of the fuselage surface and the extent to which the exterior surface of the transparent area and its frame are faired in with the fuselage contours determine the rate at which ice forms on the glass and therefore the degree of protection required.

Pitot-Static Head. The formation of ice on the exposed total-pressure pitot tube and static vents by which the airspeed indicator and altimeter are operated renders these instruments inaccurate or inoperative within a few seconds after icing conditions are encountered. Protection against such formations is provided on all aircraft in which flight in inclement weather is to be undertaken. It is possible to locate the static pressure vent in a substantially nonvulnerable position on the airplane. The pitot head must be exposed; however, the designer has some latitude in the possible shape of the component within which the quantitative need for protection may be minimized [2].

Radio Antennas. The formation of ice on radio antennas causes the exposed structures to fail because of the increased drag forces or violent vibrations induced by the flow of air over the distended shapes formed by the ice. The communication system for the airplane is guarded against such failure by submerging the antenna, by placing it in a sheltered area, or by making the structure sufficiently strong to withstand the increased loads or vibrations. The shape, location, and orientation of the antenna determine the severity of the icing problem on this type of component [8].

Vents and Air Inlets. The formation of ice on a wing leading-edge region that is forward and in the vicinity of a vent from a fuel tank will cause a loss of pressure in the ventilation system which may result in fuel-system and engine failure [16].

The formation of ice at the opening to air-duct inlets will impair the flow of air through the duct, and the severity of this aspect of the problem is aggravated by the need for a screen in the duct [1].

Protuberances and Other Vulnerable Components. The formation of ice on protuberances such as radio-antenna masts and on the forward areas of the fuselage, engine nacelles, propeller hub, and other exposed parts results in an increase in the airplane drag of as much as 25 per cent [14]. When large formations of ice which

have accumulated on forward regions of the airplane, such as engine nacelles or propeller hub, become dislodged, as will be the case when the airplane encounters air temperatures above 32F, they move with the air stream in a rearward direction. If they strike other components of the craft, serious structural damage is inflicted. Protection for areas on which the formation of ice does not seriously impair the safety of flight is not provided for in current airplane design. The need for such protection varies over a wide range depending on the aerodynamic cleanliness with which the designer is able to render the general design layout. By eliminating protuberances, by the use of flush inlet scoops and sheltered vents, and by the arrangement of the main airplane components so as to eliminate, insofar as possible, leading-edge areas having small radii of curvature, the deleterious results of icing may be significantly minimized.

Topics Needing Further Study

The hazards and nuisances presented by the airplane-icing problem have in part been met by the development of ice-protection devices and by refinements in airplane design which make the components less vulnerable. Of the various methods of removing or preventing ice—mechanical, chemical, and thermal—the thermal method has given the greatest promise of providing a solution that can be adapted to practical transport or military airplanes without undue penalty in weight, added complication, or airplane cost. It should not be construed, however, that the other methods are unworkable or impracticable. The suggestions for further research that follow pertain to the extension of knowledge on the thermal ice-prevention system, to several atmospheric studies, and to an instrument-development program which is related to meteorological observations.

Heat-Transfer Relationships. The validity and accuracy of the relationships at present employed for determining heat transfer, mass transfer, recovery temperature, and droplet interception should be examined for the full range of values of velocity, shape, and size of aircraft now under design or likely to be designed in the near future. The recovery temperature at various flow conditions on a wetted airfoil at velocities near and above the speed of sound may be of interest in future aircraft; data on this factor should be established. Of interest in this problem is the comparative significance of viscous dissipation, conductive effects, and evaporation in the regions of laminar, turbulent, and separated flow for the high-speed range. An understanding of the nature of thermal ice prevention in greater detail than now exists is needed in order that effective protection may be provided without prohibitive penalities in added equipment to operating aircraft.

Variations in the Icing Problem with Altitude. An attractive speculation is that by flying very high the icing problem may be avoided. It is therefore of interest to learn what icing conditions are to be encountered at altitudes such as in the tops of cumuliform clouds, and with what degree of severity and frequency icing is apt to be encountered. Almost all the data thus far collected have been obtained at comparatively low altitudes. In consideration of the probability that extensive future operations will be conducted at high altitudes, the altitude range of the data should be extended. Data should be collected at altitudes above 25,000 ft on the dimensions and frequency of icing conditions and on the liquid water content, droplet size, and temperature range occurring with icing at these altitudes.

Critical Geographical Areas. Reports from military and commercial transport operations supply evidence that some geographical areas are attended by unique icing conditions of severe intensity; therefore, factual data on and a theoretical basis for the variations in the icing problem with geographical locations should be established.

Physics of Icing-Cloud Stability. Inasmuch as water normally freezes at 32F, it is of prime interest to understand why the droplets in an icing cloud have not undergone the phase change to the solid state and become snow or hail. An effort should be made to identify and analyze the full range of conditions which must be met in order that water droplets may change to ice. When this topic is more thoroughly explored and understood, the very interesting problems of modifying the weather artificially can be evaluated more adequately.

Forecasts of Icing. If airmen could receive accurate and complete forecasts of icing conditions, many operations could be routed around the hazardous locality, thus eliminating the need for special flight equipment and greatly extending the usefulness of aircraft in which ice protection is not installed. Improvements in the forecasting of icing are needed as are the means for effectively communicating the forecast to flight-operation centers.

Instruments for Collecting Atmospheric Data on Icing. The understanding of the phenomenon of icing, the design of efficient anti-icing equipment, and the development of reliable techniques of forecasting the occurrence of icing require that accurate and comprehensive data on the factors that produce icing be collected. Although instruments have been developed whereby data of considerable value have been obtained [11], still further instrumental research is needed.

In the study of the physics of the icing cloud, instruments that will reveal the existence and size of nuclei, water droplets, and snow crystals are needed. Data on such particulate substances need to be collected in order to reveal their variations with space and time.

In the study of the operational aspects of the icing problem, instruments are needed with which data on the occurrence of icing over wide ranges of geographical location and altitude can be collected.

REFERENCES

1. Army Air Forces, *Report on Icing Tests for Screens in Air Inlet Ducts.* Translation F-TS 2625RE.
2. Christenson, C. M., "Aircraft Icing Revelations Lighten

Trying Tasks of Designer and Pilot." (Condensation of paper presented at Meeting of the Society of Automotive Engineers, So. Calif. Sect., Nov. 1, 1945.) *J. Soc. automot. Engrs., N. Y.*, 54:103–104 (Oct. 1946).

3. Coles, W. D., Rollin, V. G., and Mulholland, D. R., "Icing-Protection Requirements for Reciprocating-Engine Induction Systems." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1993 (1949).
4. Hardy, J. K., *An Analysis of the Dissipation of Heat in Conditions of Icing from a Section of the Wing of the C-46 Airplane.* National Advisory Committee for Aeronautics, A.R.R. 4I11a, Washington, D. C., 1944.
5. Jacobs, E. N., *Airfoil Section Characteristics as Affected by Protuberances.* National Advisory Committee for Aeronautics Rep. No. 446, Washington, D. C., 1932.
6. Jones, A. R., Holdaway, G. H., and Steinmetz, C. P., "A Method for Calculating the Heat Required for Windshield Thermal Ice Prevention Based on Extensive Flight Tests in Natural Icing Conditions." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1434 (1947).
7. Kantrowitz, A., "Aerodynamic Heating and the Deflection of Drops by an Obstacle in the Airstream in Relation to Aircraft Icing." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 779 (1940).
8. Kepple, W. L., "Determination of Aircraft Antenna Loads Produced by Natural Icing Conditions." *Res. Mem. nat. adv. Comm. Aero., Wash.*, E7H26a (1948).
9. Kimball, L. B., *Icing Tests of Aircraft-Engine Induction Systems.* National Advisory Committee for Aeronautics, A.R.R., Washington, D. C., 1943.
10. Langmuir, I., and Blodgett, Katharine B., *A Mathematical Investigation of Water Droplet Trajectories.* General Electric Res. Lab. Rep., Contract W-33-038-ac-9151, Schenectady, N. Y., July 1945.
11. Lewis, W., "A Flight Investigation of the Meteorological Conditions Conducive to the Formation of Ice on Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1393 (1947).
12. McBrien, R. L., "Icing Problems in Operation of Transport Aircraft." *J. Soc. automot. Engrs., N. Y.*, 49:397–408 (1941).
13. Neel, C. B., and others, "The Calculation of the Heat Required for Wing Thermal Ice Prevention in Specified Icing Conditions." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1472 (1947).
14. Preston, G. M., and Blackman, C. C., "Effects of Ice Formations on Airplane Performance in Level Cruising Flight." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1598 (1948).
15. Rodert, L. A., Clousing, L. A., and McAvoy, W. H., *Recent Flight Research on Ice Prevention.* National Advisory Committee for Aeronautics, A.R.R., Washington, D. C., 1942.
16. Ruggeri, R. S., von Glahn, U., and Rollin, V. G., "Investigation of Aerodynamic and Icing Characteristics of Research Fuel-Vent Configurations." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1789 (1949).
17. Tribus, M., and Tessman, J. R., *Report on the Development and Application of Heated Wings.* Army Air Force Tech. Rep. No. 4972, Add. I, 1946. (Available from Office of Technical Services, U. S. Dept. of Commerce, as PB No. 18122.)

METEOROLOGICAL ASPECTS OF AIRCRAFT ICING

By WILLIAM LEWIS

U. S. Weather Bureau, Washington, D. C.

Introduction

The formation of ice on airplanes has long been recognized as a serious problem affecting the regularity and safety of aircraft operations. With the improvement of instrument-flying techniques and navigational aids and the resulting increase in regularity of operations in inclement weather, the problems associated with icing conditions have become of prime importance to airline operators, airplane manufacturers, and meteorologists.

Until recently, very little quantitative data was available on the actual physical characteristics of icing conditions. Most of the available information on the meteorology of icing was based either on studies of pilots' reports [3, 11, 16, 17] or on theoretical inferences regarding cloud composition. In particular, the ideas of Bergeron on the composition of precipitating stratiform clouds in a warm-front situation have been used as a basis for a model of icing zones in a warm-front cloud system which has been presented in several textbooks on aeronautical meteorology (*e.g.*, [5]). A somewhat different model, later proposed by Findeisen [6], contained a more accurate representation of the typical composition of precipitating stratiform clouds. The importance of orographic effects in the geographical distribution of icing conditions was emphasized by Riley [16].

With the development of thermal methods of ice protection, the need for quantitative information on liquid-water concentration, cloud-drop diameter, and temperature in icing conditions was recognized and projects were undertaken to obtain the necessary data. The first of these was the Mount Washington Project, conducted by the Mount Washington Observatory, Harvard University, the U. S. Air Force, and the U. S. Weather Bureau. This project has resulted in the collection of a large amount of data on the composition of clouds at the summit of Mount Washington, N. H., at an elevation of approximately 6300 ft [7].

Flight measurements of liquid-water concentration and drop diameter in clouds have been made by the National Advisory Committee for Aeronautics [13–15], the Air Materiel Command of the U. S. Air Force [4], and other groups. As a result of these researches the probable range and relative frequency of occurrence of various values of the significant variables have been tentatively established and values have been chosen to be used as a basis for the design of thermal ice-prevention equipment [9]. This work has also led to a more complete understanding of the composition of clouds in general and the type and severity of icing conditions to be expected in various weather situations.

The Physical Characteristics of Icing Conditions

For most practical purposes, the physical characteristics of icing conditions may be regarded as determined by the values of three basic parameters: liquid-water concentration w, drop diameter d, and temperature T. The necessary and sufficient condition for ice formation upon a small object moving through the air with a velocity V is that w be greater than zero and T be below freezing by an amount equal to or greater than ΔT, where ΔT is the wet-air kinetic temperature rise at the velocity V. Although the value of ΔT is influenced to a slight extent by several factors, including the shape and thermal conductivity of the moving object and the value of w, it is given with sufficient accuracy for most practical purposes by the following expression:

$$\Delta T = 0.8 \left(\frac{V}{100}\right)^2 \frac{\gamma_m}{\gamma_d}, \tag{1}$$

where γ_m and γ_d are the saturated- and dry-adiabatic lapse rates, V is in miles per hour, and ΔT is in centigrade degrees. The rate at which water is intercepted by an object, expressed in grams per second per square centimeter of projected area, is given by

$$I = EwV \times 10^{-6}, \tag{2}$$

where w is in grams per cubic meter and V is in centimeters per second. This equation defines the quantity E, called the *collection efficiency*, the value of which depends upon V, d, and the size and shape of the object. In general, the collection efficiency is greatest for small objects, high air speeds, and large drops.

The variation of the kinetic temperature rise with air speed, and the dependence of collection efficiency upon drop size, air speed, and size and shape of the object, furnish an explanation of the importance of temperature and drop size in determining the severity of icing for particular components of an airplane. In the case of wings, the sizes and curvatures of the wing sections commonly used on modern airplanes, and the speeds at which these airplanes operate, are such that the values of collection efficiency vary from zero or near zero for the smallest cloud drops to 70 to 80 per cent for large cloud drops and virtually 100 per cent for freezing drizzle or freezing rain. The drop size is therefore nearly as important as the liquid-water concentration in determining the severity of wing icing.

In the case of propellers, the small cross sections and high blade speeds result in very high values of collection efficiency for almost the entire range of cloud drop sizes; hence, drop size is relatively unimportant for propeller icing. Because of the large values of the

kinetic temperature rise associated with the high speeds of propeller blades, the free-air temperature is an important factor in determining the extent of ice formation. For propellers of the size and speed ordinarily used on transport airplanes, ice formations which occur at free-air temperatures of −5C or higher do not extend far from the root of the blade and have only a very slight effect on propeller efficiency. With decreasing air temperature, the extent of ice formation gradually increases, reaching almost to the blade tips at −15C. At temperatures below −15C, serious propeller icing may occur even with low values of liquid-water concentration and small drops.

In the case of windshields, the configurations most commonly used, such as the familiar V-type, are ordinarily affected by icing conditions with average-sized drops; however, in the special case of a windshield which is flush with the fuselage contours, the collection efficiency may be so low that icing occurs only in the presence of exceptionally large drops.

The shape of the ice formations is influenced greatly by the rate of freezing. Under conditions of low temperature, low water concentration and small drops, the water freezes as it strikes without forming a liquid film, thus producing rather narrow, smooth, pointed ice accretions. On the other hand, temperatures near freezing, and larger values of water concentration and drop size, result in the formation of a film of liquid water which spreads as it freezes, giving rise to irregular ice formations with flat or concave surfaces facing the air stream. These two types of ice are generally called *rime* and *glaze*, respectively.

The Definition of Icing Intensity. The preceding discussion shows that the intensity of icing is dependent on liquid-water concentration, cloud-drop diameter, and temperature. Since the relative importance of these factors is different for different airplane components, no single function of these quantities can be used to define a scale of icing intensity that will apply equally to all components of all aircraft. The only scale now in use which defines light, moderate, and heavy icing in quantitative terms is that used by the U. S. Weather Bureau in reporting the intensity of icing observed at Mount Washington. This scale is based on the rate at which ice would form on a cylinder three inches in diameter moving at a speed of 200 mph, which represents a reasonable compromise among the various aircraft components. Because of the fact that this scale was designed to describe conditions on mountain summits, which are usually more severe than those encountered in the free air, conditions defined by it as "moderate" or "heavy" are substantially more severe than conditions described in the same terms by pilots. In fact, experience with pilots' reports suggests that the conditions defined as "trace," "light," and "moderate" according to the Weather Bureau scale would ordinarily be reported by pilots as "light," "moderate," and "heavy," respectively. For this reason some misunderstanding has resulted from the application of this scale to icing conditions encountered in flight. Accordingly, it is suggested that the scale be modified for use as a basis for quantitative definitions of icing conditions encountered in flight by assigning different descriptive terms to the four degrees of icing intensity defined by the scale. Table I shows the scale now used by the Weather Bureau for reporting icing conditions at mountain stations and the scale suggested for flight use. The definitions are given in terms of the rate of accretion on a cylinder three inches in diameter; corresponding values of liquid-water concentration for four values of drop diameter are also included.

Measurement of Liquid-Water Concentration and Cloud-Drop Diameter. Although several methods have been employed for measuring liquid-water concentration and drop diameter in clouds [10], nearly all of the reliable observations now available have been made by the rotating-multicylinder method devised by Arenberg [1]. As used in flight, this method employs a series of rotating cylindrical collectors (usually four) of different diameters which are mounted coaxially and exposed with the axis at right angles to the air stream for a measured time interval. The weights of ice collected on the cylinders are used to determine the liquid-water concentration and the average drop diameter. The calculations are based on values of the collection efficiency of cylinders calculated by Langmuir and Blodgett [12]. This method yields reliable and fairly accurate values of liquid-water concentration. The determination of drop size is reliable and reasonably accurate for values of drop diameter below 20 μ, but becomes increasingly inaccurate and unreliable for larger drops. Values of drop diameters around 30 μ are subject to considerable error and values over 40 μ are highly unreliable.

The value of drop diameter obtained in this way is called the *mean-effective diameter* and is believed to represent approximately the volume median size; half of the water is contained in larger drops, half in smaller

Table I. Icing-Intensity Scales

Rate of ice accretion on cyl. 3″ diam. at 200 mph (g cm^{-2} hr^{-1})	Liquid-water concentration (g m^{-3}) for certain values of drop diameter				Weather Bureau scale of icing intensity for mountain stations	Suggested scale of icing intensity for flight reports
	10 μ	15 μ	20 μ	30 μ		
0.00– 1.00	0.00–0.22	0.00–0.10	0.00–0.07	0.00–0.05	trace of ice	light icing
1.01– 6.00	0.23–1.32	0.11–0.60	0.08–0.42	0.06–0.30	light icing	moderate icing
6.01–12.00	1.33–2.64	0.61–1.20	0.43–0.84	0.31–0.60	moderate icing	heavy icing
>12.00	>2.64	>1.20	>0.84	>0.60	heavy icing	very severe icing

ones. By assuming the form of the drop-size distribution, it is theoretically possible to determine approximately the degree of homogeneity of drop sizes from the rotating-cylinder data. In actual practice, however, it has been found that determinations of drop-size distribution made in flight by this method are not reliable [14], although satisfactory results have been reported from Mount Washington [7]. Comparisons of values of maximum drop-diameter, obtained from the width of the area of drop impingement on a stationary cylinder, with values of the mean-effective diameter simultaneously determined by the rotating-cylinder method indicate that most clouds do not contain significant amounts of liquid water in the form of drops appreciably larger than the mean-effective diameter. For practical purposes, therefore, in the study of icing conditions, most clouds may be regarded as composed of uniform drops.[1]

The Concentration of Liquid Water in Clouds

In general, the formation of clouds is brought about by cooling due to the adiabatic ascent of air. For a strictly adiabatic process, without mixing or precipitation, the concentration of liquid water at an altitude z is given by the following equation:

$$w_z = \rho_z (x_c - x_z), \tag{3}$$

where w_z is the liquid-water concentration (g m^{-3}) at altitude z, ρ_z is the density of dry air (kg m^{-3}) at altitude z, x_c is the saturation mixing ratio (g kg^{-1}) at the condensation level, and x_z is the saturation mixing ratio (g kg^{-1}) at altitude z. Calculations from this equation indicate that the most important factors determining the liquid-water concentration in adiabatically lifted air are the temperature at the condensation level and the vertical distance above the condensation level. The actual altitude of the condensation level is of minor importance. Values of w, calculated by equation (3), are presented in Table II. These values are

Table II. Liquid-Water Concentration in Adiabatically Lifted Air*

Vertical distance above condensation level	Temperature at condensation level					
	−20F	0F	10F	20F	30F	40F
ft	*g m⁻³*	*g m⁻³*	*g m⁻³*	*g m⁻³*	*g m⁻³*	*g m⁻³*
1000	0.09	0.18	0.26	0.35	0.44	0.53
2000	0.17	0.34	0.48	0.65	0.83	1.01
3000	0.22	0.47	0.67	0.92	1.18	1.44
4000	0.27	0.57	0.82	1.13	1.47	1.82
6000	0.32	0.71	1.04	1.46	1.93	2.47

* Condensation pressure = 850 mb.

based on a condensation level at a pressure altitude of 4780 ft. In actual clouds the processes of precipitation as well as the entrainment and mixing of drier air both act to reduce the liquid-water concentration below the values based on adiabatic lifting; hence these values should be regarded as maxima which may be approached under certain conditions but are very unlikely to be exceeded.

In a study of the physical factors which influence the amount of liquid water in clouds, it is convenient to consider three classes of clouds, based on the three principal processes by which air may be lifted, namely, convection, turbulence, and horizontal convergence of the velocity field.

Clouds Formed Primarily by Convection (Cumulus and Cumulonimbus). As long as no precipitation occurs, the concentration of liquid water in cumulus clouds is determined mainly by the temperature of the cloud base, the vertical distance above the base, the amount of entrainment and mixing, and the humidity of the environmental air. Recent studies [2] indicate that the lapse rate in cumulus clouds is usually very close to that of the environment, and that the amount of entrainment increases as the lapse rate exceeds the moist-adiabatic. The conditions for maximum liquid-water concentration are, therefore, a lapse rate only slightly in excess of moist-adiabatic and high relative humidity in the environmental air. Under these conditions the liquid-water concentration sometimes closely approaches the value based on adiabatic lifting.

In general, the effect of entrainment is to reduce the liquid-water concentration below the adiabatic value, and this effect increases from the base to the top of the cloud. The result is that the greatest values of liquid-water concentration occur at a distance above the cloud base of from one-half to three-fourths the height of the cloud, instead of at the top. The horizontal distribution of liquid-water concentration across cumulus clouds has not been accurately determined because instantaneous measurements have not been made. However, it would appear reasonable to suppose that the central portions of an updraft would be least affected by entrainment and would therefore have higher liquid-water content than the outer portions.

The effect of the formation of ice crystals in cumulus clouds, as in all other types, is a rapid reduction in the liquid-water concentration. The transformation from liquid drops to snow usually proceeds quite rapidly once it begins, frequently reducing the liquid-water concentration from over 1 g m^{-3} to about 0.1 g m^{-3} in twenty minutes or less. The formation of ice crystals is one of the most important factors limiting the icing hazards in cumulus clouds. Careful observation of cumulus cloud development indicates that individual cumulus congestus towers seldom remain in that stage longer than ten to fifteen minutes; they generally either subside and evaporate or change to ice crystals within that time.

Clouds Formed Primarily by Turbulence (Stratus, Stratocumulus, and Altocumulus). Clouds of this type are formed in the upper portions of certain layers of air which are mixed by turbulence. This may be the surface turbulence layer, as is usually the case with stratus or low stratocumulus, or a higher layer which becomes unstable as a result of lifting, differential advection, or some other process. In the ideal case, in which the

1. This conclusion is based on data taken in flight. Observations on Mount Washington indicate a considerable frequency of nonuniform drop-size distributions [7, 8].

entire layer is thoroughly mixed adiabatically, the liquid-water concentration would be given by equation (3). In actual clouds, however, the observed values of liquid-water concentration are less than the calculated values, usually averaging from one-fourth to one-half and rarely exceeding three-fourths of the calculated value. In most cases the liquid-water concentration increases approximately linearly from near zero at the cloud base to a maximum just below the top. The difference between the observed and theoretical values of liquid-water content may be attributed to incomplete mixing within the layer and to the downward mixing of some dry air from above the cloud.

Ice crystals may form within a turbulent layer cloud or may fall into it from an overlying altostratus cloud. In the former case, the precipitation is usually light and the concentration of liquid water decreases quite slowly near the top and more rapidly in the lower parts of the layer. This process leads to a gradual dissipation of the cloud, proceeding from the bottom upward. In the latter case, the intensity of precipitation is frequently great enough to cause the rapid dissipation of the entire cloud layer.

Clouds Formed Primarily by Horizontal Convergence (Altostratus, Altocumulus, and Altocumulus-Altostratus). Horizontal convergence in the wind-velocity field at low altitudes is accompanied by relatively slow lifting of large masses of air. This is the type of lifting which ordinarily takes place in the forward and central portions of moving cyclones and leads to the formation of very deep and extensive cloud systems predominantly of the altostratus type. Observations have shown that supercooled liquid-water is ordinarily almost completely absent from the main altostratus portions of such cloud systems, since these are composed almost entirely of ice crystals. Around the edges of these cloud systems, however, a complex pattern of altocumulus layers at various levels frequently occurs, composed predominantly of liquid water. When these layers form in stable air, they are lenticular and usually rather low in water content. When certain layers become unstable as a result of the lifting process, altocumulus of the turbulence type, resembling high stratocumulus, may be formed. The concentration of liquid water in these clouds depends on temperature and cloud thickness, as noted in the preceding paragraph.

Freezing Rain and Freezing Drizzle. The meteorological conditions ordinarily required for the formation of freezing rain are an inversion, usually frontal, with temperatures above freezing in the warm air above the inversion and temperatures below freezing in the cold air below the inversion. The temperature range of the occurrence of freezing rain is rather small, since the raindrops freeze and become sleet (U. S. Weather Bureau definition) at temperatures only a few degrees below freezing. Unfortunately, measurements of the concentration of liquid water in freezing rain are not available. Estimates can be made, however, since it is known that freezing rain generally consists of relatively light, uniform precipitation over a fairly large area. If a precipitation rate of 0.1 in. hr^{-1} and a drop diameter of one millimeter are assumed, the liquid-water concentration would be about 0.15 g m^{-3}, which is less than is usually found in clouds.

Because freezing drizzle is only rarely encountered and the methods used to measure drop size in flight are not reliable for drops of drizzle size (thus making the correct identification of drizzle difficult), very little information is available concerning its liquid-water content. Unlike freezing rain, freezing drizzle is not limited in its occurrence to frontal conditions but may form in a variety of meteorological situations and over a wide range of temperature and altitude. The fact that large values of cloud drop size tend to occur only with small values of liquid-water concentration would suggest that low values of liquid-water content would be expected in freezing drizzle.

Statistical Data on Liquid-Water Concentration. A substantial amount of data on liquid-water concentration and drop size in clouds in the United States is available as a result of flight measurements by various organizations. Most of these observations were made by the National Advisory Committee for Aeronautics [13–15]. A considerable number were also made by the Air Materiel Command Aeronautical Ice Research Laboratory [4], and a limited number by United Air Lines, American Airlines, and the Douglas Aircraft Company. All these measurements were made by the rotating-cylinder method and represent average values over time intervals varying from about 30 sec to 10 min. The average exposure interval was about 1 min in cumuliform clouds and 3–5 min in stratiform clouds. Data from all the above-mentioned sources have been included in Tables III and IV.

Table III. Observed Frequency of Various Values of Liquid-Water Concentration in Clouds

Liquid-water concentration	*St, Sc* 327 obs.	*Ac, Ac-As* 264 obs.	*Cu, Cb* 342 obs.
g m^{-3}	%	%	%
0.00–0.09	12	50	31
0.10–0.19	32	32	
0.20–0.29	22	13	26
0.30–0.39	16	4	
0.40–0.49	12	1	22
0.50–0.59	5	0	
0.60–0.69	0.3	0	10
0.70–0.79	0.6	0	
0.80–0.89	0.3	0	7
0.90–0.99	0	0	
1.00–1.19	0	0	2
1.20–1.39	0	0	1.0
1.40–1.59	0	0	0
1.60–1.79	0	0	0.7
	g m^{-3}	g m^{-3}	g m^{-3}
Lower quartile	0.13	0.05	0.15
Median	0.22	0.10	0.34
Upper quartile	0.35	0.17	0.55
Maximum	0.80	0.41	1.71

Table III gives frequency distributions of values of liquid-water concentration observed in three principal cloud types: cumulus and cumulonimbus, stratus

and stratocumulus, and altocumulus and altostratus-altocumulus. This last group includes all middle cloud types containing liquid water since typical altostratus clouds are composed of ice crystals.

It can be seen from Table III that the average and extreme values of liquid-water concentration in cumulus clouds are much greater than in other cloud types. This is due to the much greater vertical extent of cumulus clouds. A compensating factor is their limited horizontal extent. It is noted that the concentration of liquid water is considerably lower in altocumulus and altocumulus-altostratus than in stratus and stratocumulus clouds. The difference is due to the lower temperature and smaller vertical thickness of altocumulus and also to the frequent occurrence of ice crystals in altocumulus when associated with altostratus.

The Size of Cloud Drops

Important inferences concerning the sizes of cloud drops to be expected from various conditions of cloud formation can be drawn from a theoretical analysis of the mechanism of cloud-droplet formation and growth. A detailed and quantitative treatment of the physics of cloud-drop formation has recently been given by Howell [8]. According to this study, for the case of uniformly lifted air, "the numerical concentration of drops is determined primarily by the rate of cooling during the initial stage of condensation. It depends, as a rule, only slightly on the concentration of condensation nuclei. With continued uniform cooling, the drop concentration diminishes slightly toward a fixed constant value." It is also shown that uniform lifting leads to a rather uniform drop-size distribution. The effect of evaporation, owing either to warming in descending portions of a cloud or to entrainment of dry air, is a decrease in the concentration of drops, since the smaller drops evaporate most rapidly. On the basis of these considerations, the highest concentration of drops would be expected in rapidly formed cumulus clouds and the lowest concentrations in altocumulus clouds formed slowly by convergence in stable air.

Actual measurements of cloud-drop diameter made in flight indicate wide variations for clouds of all types. Frequency distributions of observed values of mean-effective drop diameter are given in Table IV for three principal cloud types and two geographical areas. The classification on the basis of location was made because it was observed that the largest values of drop size were found mostly along the Pacific Coast. The data in Table IV are from the sources listed previously for Table III. Representative values of drop concentration were calculated from the median values of liquid-water concentration and drop diameter and are also included in Table IV.

It can be seen from an examination of Table IV that there is a slight variation of drop size with cloud type and a greater variation with geographic location. The occurrence of larger drops in the Pacific Coast region is believed to be due to the prevalence of maritime air masses with little pollution from combustion sources, and hence a very low concentration of effective nuclei. The small average drop concentration in altocumulus clouds was to be expected from condensation theory. Larger values for stratocumulus and cumulus clouds are believed to be due to greater turbulence and hence more rapid lifting at the time of condensation. The fact that cumulus clouds have smaller drop concentration than stratocumulus may be due to the greater effectiveness of such modifying factors as entrainment and precipitation, which promote evaporation and thus reduce the drop concentration.

Table IV. Observed Frequency of Various Values of Mean-Effective Drop Diameter in Clouds

Mean-effective drop diameter	Pacific Coast region			Other areas in the United States		
	Ac, Ac-As 112 obs.	*St, Sc* 60 obs.	*Cu, Cb* 220 obs.	*Ac, Ac-As* 128 obs.	*St, Sc* 267 obs.	*Cu, Cb* 110 obs.
μ	%	%	%	%	%	%
0– 9	8	5	5	20	32	6
10–14	22	36	19	32	43	31
15–19	28	25	25	30	16	35
20–24	22	17	28	12	6	20
25–29	7	7	15	5	2	5
>29	13	10	8	1	1	3
	μ	μ	μ	μ	μ	μ
Lower quartile	13.5	12.5	14.5	10	9	13
Median	18	16	19.5	14	11	16
Upper quartile	23	22	24	18	14.5	20
	cm^{-3}	cm^{-3}	cm^{-3}	cm^{-3}	cm^{-3}	cm^{-3}
Representative drop concentration	35	100	90	75	320	160

The Problem of Forecasting Icing Conditions

For the purposes of this discussion, it will be assumed that satisfactory forecasts of the type, location, thickness, altitude, and temperature of cloud masses, and the occurrence and type of precipitation can be made. It is fully realized that this assumption is frequently not valid, but it is made here in order to separate the problems peculiar to forecasting icing intensity from the general problem of weather forecasting. In other words, the problem to be treated here is that of estimating the icing intensity to be expected within a cloud of known type, dimensions, and temperature. If this can be done satisfactorily, a forecast of icing can be derived from any good forecast of weather and cloud conditions. However, regardless of whether satisfactory forecasts of cloud conditions can be made, reliable estimates of the presence of icing in the currently observed and reported cloud conditions may be of considerable value.

The importance of a consideration of cloud type in estimating icing conditions cannot be overemphasized, since entirely different icing conditions prevail in different types of clouds. A brief discussion of the icing conditions likely to be found in clouds of various types is given in the following paragraphs.

Cumulus and Cumulonimbus Clouds. Icing conditions in cumulus clouds are sharply limited in horizontal extent, highly variable, and occasionally very severe.

The most severe icing conditions occurring anywhere are found in the upper half of tall cumulus congestus clouds just before they reach the cumulonimbus stage. Ordinarily, such clouds comprise only a small portion of the total air mass and therefore they can usually be circumnavigated. Also, the limited horizontal extent of cumulus towers limits the amount of ice that can be collected during a single passage through a cloud. Because of the removal of liquid water by precipitation, icing conditions in cumulonimbus clouds are usually much less severe than in cumulus congestus clouds; however, conditions in cumulonimbus are extremely variable, and sometimes very heavy icing may be encountered for short distances.

Stratus and Stratocumulus Clouds. Stratus and stratocumulus clouds ordinarily contain moderate to low concentrations of liquid water, although fairly high values sometimes occur near the tops of unusually thick layers. These clouds are rarely characterized by large values of drop diameter. Although the icing conditions in stratus and stratocumulus clouds are usually light to moderate in intensity, they present one of the chief icing hazards to aircraft operation, since the cloud layers frequently cover large areas and it is not possible to ascend or descend without flying through them. Also, traffic control procedures may occasionally require prolonged flight in a stratocumulus layer. The intensity of icing usually increases from near zero at the base of the cloud layer to a maximum just below the top, the maximum intensity being proportional to the thickness of the layer. The factor which limits the intensity of icing is the limited thickness of the cloud layers. Individual layers of stratus or stratocumulus clouds seldom exceed 3000 ft in vertical extent. For layers of a given thickness, the liquid-water content is greater in warmer clouds, hence the greatest rates of ice accretion will be found at temperatures only a few degrees below freezing. This relation does not apply to individual clouds, however, since in any particular cloud layer the icing is most severe in the upper, colder portions.

When light or very light snow falls from a stratus or stratocumulus layer it is an indication of decreasing liquid-water concentration and icing intensity; when moderate or heavy snow occurs, the liquid water disappears very quickly and little or no icing is to be expected.

Altocumulus Clouds. Icing conditions in altocumulus clouds are generally similar to, but milder than, those in stratocumulus clouds, as the temperatures are usually lower and the average cloud thickness is less. Large values of drop diameter occur more frequently in altocumulus clouds with the result that icing conditions are occasionally considerably more severe than would be expected from the thickness of the layer. Ordinarily, however, only light and occasionally moderate conditions are encountered in altocumulus clouds and these are usually not difficult to avoid by a change of altitude.

Altostratus Clouds and Altocumulus Associated with Altostratus. Altostratus clouds usually consist of thick and extensive cloud masses which are composed almost entirely of ice crystals and therefore do not cause icing. This fact is of the greatest importance in forecasting icing conditions. The failure of forecasters to realize that altostratus clouds and the large precipitation areas usually composed predominantly of altostratus contain little or no supercooled liquid water is probably responsible for more errors in forecasting icing conditions than any other factor.

Around the edges of altostratus cloud systems, especially on the side where air is flowing upward into the system, there is frequently found a rather complex grouping of altocumulus layers. These usually merge with the altostratus, losing their liquid water as they become impregnated with ice crystals. Frequently, rather large cloud areas of mixed composition are formed in this way, for though a mixture of ice crystals and cloud drops is an unstable condition and therefore transient, a continuous supply of freshly formed altocumulus clouds is often available. Usually only light icing occurs in mixed clouds of this type.

Flight Planning and Dispatching in Relation to Icing Conditions. The consideration that should be given to icing in flight planning is largely a function of the ice-protection equipment on the airplane. Aircraft without deicing equipment of any kind should avoid all subfreezing clouds if possible. Aircraft equipped with rubber-boot deicers should avoid prolonged flight in continuous stratus or stratocumulus layers by a suitable change in altitude and should circumnavigate heavy cumulus clouds whenever practicable. Aircraft having thermal ice-prevention systems can safely fly in a great majority of icing conditions, but if exceptionally severe or prolonged icing conditions are encountered, some changes in altitude or flight path may be advisable. It should not be necessary to postpone or cancel flights because of reported or expected icing conditions when adequate thermal ice protection is available.

Because the occurrence of icing conditions is usually quite variable both in space and time, the practice of prohibiting flights in certain areas as a result of reported icing conditions is of doubtful validity. Such restrictions should not be applied to aircraft with adequate thermal ice-prevention systems.

Problems Requiring Further Research

One of the outstanding problems in the study of icing conditions is that of obtaining satisfactory measurements of the meteorological quantities involved. Much effort has been applied to this problem and noteworthy results have been achieved, but much still remains to be done. For example, there is a need for a more convenient and dependable means of determining the distribution of drop sizes in clouds, and also for a more accurate means of measuring mean-effective and maximum drop diameters when these exceed 30 to 40 μ.

There is also a need for simple and dependable automatic instruments for recording liquid-water concentration and average drop size, instruments suitable for use during routine flights to obtain statistical data.

One aspect of the icing problem in which our present

knowledge appears to be inadequate concerns the conditions under which ice crystals form in clouds. Meteorologists have been keenly interested in this problem since the advent of the Bergeron-Findeisen theories of precipitation. Thus far, however, no adequate explanation is available for the observed facts that certain types of clouds usually contain ice crystals while other types do not, nor is there a satisfactory explanation of the mechanism of the formation of ice crystals in liquid clouds.

The discovery by Schaefer [18] that liquid clouds could be converted to ice crystals by seeding with dry ice has resulted in investigations of the effects of cloud seeding by various agencies. At present there is a wide difference of opinion among investigators in this field, but there is good reason to expect that these experiments will eventually make an important contribution to a fuller understanding of the formation of ice crystals in clouds.

The modifications of clouds that have been accomplished by seeding suggest the possibility that this technique may provide a practicable means of reducing the icing hazard in limited areas, for example in the vicinity of busy airports. Further experiments will be required to establish the practical possibilities and limitations of seeding as a means of locally controlling icing conditions.

The data now available on liquid-water concentration and drop size in clouds, while sufficient to define in general terms the ranges of values usually encountered, are still inadequate to determine the frequency and severity of icing conditions to be expected in various areas under the influence of various synoptic situations. It is expected that if and when automatic recording instruments come into general use, a large body of statistical data will be accumulated which can be analyzed from a synoptic-climatological standpoint to provide information which may be of considerable value to airplane operators and practicing meteorologists.

REFERENCES

1. Arenberg, D. L., and Harney, P. J., "The Mount Washington Icing Research Program." *Bull. Amer. meteor. Soc.*, 22:61–63 (1941).
2. Austin, J. M., "A Note on Cumulus Growth in a Nonsaturated Environment." *J. Meteor.*, 5:103–107 (1948).
3. Benum, F. W., and Cameron, H., "A Study of Ice Accretion on Aircraft over the Canadian Rockies." *Bull. Amer. meteor. Soc.*, 25:28–33 (1944).
4. Brock, G. W., "Liquid-Water Content and Droplet Size in Clouds of the Atmosphere." *Trans. Amer. Soc. mech. Engrs.*, 69:769–770 (1947).
5. Byers, H. R., *General Meteorology*. New York, McGraw, 1944. (See p. 557)
6. Findeisen, W., "Meteorological-Physical Limitations of Icing in the Atmosphere." (Translation) *Tech. Mem. nat. adv. Comm. Aero., Wash.*, No. 885 (1939).
7. Harvard University, *Harvard-Mount Washington Icing Research Report 1946–1947*. A. F. Tech. Rep. No. 5676, U. S. A. F. Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio, 1948.
8. Howell, W. E., "The Growth of Cloud Drops in Uniformly Cooled Air." *J. Meteor.*, 6:134–149 (1949).
9. Jones, A. R., and Lewis, W., "Recommended Values of Meteorological Factors to be Considered in the Design of Aircraft Ice-Prevention Equipment." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1855 (1949).
10. —— "A Review of Instruments Developed for the Measurement of the Meteorological Factors Conducive to Aircraft Icing." *Res. Mem. nat. adv. Comm. Aero., Wash.*, No. A9C09 (1949).
11. Lacey, J. K., "A Study of Meteorological and Physical Factors Affecting the Formation of Ice on Airplanes." *Bull. Amer. meteor. Soc.*, 21:357–367 (1940).
12. Langmuir, I., and Blodgett, Katharine B., *A Mathematical Investigation of Water Droplet Trajectories*. Schenectady, General Electric Co., 1945.
13. Lewis, W., "A Flight Investigation of the Meteorological Conditions Conducive to the Formation of Ice on Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1393 (1947).
14. —— and Hoecker, W. H., Jr., "Observations of Icing Conditions Encountered in Flight during 1948." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1904 (1949).
15. Lewis, W., Kline, D. B., and Steinmetz, C. P., "A Further Investigation of the Meteorological Conditions Conducive to Aircraft Icing." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1424 (1947).
16. Riley, J. A., "Aircraft Icing Zones on the Oakland-Cheyenne Airway." *Mon. Wea. Rev. Wash.*, 65:104–108 (1937).
17. Samuels, L. T., "Meteorological Conditions during the Formation of Ice on Aircraft." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 439 (1932).
18. Schaefer, V. J., "The Production of Ice Crystals in a Cloud of Supercooled Water Droplets." *Science*, 104: 457–459 (1946).
19. Vonnegut, B., Cunningham, R. M., and Katz, R. E., *Instruments for Measuring Atmospheric Factors Related to Ice Formation on Airplanes*. Mass. Inst. Tech. De-Icing Res. Lab. Rep., Part I, April 1946—Contract No. W-33-038-ac-5443. Reprinted as AAF Tech. Rep. No. 5519, 1946; Part II, March 1948—Contract No. W-33-038-ac-14165. Reprinted as AF Tech. Rep. No. 5727, 1948, Air Materiel Command, Wright Field, Dayton, Ohio.

METEOROLOGICAL INSTRUMENTS

Instruments and Techniques for Meteorological Measurements *by Michael Ference, Jr.* 1207

Aircraft Meteorological Instruments *by Alan C. Bemis* 1223

INSTRUMENTS AND TECHNIQUES FOR METEOROLOGICAL MEASUREMENTS

By MICHAEL FERENCE, Jr.

Evans Signal Laboratory, Signal Corps Engineering Laboratories

INTRODUCTION

During the past several years a number of authors [15, 19, 24] have adequately and critically discussed meteorological equipment commonly employed for weather observations. The purpose of this article, therefore, is to examine the more recent equipment and techniques and focus attention on the degree of reliability of this equipment. Three things will become apparent from this discussion: first, that the major effort during recent years has been devoted to improving upper-atmosphere sounding systems; second, that electronic techniques are playing an ever-increasing and important role in equipment design; and third, that with the manyfold increase in information now possible for weather centrals an urgent need exists for a more efficient utilization of data transmission and presentation systems.

The pattern that will be followed in this article is to present a functional description of equipment or technique, followed by a more detailed examination of the major components of the equipment, with the view of analyzing errors and presenting an over-all evaluation of the system. In an article of this length it will not be possible to present all the data employed in the evaluation, but an effort will be made to supplement the discussion with appropriate references to the literature. No mention will be made of aerographs, sferics equipment, and radar storm-detection equipment, since they are described elsewhere in this Compendium.[1]

RADIOSONDE-RADIOWIND SYSTEMS

General Description. Within the past ten years more time and effort have probably been expended on the development of radiosonde-radiowind systems than on any other single meteorological system. Many systems have been proposed and several have been successfully developed and put into operational use [12, 20, 27, 32]. However, the major portion of the present survey will be devoted to the latest equipment development, and in particular to the redesigned radiosonde-radiowind system just being introduced in this country. The design features of this system are such as to lend themselves to mass-production techniques.

A complete radiosonde-radiowind system consists of a number of major components—flight equipment, including the meteorological sensing elements, radio transmitter, and balloon; a ground direction finder for precise angle measurements; and ground recorders or data-presentation equipment. As constituted at present, the system will yield information on temperature, relative humidity, pressure, and the magnitude and direction of the wind.

One of the earliest successful radiosonde-radiowind systems (placed in widespread use during World War II) is the now familiar Radio Set SCR-658. This set, operating at 400 megacycles per second (mc sec^{-1}), and now being used by the United States Weather Bureau and the United States military services, represented the best design compromise between the availability of inexpensive and efficient radio-frequency oscillators for radiosonde use and the requirement of a narrow antenna pattern. However, it became evident that two sources of error were present in this system: (1) the very broad antenna pattern which limited wind determination to within 15 degrees of the horizon, and (2) the presence of the human link in the tracking system. Both sources of error have been considerably reduced in the design of the new radiowind system operating at 1680 mc sec^{-1} and incorporating automatic tracking features. To take advantage of the increased accuracy possible in wind determination, the sensing elements of the radiosonde have also been improved. A brief, qualitative description will be given here of the over-all operation of the system, followed by a more detailed analysis. The block diagram of Fig. 1 will aid in this description.

The flight equipment of this redesigned radiosonde is basically that first introduced by Diamond and Hinman [10]. Resistance changes of the temperature and humidity elements control the audio modulation frequency (10–190 cps) of a blocking oscillator which in turn deletes 65-microsecond segments from the 1680-mc radio-frequency carrier at the audio rate. This modified 1680-mc carrier is transmitted to the ground and received by a parabolic antenna of the automatic direction finder which thus tracks the airborne transmitter (see Fig. 2). The received signal, suitably modified by the direction finder, is then divided into two parts—one used as an error signal for tracking purposes and the other (containing the meteorological message) fed to a recorder. A record is also made of the azimuth and elevation angles of the direction finder at prescribed time intervals. From these basic data, temperature, pressure, relative humidity, and wind velocity as a function of altitude may be computed. In order to analyze the errors present in this new radiosonde-radiowind system it is necessary to discuss its major components in some detail.

Radiosonde Flight Equipment. A typical radiosonde now in general field use by the Weather Service of the United States Air Force is illustrated in Fig. 3. This instrument consists of a meteorological modulator, a radiosonde transmitter, and batteries. The modulator

1. Consult articles by A. C. Bemis, M. G. H. Ligda, R. Wexler, and R. C. Wanta.

unit includes a selector switch activated by a pressure-sensitive aneroid capsule (Fig. 4). In the present design the selector switch has 150 silver conducting-reference signals for pressure measurements, and all other conducting strips up to the 104th are associated with humidity measurements. Humidity is not meas-

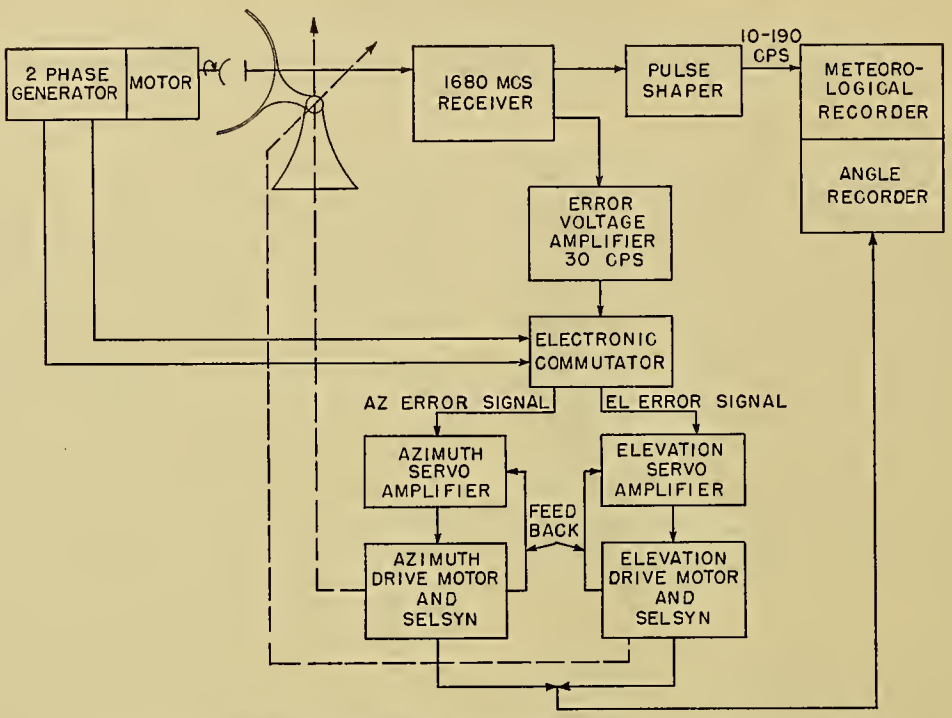


Fig. 1.—Block diagram of the latest automatic balloon-tracking direction finder.

strips, each separated by an insulating segment. Every fifth conducting strip is 0.0065 in. wide; all other conducting strips are 0.004 in. wide; while the insulating

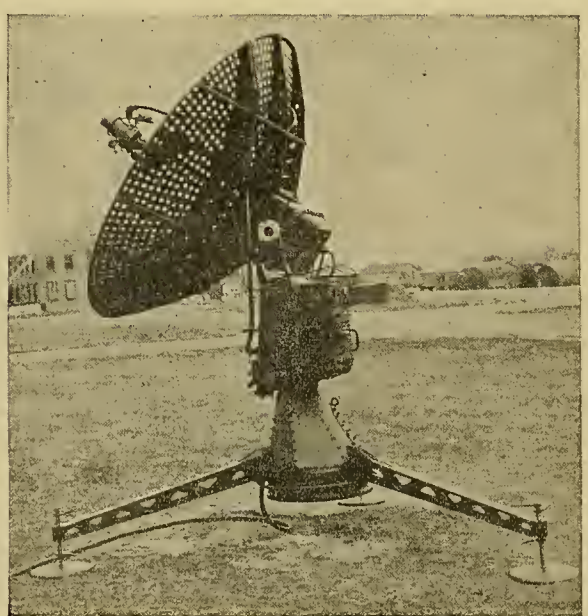


Fig. 2.—An automatic balloon-tracking direction finder set up for operation. Rawin Set AN/GMD-1.

segments are about 0.008 in. in width. Each insulating segment is associated with a temperature measurement, each 5th and 15th conducting strip is associated with

ured after the 104th contact, the remaining conducting strips being used for reference signals. The 150 conducting strips are printed on a polystyrene block to reduce fabrication costs, to avoid complicated wire connections, and to increase the uniformity of the segment alignment.

Pressure Capsule. The pressure capsule employed in this new design is temperature compensated and constructed so as to give approximately logarithmic response over the operating range—the deflection per millibar at 50 mb being approximately five times the deflection at 1000 mb. Without affecting the useful sensitivity and accuracy at high pressure, these capsules provide an appreciable increase in accuracy and sensitivity at low pressures.

The precision of pressure measurements depends upon such factors as repeatability of the mechanical motion of the aneroid and its lever system, the temperature compensation, the precision with which the surface pressure can be set on the commutator, and the reliability of the switching operation. It is estimated that the over-all probable error in pressure measurement of the present modulator units is:

±1	mb at	1000 mb
±3	mb at	500 mb
±1.5	mb at	100 mb
±1.5	mb at	10 mb

However, it should be emphasized that the incremental accuracy between contacts is very much

better than the figures quoted above; a probable error of ±0.5 mb between contacts can be assumed.

Temperature. The temperature-sensing element is a rod of ceramic material, 0.020 in. in diameter and 1 in.

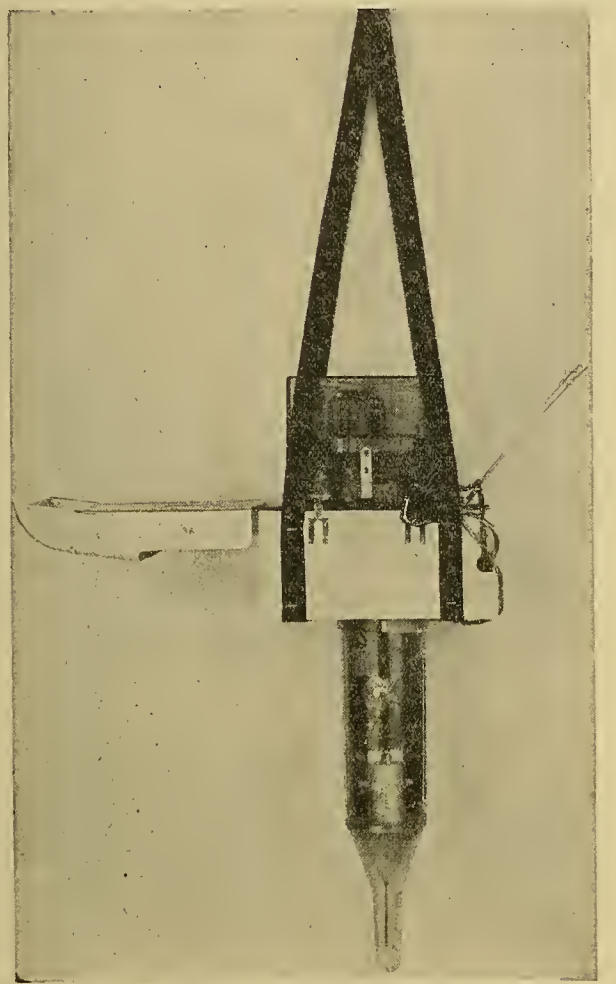


FIG. 3.—Complete radiosonde showing the meteorological modulator, the exposed temperature element, the humidity shield, and the 1680-mc transmitter.

in length. These thermistors possess very high negative resistance coefficients, the resistance varying from approximately 20,000 ohms to 2,000,000 ohms for the temperature range of from +60C to −90C. In order to decrease errors due to solar radiation, the elements are coated with a suitable lead carbonate pigment which has a reflectivity coefficient of about 0.92. The coating increases the diameter to about 0.035 in. The experimental results of Brasefield [6] indicate that such elements may be directly exposed to solar radiation without introducing an error in excess of 0.5C, except at altitudes above 80,000 ft where errors from 1.0 to 1.5C may occur. For these altitudes, however, corrections can be applied. If such corrections are applied, it is estimated that the probable error of the residual is approximately ±0.2C.

Another important parameter of a temperature element is its lag constant, defined as the time required for the element to respond to approximately 63 per cent of the interval between an initial temperature and a final stable value. For the new white temperature element this constant is about 3.5 sec at sea level and for a ventilation rate of 1000 ft min^{-1}. At an altitude

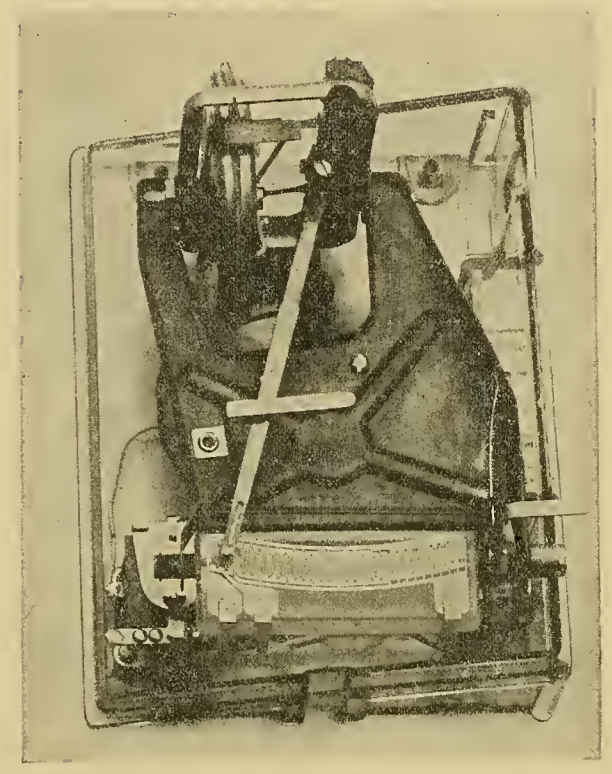


FIG. 4.—The selector switch with pressure capsule.

of 100,000 ft the lag is about 22 sec proportional to $\rho^{-0.4}$, where ρ is the density. If the temperatures are corrected for lag, then the residual lag error introduces a probable error of about ±0.2C in the temperature determination.

To obtain a reasonable estimate of the over-all probable error in determining temperature, it is necessary to consider the following sources of random error: solar radiation (±0.2C), residual lag (±0.2C), transmitter error (±0.2C), lock-in error (±0.1C), calibration error (±0.2C), recorder error (±0.2C), and finally, evaluator error (±0.2C). From the theory of the propagation of errors, it follows that the over-all probable error in temperature measurement is about ±0.5C.

Humidity. Unquestionably the weakest link in the present radiosonde is the humidity element. The element employed is a modification of the original Dunmore [13] design and consists of a polystyrene strip 4 in. long, 2 in. wide, and $\frac{1}{16}$ in. thick, coated with an electrolytic film of lithium chloride dissolved in polyvinyl acetate or alcohol and provided with two electrodes along the two long edges. At room temperature the resistance of the strip varies from approximately ten million ohms to five thousand ohms for a relative-humidity change of from 15 per cent to near saturation. In addition, the resistance of the humidity strip varies with temperature as well as vapor pressure, and ac-

cordingly corrections must be applied for the temperature effect.

As in the case of the temperature element, the absolute values of the humidity-resistance relation may vary over wide limits between elements, provided the shape of the humidity-resistance curves remains reasonably unaltered. This latter requirement permits use of a simple evaluator similar to that employed for temperature reduction.

The humidity strip is subject to several serious errors, some of which can be evaluated and corrections then applied. Others are not easily evaluated, and corrections are therefore difficult to apply. The more important sources of error are polarization, "washing out effect," variation in the shape of the humidity-resistance curve, effect of direct exposure to moisture, lag, insensitivity to humidity changes at very low temperatures, and insensitivity to humidity changes at very high humidity.

Polarization is a phenomenon exhibited by the strip whenever direct current passes through it. The effect is an increase in the resistance of the strip and a consequent lower humidity indication. In the early design of humidity elements, polarization was a rather serious source of error. However, with the use of wide strips (2 in. wide) and consequently higher resistance, polarization has been all but eliminated. For example, a change of less than 1 per cent relative humidity takes place during a 30-min continuous exposure of the element to constant humidity and temperature.

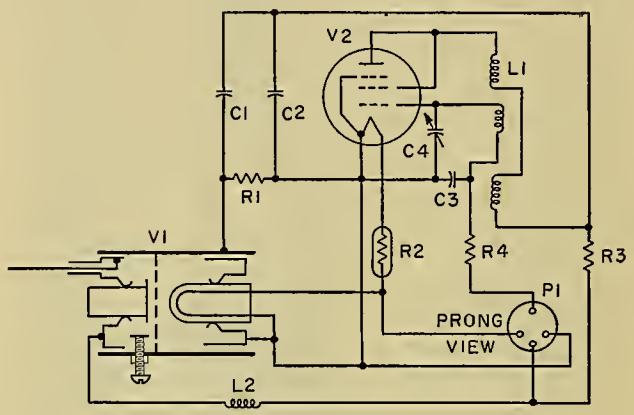
Humidity elements exposed to near-saturation (96 per cent relative humidity) for 30 min and then to air at 70 per cent relative humidity retained their calibration within 1 per cent relative humidity. However, the effect of direct exposure to moisture is far more serious. At this writing there is no unanimity of agreement among investigators as to the extent of damage caused by water droplets on the strip. This point certainly requires more detailed treatment.

A very important characteristic of any humidity element is its lag constant. According to Wexler [33] this constant for the lithium chloride strip (4 in. by ¾ in. by 1/16 in.) depends upon the magnitude and direction of the relative-humidity change and upon the relative humidity from which the change was made, in addition to the marked dependence upon temperature. Although the response characteristics of the strip do not follow an exponential law, for convenience the lag constant will be given for a 63 per cent change in humidity, for the range of 50 to 70 per cent relative humidity. This procedure will give low lag constants.

For ascensional rates of 1000 ft min^{-1}, the lag constant of the present element is approximately 4.0 sec for a temperature of 25C, increasing rapidly with a decrease in temperature, reaching values of 15 sec at 0C and values in excess of 120 sec at −30C. Although these high lag constants of the electrolytic elements leave much to be desired, they are superior to the hair hygrometer in this respect.

Under proper conditions the electrolytic strip can be quite reliable. If the element is not subjected to moisture condensations, it will give relative-humidity measurements with a probable error of ±2.5 per cent for temperatures to −10C and a humidity range of 15 to 96 per cent.

Radiosonde Transmitter. A diagrammatic sketch of the radiosonde transmitter is shown in Fig. 5. This



REF. SYM.	DESCRIPTION				
C1	CAPACITOR,	FIXED,	PAPER	.1 MFD	± 10%
C2				.001	
C3				.02-.03	
C4		ADJ.,	CERAMIC	7-45 MMFD	
L1	COIL	RADIO	R.F.		
L2		R.F.	CHOKE		
P1	CONNECTOR,	MALE			
R1	RESISTOR,	FIXED,	COMPOSITION	1500 OHMS	
R2			AMPERITE	D6M2 REG	
R3				2200 Ω ± 10%	
R4			WIRE	100,000 Ω ± 1%	
V1	TUBE	R.F.	CAVITY	OSC	JAN 5794
V2		JAN	1U4		

FIG. 5.—Circuit diagram of the present 1680-mc transmitter. V-1 is the new cavity oscillator.

transmitter consists of a radio-frequency oscillator operating at 1680 mc sec^{-1} and an electronic modulator for imposing intelligence onto the radio frequency carrier. As shown in the diagram, the radio-frequency oscillator is of the cavity type and is capable of yielding ½ watt of radio-frequency power.

The electronic modulator is basically a blocking oscillator whose repetition rate of 10 to 190 cps is controlled by the variable resistance of the meteorological sensing elements. From basic considerations, it can be shown that the oscillation time of the blocking oscillator should be as short as possible. To meet this requirement, the time constants have been adjusted so that the oscillation time is about 65 microseconds. While the blocking oscillator is on, the radio frequency carrier is quenched, and thus the radio signal reaching the ground has 65-microsecond segments deleted at an audio rate of 10 to 190 cps. To the direction finder, this represents an essentially continuous wave.

The basic reason why the Diamond-Hinman radiosonde lends itself so successfully to mass production is the fortunate relationship that exists between the audio frequency and resistance parameters of the blocking oscillator. It can be shown from an analysis of the circuit that the audio frequency f is given by the following general relation:

$$f = \frac{1}{\tau + bRC}, \qquad (1)$$

where f is the audio frequency, τ a constant proportional to the pulse width, b a constant, R the grid-circuit resistance, and C the grid-circuit capacity. By reducing the pulse width to 65 microseconds and maintaining the product bRC constant, τ was made quite small compared with bRC. Thus, the inverse proportionality between audio frequency and resistance was markedly improved. The accurate evaluation of the meteorological data is based upon maintaining a strict proportionality between two frequencies for two values of R.

The power requirements for this transmitting unit are supplied by a newly developed cuprous chloride-magnesium battery, whose capacity and discharge characteristics are superior to the lead acid cells. When required, the battery is actuated by the addition of water. Considerable internal heat is generated during discharge, hence the usual precautions of thermal insulation for protection against low ambient temperature are unnecessary. A battery pack of sufficient capacity to operate the radiosonde for three hours weighs less than 600 g. The "A" portion of the battery can supply 6.8 v with a 250-ma drain and the "B" section 115 v with a 30-ma drain. A comparison of the discharge characteristics of the lead acid cell and the new cuprous chloride-magnesium cell is shown in Fig. 6.

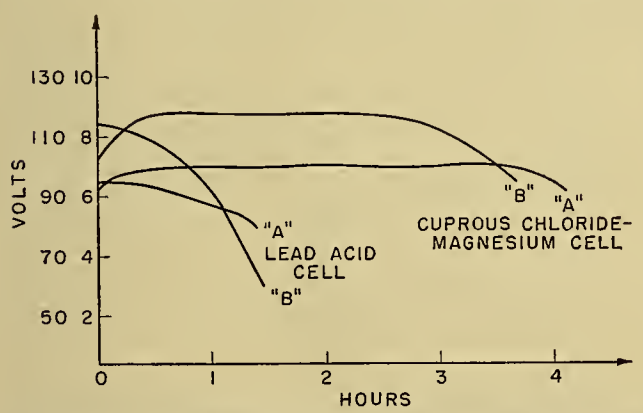


FIG. 6.—A comparison of the discharge curves of the new cuprous-chloride magnesium cell and the standard lead acid cell.

The ambient temperature of the cell was changed during discharge from room temperature to −58F.

The radio transmitter represents an important link in the radiosonde-radiowind system, for serious errors may occur in the meteorological measurements if extra precautions in design are not exercised. For example, it is necessary to keep the frequency drift of the 1680-mc oscillator to within a few megacycles in order that direction finding should not be impaired. In the present design, this drift is maintained in the range of −2 to 4 mc sec^{-1}. Also, changes in the blocking-oscillator rate due to causes other than true changes in the meteorological elements must be kept to a minimum. Audio-frequency shifts of 2 cps on release of the radiosonde and a gradual change of 2.5 cps during an hour flight have been noted. Compensation for these shifts can readily be made in the ground recorder. However, it is estimated that random variations in the blocking-oscillator rate will introduce probable errors in temperature measurement, for example, of about ±0.2C. These newest radiosonde transmitters are quite efficient and reliable and represent a real advance in the telemetering art.

Radiosonde Ground Equipment. The ground direction-finder shown in Fig. 2 serves the dual role of an electronic theodolite [18] and a receptor for radiosonde data. These two functions are discharged by the direction finder in this manner. The essentially continuous waves from the radiosonde transmitter are received by the seven-foot parabolic antenna and its scanning system. As indicated in the block diagram of Fig. 1, the scanning mechanism places a 30-cps modulation on the carrier, the amplitude and phase of the modulation being a function of the vectorial deviation of the target from the axis of the parabola. For radio signals arriving along the axis, the amplitude of the modulation is zero. The modulated carrier is amplified by the receiver, detected, and passed through a band-pass filter to an electronic commutator which yields a pair of d-c voltages, one indicating an elevation error and the other an azimuth error. These error voltages are then applied to their respective thyratron amplifiers for control of the elevation and azimuth drive motors. The receiver also detects the pulses that contain the meteorological data and passes them along to special circuits that amplify and shape the pulses suitably for the remote meteorological frequency-meter and recorder.

As an electronic theodolite [18], the direction finder measures elevation and azimuth angles as basic data. To determine wind speed and direction, it is necessary to know either the height of the radiosonde above a ground plane or the slant range. In the present system, the height of the radiosonde is computed by means of the hydrostatic equation from data obtained during flight. Given the height above the ground for prescribed time intervals and the corresponding elevation and azimuth angles, the horizontal wind velocity may be computed by following the standard theodolite procedure.

In analyzing the sources of errors present in wind measurement, it is necessary to consider the internal precision of angle measurement of the theodolite under dynamic conditions, the errors introduced through the radio transmitting link, and the errors introduced through height determinations. It is estimated that the over-all probable error in elevation and azimuth angle measurements is ±0.05 degrees. This figure takes into account random errors due to the swinging of the radiosonde, propagation errors, dynamic errors in angle measurements at the ground, and errors present in the angle recorder. In estimating this error it is assumed that the elevation angle is in excess of 6 degrees and that a line-of-sight range of approximately 125 miles is not exceeded. Tracking accuracy deteriorates rapidly for elevation angles less than 6 degrees due to ground reflections. For distances in excess of 125 miles, the

signal-to-noise ratio at the receiver becomes unfavorable.

To describe the error in wind-velocity measurement, use will be made of the standard vector-error σ_v. This quantity is defined by the equation

$$\sigma_v = \sqrt{\sigma_s^2 + |V|^2 \sigma_\alpha^2} \text{ (mph)}, \qquad (2)$$

where σ_s is the standard wind-speed error, $|V|$ the wind speed, and σ_α the standard wind-direction error. The standard vector error then represents the radius of the error circle associated with each wind-velocity measurement. Errors in speed and direction are thus conveniently combined. The measurement of σ_s, however, requires knowledge of azimuth elevation angle and height errors, and is influenced by the time interval used for averaging the wind speed $|V|$. With the present equipment, 2-min time intervals are used up to 40,000 ft, 5-min intervals up to 70,000 ft, and 10-min intervals up to 100,000 ft.

It is clear that to ascribe a single standard vector-error to the wind-measuring system is not meaningful unless considerable auxiliary knowledge is also available. Sufficient information for vector-error computation is given in the two graphs of Fig. 7. These graphs represent two wind cross-sections observed at Belmar, New Jersey, Fig. 7(*A*) during December, and Fig. 7(*B*) during July. For ease in computation, only the gross features of this wind structure are shown. In each example the wind speed and direction are shown, together with the computed slant range of the radiosonde. It was assumed that the rate of rise of the balloon was 1000 ft min^{-1}, that the probable incremental error in pressure measurement was ± 1.0 mb (a very conservative estimate), and that the probable error in angular measurement was ± 0.05 degrees. These data are sufficient for computing the standard vector error by equation (2). The numbers at the side of the wind-speed curve represent standard vector errors in miles per hour. Figure 7(*A*) represents an unfavorable situation for wind measurement, Fig. 7(*B*) a favorable one. In (*A*) very strong westerly winds were prevalent, reaching speeds of 100 mph at 40,000 ft. The balloon burst at approximately 120,000 ft nearly 65 miles from the direction finder. The sudden decrease in error from 40,000 to 60,000 ft is due principally to the use of 5-min rather than 2-min averaging intervals. Above 80,000 ft, the errors in height determination become critical. By a judicious use of the rate of rise of the balloon, this error can be reduced by a factor of two or three.

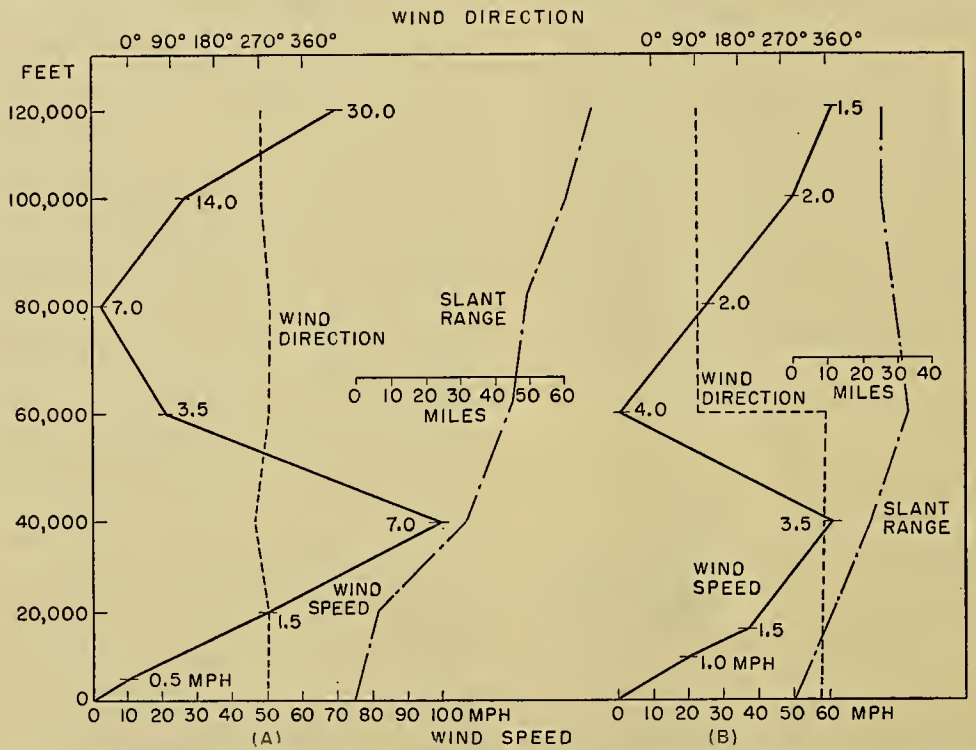


FIG. 7.—To compute the standard vector error expected from the radiosonde-radiowind set, two wind structures observed at Belmar, N. J., in December and July, are shown. In *A*, a strong jet is present at 40,000 ft, with the winds predominantly westerlies up to 120,000 ft. In *B*, the winds from 60,000 to 120,000 ft are strong easterlies. Computed standard vector errors are placed alongside selected height intervals of the wind-speed curves.

In Fig. 7(*B*), the winds were from the northwest, increasing in magnitude to 40,000 ft, then decreasing to nearly zero at 60,000 ft. Above this level the winds steadily increased, blowing from the east and returning the radiosonde to the vicinity of its launching point. The balloon burst at approximately 120,000 ft. For this case the elevation angles for the high altitudes were in the vicinity of 80 degrees, and errors due to height inaccuracies did not predominate; hence the rather low standard vector-error above 80,000 ft.

To improve the accuracy of wind measurements above 80,000 ft, it will be necessary either to improve the accuracy of pressure measurement at these altitudes and take advantage of the rather uniform ascensional rates of the balloon, or to resort to transponder techniques and obtain accurate slant ranges. It is still too early to determine which method is more economical.

Meteorological Recorder. As described above, the properly shaped pulses containing the temperature, pressure, and relative humidity data are fed to a frequency meter which converts the incoming frequency, over the range from 8 to 200 cps, to d-c voltages linearly proportional to frequency. These voltages are measured by a modified Leeds and Northrup-type recording potentiometer. The unit includes circuits for automatic range adjustment. Drifts occurring in the airborne transmitter are automatically corrected by this adjustment. The equivalent frequency is printed on a 10-in. strip chart that covers the range from 0 to 200 cps. The chart can be read within 0.1 of a division, producing an error in temperature of about ±0.2C.

Miscellaneous Accessories. In order to improve the reliability of base-line checks, a portable chamber has been designed to maintain reasonably constant values of temperatrue, pressure, and relative humidity. These quantities are known within 0.1C, ¼ mb pressure, and 1 per cent relative humidity. A circular evaluator that permits the simultaneous lock-in of both temperature and relative humidity is also available. This important innovation was made possible by recognizing that the shapes of the resistance-humidity curves of the humidity strips are nearly identical. The dew-point temperature for any relative humidity may be read directly off the scales. By use of this evaluator the efficiency and accuracy of humidity computations have been significantly increased.

General Remarks. In appraising the over-all performance of this radiosonde-radiowind system, it is important to bear in mind that the design features are such that no individual calibration of the radiosonde after leaving the factory is neccssary, and furthermore that each of the components can be mass produced—on the order of 100,000 per year. The manufacturing tolerances placed on the individual components are indeed quite severe. If the radiosondes were to be used for research purposes, where individual calibration of a small number could be justified, the over-all accuracy of the measurements would be greatly improved—the temperature error reduced to approximately ±0.2C, the pressure errors to ±0.5 mb, and the wind errors for altitudes above 40,000 ft to about one half (or less) of those computed in Fig. 7.

Because of the uncertainties of the solar radiation correction, the effects of moisture condensation, and the lag, it does not appear that temperature measurements with the present thermistors can be made with probable errors less than ±0.2C up to altitudes of 100,000 ft. Very little data are available on the accuracy of temperature measurements within clouds, using the exposed thermistors. The problem has not received the attention that its importance merits.

In order to improve the pressure measurement at high altitudes, serious consideration should be given to the use of the hypsometer. Preliminary measurements on hypsometers, designed for radiosonde use, indicate that in the range of from 50 mb to 1 mb an accuracy of about 0.1 mb should be expected. Furthermore, the use of the hypsometer for the entire pressure range is feasible, provided pressure measurements at 1000 mb to within 7 mb are acceptable. Pressure measurements at the lower pressures would be made with the same percentage error. Some modification of the meteorological recorder would also be necessary.

The problem of humidity measurements in the free atmosphere is indeed the most difficult one. As outlined above, the three major objections to the present lithium chloride strip are (1) the deterioration of the element on direct exposure to water droplets, (2) the very high lag constant at low temperatures, and (3) the insensitivity at very high humidities. Although various salts have been suggested as substitutes for lithium chloride to improve the performance of these elements, it does not appear that the use of electrolytes will overcome the basic difficulties of low-temperature hygrometry and of irreversible dilution when exposed to water droplets. One encouraging development on the horizon is the possible use of a carbon humidity strip. Although very little information is available at present, it does appear that the carbon elements have lag constants of the order of 0.4 sec at 25C; furthermore they are not subject to the dilution difficulties of the electrolytic strips, and are quite sensitive to humidity changes at high relative humidity. The opportunities for research in hygrometry are still very great.

Radar-Wind and Radarsonde Systems. In the preceding paragraphs a description has been given of a wind measuring system that required a knowledge of azimuth angle, elevation angle, and height. Another system in use, capable of accurate wind measurement, employs a microwave radar. Measurement is made by automatically tracking a balloon-borne passive target, in the form of a corner reflector, with a suitable ground radar set operating at either 3000 or 10,000 mc sec^{-1}. Measurements are then made of azimuth angle α, elevation angle θ, and slant range r. In addition, the time derivatives of these quantities, $\dot{\alpha}$, $\dot{\theta}$, $\dot{r}$, are available. With this information, wind velocity can readily be computed. The accuracy of the angular measurement is comparable to that of the direction finders; however, the accuracy in range measurement is greater than the corresponding accuracy in height determination, especially at high altitudes. Furthermore, in the radar system the cosine of the elevation angle is used to compute horizontal distance; in the radio direction-finding system, the cotangent is used. For low-elevation angles, therefore, the radar system is capable of greater accuracy in wind measurement than the radio direction-finding system. The possible limitation of this system for wind measurement is that of range. However, by replacing the passive target with an active transponder or beacon, this limitation can readily be over-

come. The British [17], for example, have developed such a transponder system, and a similar system has been developed for the Navy [22] by the National Bureau of Standards. Wind measurements have been made by the British to ranges of 100 miles and altitudes of 70,000 ft, the latter limitation being that of the balloon. Although no accuracies have been quoted, an analysis of the basic system suggests that the errors in measurement of wind should be less than those of the radiosonde-radiowind system.

Although measurement of wind alone is an important function, the value of a sounding is increased if the temperature, pressure, and relative humidity are also known. In the British [17] system, the incorporation of meteorological sensing elements into the flight equipment in such a manner that the pulses are used as the telemetering channel is now contemplated. Although no entirely satisfactory system has so far been described, there is no basic reason why such a system cannot be developed. The author is of the opinion that an all-electronic radarsonde system with meteorological computers will represent another major step forward in the radiosonde art.

The radiosonde-radiowind system described previously is capable of modification to an equivalent radarsonde system by the addition of a ranging unit. Whether such a modification is desirable, or whether a complete redesign starting with a basic radar set is necessary, will be determined to a great extent by economic rather than technical difficulties.

METEOROLOGICAL BALLOONS

A full description of meteorological balloons and their uses, from the time of the first balloon in 1783 until 1935, has been given by Reger [19]. In these early years, balloons were fabricated from a variety of materials, such as oiled silk, paper, goldbeater's skin, and sheet rubber. The use of rubber latex was restricted to the smaller balloon sizes, such as pilot balloons.

With the advent of the modern radiosonde, tremendous strides have been made in the manufacture of large balloons (350–10,000 g). Two methods have been developed for fabricating these large balloons. In the first method, a rubber gel is formed by introducing latex into a hollow spherical mold immersed in hot water, and then automatically rotating it in such a way that the latex is deposited rather uniformly on the interior wall of the mold. In the second method, a suitable form made of plastic or aluminum is dipped into a coagulant and then into the latex. The latex solidifies and forms a gel on the mold. In both methods the gel is inflated while still wet and soft, and then carefully dried, deflated, and cured. From a wet gel pulled off a 20-in. dipping mold, it is possible to produce a finished balloon 60 in. in diameter that in turn can be inflated to a diameter in excess of 25 ft before bursting. It appears from the rather sparse data available that balloons made from a hollow mold are somewhat stiffer than those produced by the dipping process. Recently, plastic materials such as polyethylene have been successfully fabricated into balloons. It is still too early, however, to evaluate this balloon development properly.

Altitude Performance. The present standard meteorological balloon used to carry radiosondes aloft is made of a synthetic rubber called *neoprene*. The exact ingredients and the technique of fabrication, including curing times and temperatures, are considered trade secrets by the balloon manufacturers. Therefore, it is not possible to correlate these important factors with such balloon parameters as ascensional rates and bursting altitudes. It can be stated, however, that the balloons now supplied in this country for radiosonde work are of remarkable quality. The average daytime performance of these balloons, based on several hundred flights, can be summarized as follows:

Weight of balloon	1000–1400 g
Pay load	1700 g
Free lift	1000–1500 g
Amount of H_2 used	120 ft^3
Approximate initial diameter	6$\frac{1}{4}$ ft
Approximate diameter at 90,000 ft	25 ft
Unstrained thickness	0.004 in.
Thickness at burst	0.0002 in.
Ascensional rate	1000 (+100 ft min^{-1})
	(−50 ft min^{-1})
Altitude attained	90,000 ft (80% of the time)
	100,000 ft (37% of the time)

Very large balloons, fabricated of synthetic rubber, have reached altitudes of from 120,000 to 140,000 ft during the daytime. These experimental balloons, carrying a pay load of about 2000 g, weigh approximately 10,000 g and require nearly 700 ft^3 of gas. The balloons are only partially inflated at the ground and become fully extended at about 30,000 ft. The rate of climb is about 800 ft min^{-1} to 30,000 ft, and then 1100 ft min^{-1} to bursting altitude. However, the performance of these few experimental balloons has been extremely erratic. If fabrication techniques can be developed that will yield balloons of uniform thickness and quality, it is estimated that altitudes of about 150,000 ft can be attained. Balloons of such altitude performance will open up new frontiers for upper-atmospheric research.

Temperature Effects. In the course of testing balloons for performance, it has been observed that differences in temperature between the inside and outside of the balloons for daytime flights may reach values as high as 40C at 50,000 ft, and remain sensibly constant to the bursting altitude. At night, the temperature within the balloon remains about 1C warmer than the outside up to altitudes of 50,000 ft. The temperature gradient then reverses, the inside becoming progressively cooler and reaching temperature differences of about 10C at altitudes around 90,000 ft. Natural-rubber balloons were used for these nighttime flights.

Although the daytime performance of the neoprene balloons can be considered excellent as far as altitude is concerned, the nighttime performance of the synthetic balloons leaves much to be desired. The average height that can be expected is from about 55,000 to 60,000 ft; this limitation is due principally to the freezing of the balloon. By using natural rubber latex,

altitude performance comparable to the daytime performance of the neoprene balloons has been attained.

A curious variation of the average bursting height of the natural rubber balloons with the seasons has recently been noted. During the summer and autumn about 55 per cent of the balloons reached altitudes of about 90,000 ft. During the winter and spring, only 8 per cent reached this altitude. Whether or not this variation in balloon performance can be attributed to the seasonal variation in ozone remains to be determined.

Ascensional Rates. One important way of improving the over-all performance of the present radiosonde-radiowind system is to increase the ascensional rate of the balloons to about 2000 ft min^{-1}. With values of this order, very favorable elevation angles can be attained for wind determination. This objective should be attained without sacrificing the bursting altitude of the balloon and without using an excessive amount of hydrogen. This latter consideration is of practical importance. To date, the use of rubber balloons to attain high ascensional rates has not been entirely successful. However, the use of balloons made of polyethylene material 0.0015 in. thick, and designed to conform to an optimum aerodynamic shape, has given rather interesting results. Teardrop balloons, 22 ft long and 5½ ft maximum width, have attained stable rates of climb in excess of 2000 ft min^{-1} up to altitudes of 45,000 ft.

In studies of the ascensional rates of the 2000-g balloons for varying free lifts Sharenow [29] has found that the ascensional rates of spherical rubber balloons do not continue to increase with increasing free lift. The optimum free lift for the 2000-g balloon is roughly 5000 g, providing the maximum average ascensional rate and a high bursting altitude. Greater free lifts reduce the average rate of ascent and also the bursting altitude. In analyzing the reasons for these effects, Sharenow concludes that the back pressure of the balloon controls in part the initial ascensional rate of a given balloon and thus the over-all average rate.

According to Väisälä [24, p. 151], the back pressure of a balloon of thickness d and unstretched radius r_0 is given by

$$\Delta p = \frac{2d}{r_0} P(n), \tag{3}$$

where $P(n)$ is characteristic of the material and $n = r/r_0$, with r the radius at any time during inflation. In order to improve Δp appreciably, it appears necessary to change the characteristics of the material, since the thickness d and unstretched radius r_0 are reasonably fixed by pay-load and altitude requirements. Sharenow suggests investigating the effect of cure on back pressure.

Summary. These basic problems in balloon development are in need of solution: (1) substantial increase in ascensional rates of balloons without impairing other desirable performance characteristics, (2) design of a balloon to reach heights of about 150,000 ft economically and with a high degree of reliability and, (3) improvement of the nighttime performance of the synthetic balloon.

PARACHUTE RADIOSONDES

Introduction. The parachute radiosonde was designed to be launched from a weather-reconnaissance plane, operating at a maximum altitude of 30,000 ft and with a maximum speed of 300 mph. The basic data of temperature, pressure, and relative humidity, obtained from the flight level to the earth's surface, are transmitted in Morse code signals back to the plane. The equipment consists of the telemetering device, sensing elements, transmitter, and parachute assembly.

Telemetering System. One of the early requirements of this type of radiosonde that greatly influenced its design was that the telemetered data must be received in the plane with the minimum amount of recording equipment. A very simple telemetering system using phonograph-type disks was suggested by J. M. Brady and developed into a practical parachute radiosonde by Brailsford [5]. In the present system, a vinylite disk with 200 grooves is used. Each groove in turn is engraved with a unique set of Morse code letters, so that the exact position of a set of pickup arms can be determined from the code letters. The three pickup arms, arranged around the circumference of the disk at about 90-degree intervals, are mechanically linked to an aneroid capsule, a bimetal, and a hair hygrometer. As the sensing elements respond to changing meteorological conditions, their respective pickup arms sweep across the disk.

Only an 85-degree sector of the disk contains the coded information. As the disk rotates, this sector comes into contact with each stylus of the pickup arm successively, so that the transmitted signal will consist of three code groups representing pressure, temperature, and humidity, followed by a pause. In order that each pickup arm can move freely in accordance with atmospheric changes, the 85-degree sector is raised about 0.62 in. above the rest of the disk. By this device, all arms are free except the one transmitting a signal.

The over-all sensitivity of the coding system is dependent on the number of grooves and the range of the sensing elements. For this instrument, the pressure sensitivity is 5 mb per groove width, the temperature sensitivity is 0.8C per groove, and the relative-humidity sensitivity is 2 per cent per groove. The self-starting unidirectional motor drives the disk at a rate of 12 rpm, so that with a descent rate of 1200 ft min^{-1}, the atmosphere is sampled once every 100 ft.

Temperature. In designing the various components of the radiosonde, it is important to bear in mind the transient shock of $20g$ to which the instrument is subjected at release. This is particularly true of the sensing elements. The bimetal used for temperature measurements is in the form of an elliptical spring, with the inner surface of each leaf as the high-expansion side. With this construction it is possible to use quite thin materials for low temperature lags and at the same time retain a reasonably stiff structure. In addition, it turns out that the element possesses linear

response characteristics over the temperature range of +40C to −70C. The temperature lag constant of the radiosonde is about 15 sec. The over-all probable error in temperature measurement of this instrument is ±1.5C.

Pressure. The aneroid pressure-cell will operate over a pressure range of from 200 to 1060 mb. The probable error in pressure measurement is ±5 mb.

Humidity. The humidity measurements are subject to the inherent limitations of hair hygrometers. Unfortunately, in most operations the hair is exposed very quickly to low temperatures, with the result that the lag constant becomes prohibitively high. Under these conditions the humidity indication can only be considered as qualitative.

Transmitter. The radio transmitter used in this equipment is a one-tube standard continuous-wave crystal-controlled oscillator, operating at specific frequencies in the range of from 2 to 12 mc sec^{-1}. The oscillator is interrupted by a keying relay actuated by the vibrating styli of the pickup arms. The radio frequency output is about ¼ watt, ample for the operational range of the airplanes. The transmitting link introduces no appreciable errors in the meteorological message.

Parachute. The parachute assembly contains a 5-ft nylon parachute with a 135-ft load line. The transmitting antenna is woven into the load line. In addition, the parachute is provided with a time-delay release mechanism to delay the opening of the parachute from 3 to 5 sec. The over-all dimensions of the "parachute radiosonde," including parachute and batteries, are about 10 in. by 6 in. by 19 in., and it weighs approximately nine pounds.

General Remarks. The equipment described above represents a first attempt to design a parachute radiosonde to meet rather severe operating conditions. In spite of the fact that the humidity response is disappointing, as it is in most equipment, the radiosonde does give reliable and meaningful data on temperature and pressure. In addition, the simplicity of data transmission and reception has much to commend it. To improve the humidity response, the use of the lithium chloride strip, or better still the new carbon element, appears as a possibility. The new white-coated thermistors would improve the temperature measurements. However, use of these electrical elements would necessitate either an electrical-mechanical adapter kit, a very undesirable solution, or a complete redesign of the telemetering system. There are many elegant telemetering schemes available today and possibly a suitable one can be adapted for this use.

WIRE SONDES

Introduction. The wire sonde is a specialized piece of meteorological equipment designed to provide temperature and humidity data from the ground to a height of approximately 1000 ft for micrometeorological purposes and, more particularly, for microwave-propagation studies. The typical system includes sensing elements for temperature and relative humidity, ground recording devices, and captive-balloon equipment. Although several systems [1, 2, 28] have been developed and are in use at present, basically they have much in common and differ only in details. The typical system that will be described and used to illustrate sources of error is one that was developed for the Signal Corps.

Airborne Unit. The airborne unit consists of the sensing elements, radiation shields, blower, and batteries; the entire unit weighing approximately ¾ lb. Temperatures are measured by means of a small ceramic bead, similar to that employed in the standard radiosondes. Relative humidity is measured in two different ways, depending on the ambient temperature. For temperatures above freezing, the wet-bulb temperatures are read. A thermistor bead identical to that used for dry-bulb measurements is covered with a light cloth wick immersed in a well of water. For temperatures below freezing, the lithium chloride electrolytic element, previously described, is used. Proper ventilation for these elements is supplied by a small blower, powered by a six-volt cuprous chloride-magnesium water-activated battery.

Recording Unit. The recording device consists essentially of a Wheatstone bridge with the balancing dials calibrated to read either the temperature directly or, by throwing a switch, the wet-bulb temperature or relative humidity. The resistance network of the bridge is compensated for the resistance of the cable leads. In operation, the bridge is electrically connected to the thermistors through a nylon-encased three-conductor cable, 2000 ft long and weighing about 4.5 lb.

Balloon. To maintain the airborne unit at a reasonably constant height during observations, a kite-type balloon called the "kytoon"[2] is used. The kytoon is streamlined and combines the aerodynamic properties of a balloon and a kite. The balloon is approximately 6½ ft long and 39 in. in diameter. During a strong wind, the kytoon lifts like a kite and is not driven downward as easily as a captive spherical balloon. Under average conditions the kytoon has a lift of about three pounds. Calculation of the height of the balloon is made by measuring the length of cable payed out and the angle of the cable leaving the reel. Difficulties in computation arise if a variable wind is present.

Accuracy of Measurements. With the proper exposure and ventilation of the elements, temperature may be measured with a probable error of about ±0.1C over the temperature range of from +40C to −40C. The wet-bulb reading can be made with a probable error of ±0.2C for temperatures above 0C. The lithium chloride strip will yield values of relative humidity to within ±2 per cent for temperatures to −10C and for humidity ranges of 15–96 per cent. Below this temperature, the element rapidly becomes sluggish and its accuracy will strongly depend on the rate of humidity change as indicated earlier in this survey.

General Comments. As stated above, many variations have been suggested for wire-sonde equipment. Investigators at the Naval Electronics Laboratory [1]

2. The kytoon is manufactured by the Dewey and Almy Chemical Co., Cambridge, Mass.

use the lithium chloride strip for humidity over the entire temperature range. At the National Bureau of Standards [28] a wire sonde has been completed that will give accurate heights through the use of a sensitive pressure-altimeter in the airborne unit. This latter development will ease the serious problem of height determination. Practically all of the systems in use today will yield useful information within the lag-constant limits of a few seconds.

AUTOMATIC WEATHER STATIONS

Introduction. Although the need for automatic weather stations has been recognized by meteorologists for a good many years, it has not been until rather recently that developments in radio communications, telemetric techniques, and power units have advanced sufficiently to insure a practical system. In general, a system can be considered practical if it can operate unattended for at least three months and preferably six months under a variety of climatic conditions and furthermore be reasonably transportable.

Several successful automatic weather stations have been designed to meet particular needs [7, 25]. However, for general use in isolated regions where standard meteorological data at the surface are required, two systems have been developed sufficiently to warrant discussion here. One system uses the telemetric technique, now extensively employed in the audio-modulated-frequency radiosonde [9]. The other is based on the "comb" principle of coding, originally suggested by the work of Moltchanoff [27]. The audio-modulated system has recently been described in some detail by Wood [35]. The present discussion will be centered around the code-type station which will be used to illustrate the accuracy of the data obtained from unattended stations and the problems peculiar to this type of research.

Code-Type Automatic Weather Station. The essential components of any unattended station are programming and coding devices, sensing elements, power supply, and communications equipment. Most systems differ in the coding device employed.

In the present system, measures of temperature, pressure, relative humidity, precipitation, wind speed and direction, and light from the sky are transmitted via a two-letter Morse code. In order to accomplish this coding, each indicating element has been so designed that changes in meteorological conditions are translated into angular positions of an arm. To measure the angular position of the arms, use is made of the "comb" principle. The freely moving arm ranges over a series of contacts insulated from each other and arranged in the arc of a circle. Two sets of combs in juxtaposition are utilized, one containing 100 narrow contact pins, and the other containing 10 contacts sufficiently wide for each to subtend 10 of the narrow pins. Associated with each wide contact, therefore, are the 10 narrow contacts, so that a total scale of 100 unique units is provided by the combs. During the measurement of temperature, for example, a depressor bar is actuated and forces the freely moving temperature arm into the teeth of the comb, necessarily touching one wide contact, say, number 30, and one of the narrow contacts associated with it, say number 6. The number "36" then represents a measure of temperature in arbitrary units. At the instant the arm touches the contacts, an electrical circuit through the depressor bar is closed, so that keying of the transmitter with the two-letter Morse code, characteristic of "36," is made possible.

Programming of the station is controlled by a marine chronometer, electrically wound. About two minutes before transmission time the clock turns on the power through a series of relays. While the gasoline engine used to power the station is warming up, various other electrical devices, such as radio tubes, are placed in operational readiness. At the end of two minutes a series of motor-driven cams for switching the sensing elements into the keying circuit are started. During the time the depressor bar keeps the arm of a sensor element in the teeth of the comb, coded signals at the rate of 12 per min are transmitted. The total transmission time required for the meteorological message and station identification is about two minutes. The station transmits once every three hours. The received signal may be taken down by a trained radio operator or fed into a tape recorder of standard design.

The power required to operate this station is supplied by a gasoline-engine-driven generator having a capacity of about 7½ kilowatts. The engine employs a four-cylinder four-cycle liquid-cooled system. The liquid that circulates through the engine is part of a 30-gallon heat-ballast tank required to maintain the temperature of the weather station at about 40F. For operations in cold regions, about one gallon of gasoline is consumed each day.

Conventional radio-transmitting equipment has been adapted for this use. However, since the transmitting ranges may vary between 100 and 1000 miles and since transmitting conditions vary with location and season, the selection of appropriate radio gear has been on an individual station basis. It is expected that 75- and 500-watt transmitters, covering the frequency range of from 2 to 20 mc sec^{-1}, will provide satisfactory ranges for most applications. Provisions are available in the programming unit for converting automatically to appropriate daytime or nighttime frequencies in order to obtain optimum results.

Measurements of temperature, pressure, relative humidity, winds, and precipitation are made with standard weather-station equipment, slightly modified for the coding device. A Bourdon-tube measuring element is employed for the measurement of ambient temperature. The element itself is exposed in a standard instrument shelter. Two sets of combs in series are used to cover the temperature range of from +110F to −60F. The probable error of the temperature measurement is about ±1F.

The pressure unit is a sensitive one-cell aneroid barometer whose indicating arm travels through 720 degrees for a pressure range of from 1050 to 700 mb. For this pressure range, three sets of combs are

combined in series in a circle. Because of the frequency of measurement, no ambiguity will arise in using the same contacts for each 360 degrees of motion of the pressure arm. Probable error in pressure measurement is ±0.5 mb.

For humidity indications, a standard hair hygrograph has been modified by replacing the clock and chart drum with the coding device. For the range of from 10 to 100 per cent relative humidity only fifty contacts are used. Although humidity changes in increments of 2 per cent can be recorded, the accuracy of the measurements is about ±7 per cent for temperatures above freezing, due principally to drift errors of the hair. It should be noted that for all these measurements lag constants are not as critical as they are in radiosonde observations.

A tipping-bucket type of rain gage is employed to determine the accumulated rainfall between transmission times. Accumulated rainfall up to 3 in. can be measured in steps of 0.03 in. Standard rotating-cup anemometers and wind vanes have been adopted for surface-wind measurements. Over the range of from 2 to 150 mph, values are transmitted in increments of 1.5 mph and wind direction in increments of 3.6 degrees. As soon as the equipment is available, measurements of the visibility and of the amount of light from the northern sector of the sky will also be made.

General Remarks. To utilize the full potential value of an automatic weather station, the installation should be operable unattended for as long a time as possible. The design goal of this equipment is from six months to one year. As a matter of fact, this equipment has operated unattended for a little over six months. This achievement, however, should be taken merely as an indication of the possibilities of the newly developed automatic weather station.

One serious problem has always been that of an adequate power supply—a reliable source of power that can operate automatically, intermittently, and under a wide variety of climatic conditions for a long period of time. To date, the best results are obtained with gasoline-engine-driven generators, especially where large amounts of power are required. Wind-driven power plants offer promise for locations where the average winds are high enough to operate the wind generators efficiently and maintain the storage batteries fully charged. Also, for systems with modest over-all power requirements of several hundred watts, the use of banks of Edison primary cells in underground vaults is possible.

For locations where icing is prevalent, serious problems of instrument design arise, particularly for wind measurements. Most rotating-cup anemometers and wind vanes can become totally inoperative and permanently damaged when subjected to severe icing. A specially designed heated Pitot-tube anemometer and wind vane have been designed and are satisfactory, except that excessive continuous power is required to keep them deiced. In addition, the Pitot-tube anemometers are rather insensitive to low wind speeds.

A properly designed instrument shelter is required for the arctic use of unattended weather stations. Rime ice and wind-blown snow can cause havoc with instruments exposed in present standard shelters.

To expand the usefulness of unattended stations, additional instrument research is necessary. There is a need for a reliable humidity indicator for the temperature ranges encountered. A pressure-tendency indicator, a snow gage, a cloud-base indicator, and a visibility meter that are readily adaptable to automatic weather station techniques would collectively expand the importance of the unattended weather stations manyfold.

CEILOMETERS—PULSED LIGHT AND RADAR

Introduction. The general requirement for the measurement of the base of clouds during the daytime as well as at night has been satisfied with the development of the elegant ceilometer by Laufer and Foskett [21]. The problem of the overwhelming background light scattered by a cloud during daytime observations, with the standard light projector, was solved by utilizing a modulated light beam in the projector and a specially designed photoelectric pickup device in the telescope. The projector and telescope were separated by a 1000-ft base line. These ceilometers are in widespread use by the United States Weather Bureau, Navy, and Air Force.

There are, however, two other important requirements in meteorology for cloud height measurements. One is a method for measuring the base of clouds from a single station. The second is a method for measuring the tops as well as the bases of clouds from the ground. In this section, a brief description will be given of two pieces of experimental equipment that show great promise of meeting these additional requirements. One is a pulsed-light cloud-base indicator and the other a microwave radar set capable of depicting cross sections of cloud systems.

Detection of Clouds by Pulsed Light [17, 26]. The principle of operation of this British [17] device is briefly this. Light pulses of about one-microsecond duration are produced by a high voltage spark placed at the focus of a good quality 36-in. Cassegrain mirror. The light pulse reflected from a cloud deck is then received by a second mirror which focuses the light on a photocell. The output of the cell is placed on a cathode-ray tube, using the usual radar techniques of display.

The electric spark produced between aluminum electrodes has a peak power of about 11 megawatts. It is estimated that a flux of about four million lumens is placed on a cloud deck. With the first experimental model, clouds up to 18,000 ft were detected in daylight. It is expected that this device, designed at the Telecommunications Research Establishment in England, will be modified to incorporate a means for recording on paper the height of the base of clouds.

Radar Cloud-Base and Cloud-Top Indicator. It was concluded from theoretical studies [34] that with a radar set operating on a wave length of about one centimeter and with sufficient, though reasonable, peak

power and receiver sensitivity it should be possible to detect and locate clouds and cloud layers. These studies were based on the concept that the microwaves are scattered by water droplets of radii between 10 and 30 microns.

These results were strikingly verified with an experimental radar set designed to operate on a wave length in the one-centimeter region. The equipment was designed so that the radar beam was directed vertically and the returned cloud echoes were presented on a standard "A" type cathode-ray tube. According to Gould [16], clouds were detected to heights in excess of 45,000 ft. It has also been possible to locate clouds through several thousand feet of light rain. However, the minimum altitude that can be detected is about 800 ft, this limitation being due to the length of the pulse and the finite recovery time of the radar receiver.

In order to obtain continuous records of the cloud tops and bases, a conventional-type facsimile recorder has been modified to record the cloud echoes. The recording device gives a plot of altitude of the cloud versus time. In this way a good cross section of the clouds that go by overhead is obtained.

Figure 8 is a typical record made with this radar set. At the top of the photograph is shown the fine structure of a cloud resolved by the equipment. At the bottom of the photograph is a record of a thunderstorm of 23 July 1948. This cumulonimbus built up to approximately 45,000 ft before it gave rise to an intense rainfall. The white vertical lines are due to instrumental difficulties of detuning. The very deep hole to the right of the white lines is probably due to excessive attenuation by the rain. An example of a recording showing two distinct cloud decks is illustrated in Fig. 9. The development of the radar cloud-base and cloud-top indicator represents one of the most significant advances in meteorological instrumentation of the past fifteen years.

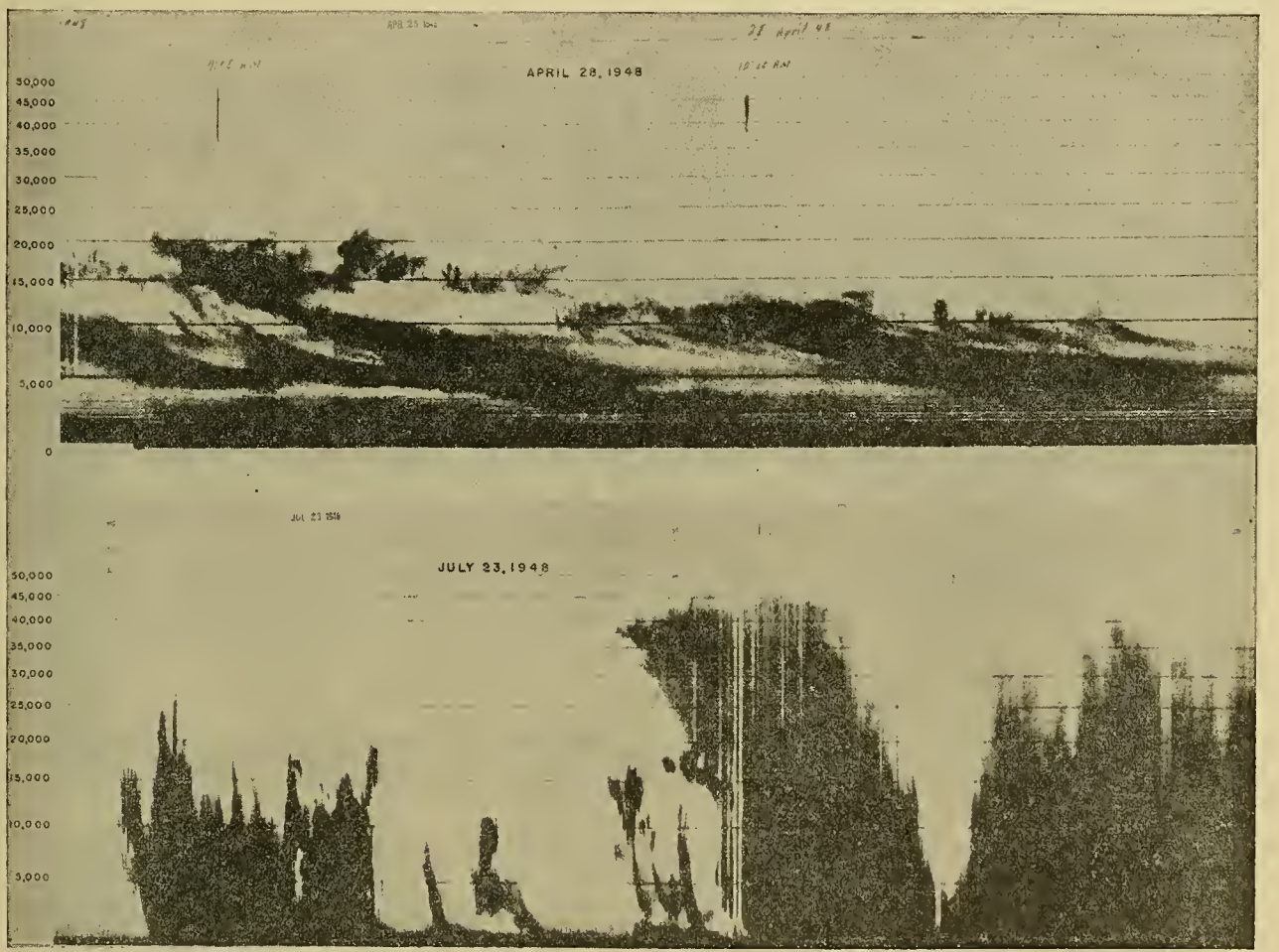


FIG. 8.—Typical cross section of cloud structures taken with an experimental radar cloud-base and cloud-top indicator. Recordings were made at the Evans Signal Laboratory, Belmar, N. J. (*Courtesy, Signal Corps Engineering Laboratories.*)

GENERAL TRENDS

During the past ten years many new and clever experimental techniques have been developed and made available for the solution of meteorological problems. For example, a few of the more obvious advances the last decade has seen are (1) the introduction of radar capable of radiating megawatts of electromagnetic peak power and detecting micro-microwatts of scattered power, (2) basic advances in electro-mechanical systems, (3) basic advances in sensitive infrared devices, (4) development of versatile and rapid analogue and digital

computers, and (5) introduction of new concepts in communication theory. Some of these techniques have already markedly influenced the design of meteorological equipment and increased our understanding of the atmosphere. An even greater use of this fund of knowledge will, of course, be made in the future. In the following paragraphs, a few experimental problems that lie within the scope of this survey will be discussed and possible approaches for their solution suggested.

grometer and have placed it in the forefront as a low-temperature humidity indicator. Dobson, for example, has successfully used this instrument for measurements of water vapor from high-altitude airplanes, and recently Barrett and Herndon [4] have flown the hygrometer to 100,000 ft using the newly developed plastic balloons. Although the equipment used by Barrett weighed approximately 46 lb, there is no reason why this weight could not be materially decreased by the

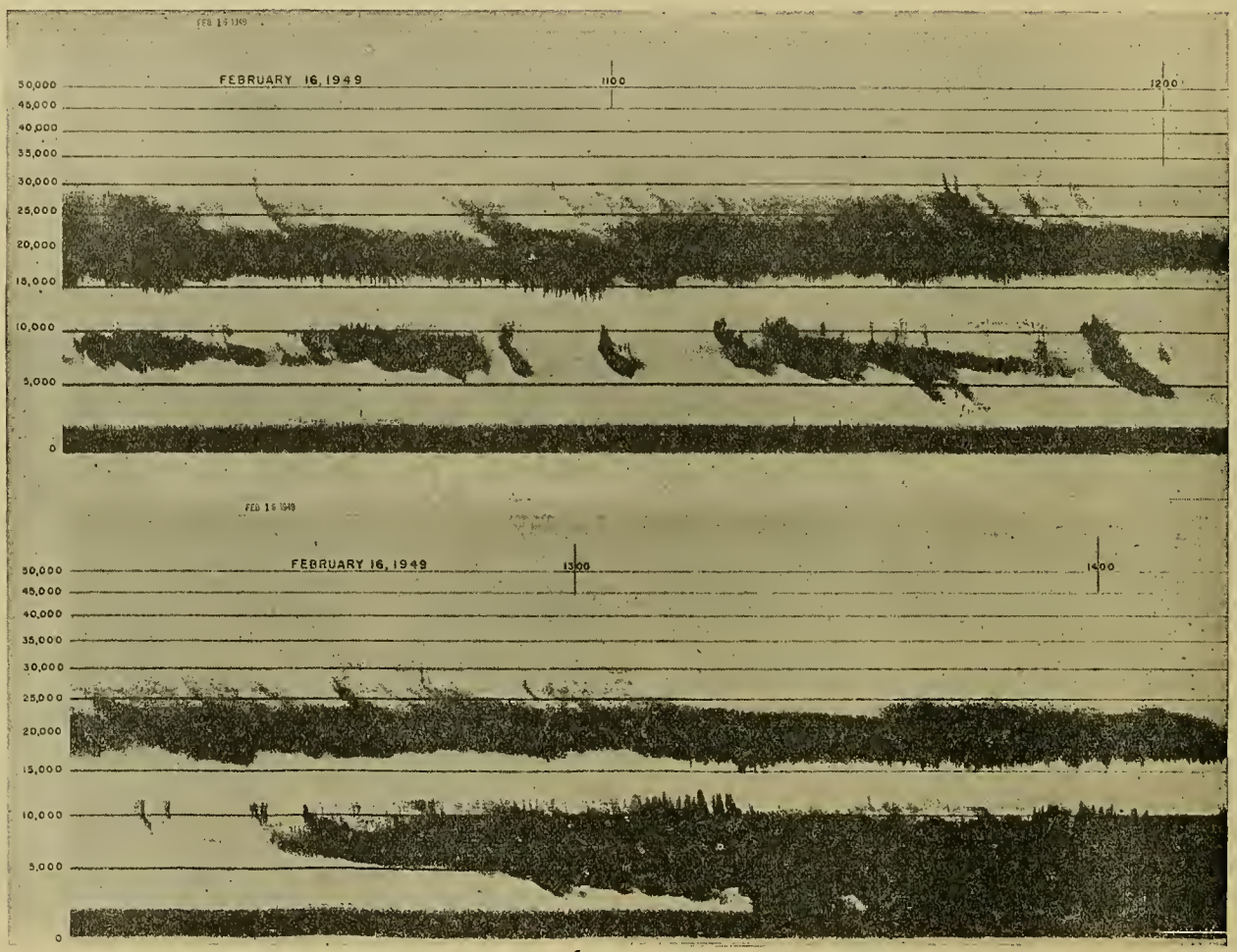


Fig. 9.—An example of a recording showing two distinct cloud decks. (*Courtesy, Signal Corps Engineering Laboratories.*)

One of the most exasperating if not the most important experimental problem in need of a fresh approach is that of low-temperature hygrometry—the measurement of water vapor to temperatures as low as −80C, accurately, rapidly, and economically. As is well known, such standard devices as the hair, goldbeater's skin, electrolytic strip, and wet-bulb thermometer are inherently poor indicators of humidity at these low temperatures. There are, however, two quite promising techniques that should be more fully exploited, namely the dew-point hygrometer and the infrared spectral hygrometer. Such investigators as Dobson [11], Thornthwaite and Owen [30], and the University of Chicago group [31], have contributed significantly to the development of the dew-point hy-

use of printed circuits, light-weight cuprous chloride-magnesium batteries, and a refrigerant such as Freon 13, and thus make possible flights to 140,000 ft, using rubber balloons. The Freon may also be used as a hypsometer. With properly employed surface-temperature measuring techniques, frost-point temperatures to −80C can readily be measured with a probable error of about ±0.5C. In addition to research requirements, the dew-point hygrometer should find wide-spread use for aerographs and for arctic station observations.

Although a spectral hygrometer using the 6.5-μ water-vapor band would appear to be an ideal instrument for humidity measurements, it has not to date been sufficiently practical for field use, due in part to the lack of

sensitivity and ruggedness of the infrared detectors and in part to the lack of precise water-vapor absorption coefficients for low temperatures and humidity. However, this former difficulty should be somewhat eased with the use of the newer photoconductive cell and especially the nonselective pneumatic cell. Both of these detectors can readily be incorporated into stable electronic amplifiers and should result in a quite sensitive and accurate hygrometer.

The possibility of probing the troposphere and middle stratosphere without the use of expendables is taking an encouraging turn, particularly through the use of radar techniques. Experimental equipments have been devised to probe cloud structures and locate the zero degree isotherm within the clouds whenever present. The location of precipitation areas through radar is also well known, and there are data on hand to suggest that some temperature inversions can be detected through the use of radar. For example, the work of Friend [14], using radar of frequencies ranging from 100 to 10,000 mc sec^{-1}, points to the real possibility of discerning air-mass boundaries, surfaces of water-vapor transitions, and other fine structures associated with dielectric changes of the atmosphere, such as turbulence. Furthermore, Jones [17] has made a most interesting suggestion of utilizing pulsed-light techniques for deducing the density field as a function of height.

Recent advances in the art of balloon fabrication and in techniques of radio direction finding will practically insure measurements of winds to about 35 km. Studies of the motion of noctilucent clouds in northern latitudes has given some information on wind velocities from 70 to 90 km. In addition the possibility of estimating winds in the vicinity of 160 km through doppler studies of meteor trails has been reported by Manning and Villard [23]. Recently, Crary [8] outlined a method of obtaining wind estimates from 30 to 60 km by the study of the anomalous propagation of sound. However, in spite of these most interesting techniques, a real need exists for a method of measuring winds on a semiroutine basis at the critical altitudes of 35 to 80 km. The use of smoke trails from V-2 rockets has not been successful at these very high altitudes.

In an earlier section of this article it was stated that the white thermistor element could measure temperatures with an accuracy of ± 0.5C. Two of the important corrections that must be applied are those of lag and radiation, these errors becoming increasingly important at very high altitudes. According to Barrett and Suomi [3], one method of suppressing these sources of errors is to use a sonic thermometer. These investigators have described a very clever sonic thermometer based on a pulse feed-back principle and have claimed such improvements as absence of radiation error and freedom from lag errors. In addition to these advantages, it is emphasized that a sonic thermometer permits space integration of air temperatures as a means for obtaining air temperatures more representative than is possible with point measurements. The question of space integration versus point measurements of temperature and winds is in need of clarification.

The time has perhaps arrived to give serious consideration to the important problem of data-presentation schemes. The amount of raw data presented to the forecaster for analysis and study is truly staggering. He may now have available to him for consideration, in addition to standard surface observations, complete radiosonde and radiowind data to 100,000 ft, detailed data on bases and tops of clouds from various strategic observation stations, sferics data from a world-wide network, detailed radar storm information, and other miscellaneous data. From these raw data he may draw a multitude of thermodynamic diagrams and charts, wind trajectories, velocity fields, acceleration fields, etc., for a variety of pressure levels, for an entire hemisphere, and for several time intervals. Presumably the forecaster would want a three-dimensional presentation of the current state of the atmosphere and of significant variations therein. There appears to be general agreement that atmospheric changes are sufficiently complex that a simple representation of them is not feasible at best.

It is suggested, therefore, that a careful over-all study of systems be undertaken at the earliest possible moment. Such a study would include an analysis of methods of obtaining raw data, of systems for data transmission, of possible utilization of analogue and digital computers, and finally a recommendation of an integrated data-presentation scheme. In the opinion of this writer such a cooperative study between system engineers and meteorologists is necessary to insure continuing progress in forecasting techniques.

It is a pleasure to acknowledge the many valuable criticisms and suggestions given to the author by his colleagues in the Meteorological Branch of the Evans Signal Laboratory.

REFERENCES

1. Anderson, L. J., "Captive-Balloon Equipment for Low-Level Meteorological Soundings." *Bull. Amer. meteor. Soc.*, 28: 356–362 (1947).
2. Anderson, P. A., and others, *The Captive Radiosonde and Wiresonde Technique.* N.D.R.C. Project P.D.R.C. 647, Report No. 3.
3. Barrett, E. W., and Suomi, V. E., "Preliminary Report on Temperature Measurement by Sonic Means." *J. Meteor.*, 6: 273–276 (1949).
4. Barrett, E. W., Herndon, L. R., Jr., and Carter, H. J., "A Preliminary Note on the Measnrement of Water-Vapor Content in the Middle Stratosphere." *J. Meteor.*, 6: 367–368 (1949).
5. Brailsford, H. D., "A New Code Transmitting Radiosonde." *J. Meteor.*, 6: 360–362 (1949).
6. Brasefield, C. J., "Measnrement of Air Temperature in the Presence of Solar Radiation." *J. Meteor.*, 5: 147–151 (1948).
7. Carson, E., "Automatic Weather Stations." *Daily Weather Maps, Wash.*, July 27, 1948.
8. Crary, A. P., "Stratosphere Winds and Temperatures from Acoustical Propagation Studies." *J. Meteor.*, 7: 233–242 (1950).
9. Diamond, H., and Hinman, W. S., Jr., "An Automatic Weather Station." (Research Paper 1318) *J. Res. nat. Bur. Stand.*, 25: 133–148 (1940).

10. —— and Dunmore, F. W., "A Method for the Investigation of Upper-Air Phenomena and Its Application to Radio Meteorography." (Research Paper 1082) *J. Res. nat. Bur. Stand.*, 20: 369–392 (1938).
11. Dobson, G. M. B., Brewer, A. W., and Cwilong, B., "Meteorology of the Lower Stratosphere." *Proc. roy. Soc.*, (A) 185: 144–175 (1946).
12. Duckert, P., "Die Entwicklung der Telemeteorographie und ihrer Instrumentarien." *Beitr. Phys. frei. Atmos.*, 18: 68–80 (1932).
13. Dunmore, F. W., "An Electric Hygrometer and Its Application to Radio Meteorography." (Research Paper 1102) *J. Res. nat. Bur. Stand.*, 20: 723–744 (1938).
14. Friend, A. W., "Theory and Practice of Tropospheric Soundings by Radar." *Proc. Inst. Radio Engrs., N. Y.*, 37: 116–138 (1949).
15. Glazebrook, R., *Dictionary of Applied Physics*, Vol. 3. London, Macmillan, 1923.
16. Gould, W. B., *Cloud Detection by Radar*. Paper presented at Amer. Meteor. Soc. Spring Meeting, 1949.
17. Jones, F. E., "Radar as an Aid to the Study of the Atmosphere." *J. R. aero. Soc.*, 53: 437–448 (1949).
18. Kirkman, R. A., and LeBedda, J. M., "Meteorological Radio Direction Finding for Measurement of Upper Winds." *J. Meteor.*, 5: 28–37 (1948).
19. Kleinschmidt, E., *Handbuch der meteorologischen Instrumente*. Berlin, Springer, 1935.
20. Lange, K. O., "The 1936 Radio-Meteorographs of Blue Hill Observatory." *Bull. Amer. meteor. Soc.*, 18: 107–126 (1937).
21. Laufer, M. K., and Foskett, L. W., "The Daytime Photoelectric Measurement of Cloud Heights." *J. aero. Sci.*, 8: 183–187 (1941).
22. Lyons, H., Freeman, J. J., and Heberling, D. E., *Radar Pulse Repeaters for Upper-Air Wind Sounding*. Paper presented at Amer. Meteor. Soc. Meeting, Jan., 1947.
23. Manning, L. A., and Villard, O. G., Jr., *Meteor Ionization, Wind Studies*. Quart. Progr. Rep. No. 7, Electronics Research Laboratory, Stanford University, 1949.
24. Middleton, W. E. K., *Meteorological Instruments*. Toronto, The University of Toronto Press, 1941.
25. —— and Coffey, L. E., "A Buoy Automatic Weather Station." *J. Meteor.*, 2: 122–129 (1945).
26. Moles, F. J., "The Cloud Range Meter." *Gen. elect. Rev.*, 49: 46–48 (1946).
27. Moltchanoff, P., "Zur Technik der Erforschung der Atmosphäre." *Beitr. Phys. frei. Atmos.*, 14: 39–48 (1928).
28. Randall, D. L., and Schulkin, M., "Survey of Meteorological Instruments Used in Tropospheric Propagation Investigations." *Bull. nat. Bur. Stand.*, C.R.P.L. Report 2-1 (1947).
29. Sharenow, M., *Ascensional Rate Characteristics of Large Spherical Balloons*. Paper presented at Amer. Meteor. Soc. 30th Anniversary Meeting, Jan., 1950.
30. Thornthwaite, C. W., and Owen, J. C., "A Dew-Point Recorder for Measuring Atmospheric Moisture." *Mon. Wea. Rev. Wash.*, 68: 315–318 (1940).
31. University of Chicago, Dept. Meteor., *A Method for the Continuous Measurement of Dew Point Temperatures*, 1945.
32. Väisälä, V., "The Finnish Radio-Sound and Its Use." *Mitt. meteor. Zent-Anst. Helsingf.* No. 35, 28 pp. (1937).
33. Wexler, A., "Low-Temperature Performance of Radiosonde Electric Hygrometer Elements." (Research Paper 2003) *J. Res. nat. Bur. Stand.*, 43: 49–56 (1949).
34. Wexler, R., and Swingle, D., *Theoretical Optimum Wave Length for Cloud Base and Top Indicator*. TM-191-R, Evans Signal Laboratory, Belmar, N. J., Mar. 1946.
35. Wood, L. E., "Automatic Weather Stations." *J. Meteor.*, 3: 115–121 (1946).

AIRCRAFT METEOROLOGICAL INSTRUMENTS

By ALAN C. BEMIS

Massachusetts Institute of Technology

INTRODUCTION

The first scheduled weather flights were made in the early 1930's and were essentially vertical soundings. The airplanes carried simple meteorographs which recorded temperature, pressure, and humidity in much the same manner as the balloon-borne meteorographs of that period. Radiosondes being as yet undeveloped, the use of airplanes was the most practical method of obtaining vertical measurements promptly for forecasting purposes. At present the airplane, with its greater speed and cruising range, is more commonly used for measurements aloft along a chosen horizontal line, and it is for flights of this type that most instrumentation is now designed.

Much of the instrumental development to date has been carried out in connection with highly specialized research programs of various sorts. The study of aircraft icing, for example, has required an extensive program of instrumentation, much of it designed for important meteorological quantities never before measured in the free air [4, 27, 41, 52]. For such a program it is often practicable to use very complex laboratory instruments, expensive to build and operate. The author believes we now suffer from a serious lack of simple, well-engineered instruments, which might be relatively inaccurate, but nevertheless of great value when commonly available. For example, the Thunderstorm Project [5] operating in 1946–47 was unable to make any in-flight measurements of liquid-water content because the only available instruments for the purpose were still experimental. Even temperature measurement was a problem, particularly in the presence of liquid water. In short, aircraft meteorological instrumentation was then and still is in the laboratory.

A more vigorous attack on these problems is recommended. Weather reconnaissance by aircraft is an expensive and important operation, yet relatively small sums have so far been spent on instrumentation for those aircraft. Though many important meteorological quantities are best observed visually, others require instruments, preferably recording instruments. There is need here for inventive genius, good engineering, and funds for development work.

THE PROBLEM

For any observation to be useful it must be located in time and space. Let us assume that location in time is no problem and that ordinary navigational aids will locate the measurements in a horizontal plane with sufficient accuracy. We may also assume that some form of radio altimeter will measure true altitude with sufficient accuracy to give significance to local barometric pressure readings. Then the measurable quantities which are most important to the meteorologist are temperature, humidity, pressure, wind velocity, and liquid (or solid) water content. Also of considerable importance are droplet or snowflake size, turbulence and drafts, and electric field strength. Most of these quantities can be either measured directly or calculated from several related quantities, but are specified above in something approaching fundamental parameters. More precisely, the basic quantities referred to are:

1. *Air temperature* T, in degrees centigrade.
2. *Humidity* q, in grams of water vapor per kilogram of moist air.
3. *True altitude* h, in meters above mean sea level.
4. *Air pressure* p, in millibars.
5. *Wind velocity* V, referring to both the direction in degrees and the speed in knots.
6. *Liquid (or solid) water content* M, in grams per cubic meter.

All of the above quantities can be measured with relative ease at the ground. Most of the difficulties aloft stem directly or indirectly from the high velocity of the measuring "platform." Problems of instrumental lag, accurate location of the point of measurement, and automatic recording are far more acute, and new problems involving kinetic energy arise. Aircraft velocities are about ten times greater than the wind velocities and raindrop fall velocities encountered at a ground station. Quantities involving energy transfer are then one hundred times greater. This means that adiabatic heating effects, and difficulties caused by the presence of liquid water, are so severe as to be unsolved for some conventional instruments. These problems become particularly acute in measuring temperature, and will be treated at some length in the next section, where several new and promising methods are also described.

MEASUREMENT OF TEMPERATURE

The measurement of temperature from aircraft at speeds over 100 knots is seriously complicated by dynamic heating of the thermometer element. The heating is partly due to adiabatic compression of the air and partly to friction. It differs, therefore, for each type of thermometer housing, or lack of housing, and for various mounting locations on the aircraft, but is generally proportional to the square of the true air speed.

This source of error has received careful study [21, 29]. Theoretical and empirical correction factors as functions of air speed are available for the most common thermometer types and are reasonably satisfactory in dry air [47]. Some engineers have even designed built-in compensators working from air-speed devices. A more reasonable approach has been made by others [44, 51] using expansional cooling in the housing to

compensate for the heating at the entrance to the housing and at the thermometer bulb. The most promising is that which uses a vortex chamber for the required cooling, placing the thermometer at the center of the vortex (see Fig. 1). By proper construction of the

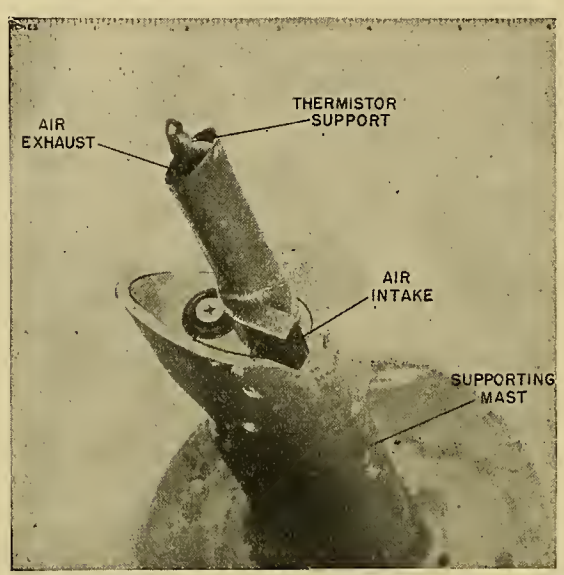


FIG. 1.—Photograph of a "vortex thermometer" as designed by P. J. Harney for measuring true air temperature from an airplane. (*Courtesy of Signal Corps Weather Radar Research at M.I.T.*)

housing and controlled throttling of the flow, the thermometer can be made to indicate free-air temperature correctly over a wide range of air speeds [51]. If another thermometer element is mounted on the same aircraft so as to receive the full adiabatic heating effect, the difference between the two will be a measure of true air speed. The vortex-type housing also provides good mechanical protection for the element. All factors considered, it appears to be the most promising recent development in aircraft thermometry.

Radiation effects, particularly radiation from the sun, may be another source of error. However, with a well designed radiation shield, and with the high ventilation speeds available on aircraft, radiation effects may be made negligible.

The most serious problem is accurate measurement in the presence of liquid water. The effects of liquid water are various. First, any housing so far devised which provides adequate ventilation will allow the thermometer element to become wet, and it will remain wet for a time after emergence into dry air. (First tests of the vortex thermometer in clouds indicate that either its element is wetted to some extent, or water present when the pressure is high causes evaporational cooling "upstream" from the element.) The thermometer will then read some sort of wet-bulb temperature. If it indicated the true wet-bulb temperature, it would be a useful quantity, but it might not be completely wetted and would undoubtedly be affected by the temperature of the water, as well as that of the air. One simple, but only partially adequate, solution is to use two elements, one coated with a hydrophobic substance, the other with a hydrophilic substance, and both exposed to the direct air flow [7]. In air of low liquid-water content, the hydrophobic bulb is then assumed to indicate the dry-bulb temperature (with appropriate corrections). In air of high water content, the hydrophilic element is assumed to measure the wet-bulb temperature (again with appropriate corrections). (See Fig. 3.) The difficulties are (1) obvious uncertainties in the case of intermediate water contents, (2) uncertainties in the appropriate corrections to use, particularly for the wet-bulb temperature, and (3) uncertainties regarding the effects of the temperature of the water striking the thermometer. This general method might be slightly improved by the use of a constantly wetted wick instead of a hydrophilic coating, but it would remain inadequate.

Second, we must consider the effects of water on the temperature of the air before it reaches the bulb, as well as the effects of water on the bulb itself. All heat-transfer-type thermometers must have adiabatic heating effects at the thermometer element or upstream from it. If there is liquid water present, there will be some evaporation which may or may not be at the moist-adiabatic rate. The vortex thermometer corrects for dry-adiabatic heating by subsequent cooling, but it is not yet clearly established that the cooling process will introduce the right correction in the presence of liquid water.

In spite of these heating and cooling uncertainties, it seems worth while to make a further study of means of keeping the thermometer element dry. This would not seem impossible, although the usual means of removing water from the air mechanically requires deflecting surfaces, upstream from the element, which themselves become wet and consequently lower the temperature by evaporation. Even when the air is originally saturated, the temperature rise at the deflecting surfaces due to dynamic effects will lower the relative humidity and permit evaporation. Possibly some modification of the vortex thermometer with porous capillary surfaces to remove the water would be worth trying.

This general problem of temperature measurement in the presence of liquid water has been dealt with at some length because it presents such a challenge. Even if one succeeds in conquering it at temperatures above freezing, then a whole new set of difficulties arises under icing conditions. (Dry snow appears to cause no trouble.) All these difficulties naturally make one turn away from thermometers operating on conductive heat transfer from the air. At least two entirely different principles have been used. In one the speed of sound [1, 49], which is a function of temperature and humidity, is measured. If one defines temperature in terms of molecular motion, the velocity of sound is a more direct measure of that temperature than observing the temperature of some secondary material, as one does with a conventional thermometer. Several very ingenious variations of this sonic-velocity idea have been devised, some of which are suitable for aircraft use. The presence of liquid water, even under icing condi-

tions, would introduce only secondary effects and should pose no insurmountable problem. High accuracy is possible with very short time lags, and there would be no radiation effects. The difficulties seem to be those of a relatively cumbersome installation, the usual but in this case not very difficult deicing problem, interference from sound produced by the aircraft, and an over-all instrument that is very complex compared to a common thermometer. At its present stage of development it stands as an interesting and valuable instrument for research purposes, but it is too complicated for routine use. However, its potentialities are very great.

A second very different principle would use radiant energy. Many radiant-energy thermometers have, of course, been developed, but the author knows of no published work on one for the measurement of free-air temperature from aircraft. However, an instrument has been devised for measuring the temperature difference between the air at flight level and any clouds within the line of sight of the instrument [54]. If the energy received by such a device could be confined to wave lengths strongly absorbed and radiated by the atmosphere (carbon dioxide and water-vapor bands), an accurate free-air temperature device might result. On paper such an instrument appears barely possible with some of the new infrared radiation devices developed over the past ten years [20, 55]. It would be less accurate at low temperatures where energy levels and humidities are also low. In dense cloud, rain, and snow, it would tend to indicate the temperature of the particles instead of the free-air temperatures.

It seems unnecessary to discuss here the relative advantages of thermocouples, resistance thermometers, bimetallic strips, and other types of temperature-measuring devices. The choice is primarily one of the simplicity of auxiliary equipment and time of response. The latter requirement is largely one of the mass and shape of the element. The lowest time lags are usually obtained by using small thermocouples or resistance elements and electronic amplifiers. With such devices, time lags can be reduced to a tenth of a second or less.

It becomes important then to consider what time lags are desirable. We will assume an aircraft travelling at 100 m sec^{-1}. Temperature discontinuities are of two types. Those produced by instability and turbulence will cover a wide scale range. The shorter the time lag the finer the observed temperature detail. For most purposes we are interested only in the larger convective motions with dimensions of several hundred meters, so that a time lag of one second is adequately short. Temperature discontinuities produced by horizontal stratification may amount to several degrees within a few meters but lie nearly parallel to the direction of motion of the aircraft. However, interesting wave phenomena are often observed in such surfaces and may be of importance. Such wave structure is sometimes made visible by clouds, and, if this be a criterion, is seldom less than 100 or 200 meters in wave length. In such cases, to obtain at least a rough measure of the sharpness of the inversion, the time lag should be of the order of a few tenths of a second. Most aircraft thermometers in use today have lags of several seconds, so there is much room for improvement here.

MEASUREMENT OF HUMIDITY

There is a great variety of methods of measuring humidity, both at the ground and aloft. Most aircraft humidity instruments are beset with the same problems as aircraft thermometers, and in addition must be accurate at very low values. Wet bulbs, electrolytic conductivity elements, hair hygrometer types, and all their counterparts become very sluggish and inaccurate at low temperatures and humidities. Some sluggishness is inherent in these types due to the necessity for moisture transfer between the measuring element and the air. This moisture transfer requires ventilation, and again, as in the case of temperature measurement, dynamic heating effects cause errors. The "standard" aircraft instrument at present in use in the United States is the military ML-313/AM, which is simply wet- and dry-bulb mercury thermometers in a special housing [29, 47]. Its use below 0C is awkward, slow, and of questionable accuracy. It has the high time lag characteristic of mercury-in-glass thermometers. In the presence of liquid water its dry-bulb reading has no significance.

Much development work has gone into improvement of electrolytic and hair hygrometers (using the term hair hygrometer in its broadest sense). It seems reasonable to expect that an electrically conducting material may some day be developed which, as an electric hygrometer, will have low lag and constant calibration. The use of carbon shows promise.[1] However, the liquid-water problem must be kept in mind. A drenching in liquid water changes the calibration of some elements, requires a long drying-out period for others, and certainly results in a false reading for most elements when flying through rain in unsaturated air.

As in the case of temperature measurement, the most accurate instruments under all flight conditions are the most complicated. One of these automatically measures and records the dew point [14, 45]. The dew point obviously is unaffected by transient changes in temperature and pressure, so that the sampling problem is relatively simple. When liquid water is present, it will be difficult but perhaps not impossible to obtain a representative sample. The problem is to avoid evaporation or condensation while separating the liquid water from the moist air. Separation of the larger drops is relatively easy, and fortunately more important. When many small cloud drops are present, nearly saturated air can be assumed. With this instrument there may be uncertainty between dew point and frost point due to supercooling. Except for this factor, which can perhaps be eliminated, it is accurate to about 1C and is useful down to dew points of −40C or below. Although this instrument also requires exchange of moisture between the air and the sensing element, it is only a surface phenomenon. Most of the lag occurs in the heating and

1. See "Instruments and Techniques for Meteorological Measurements" by M. Ference, Jr., pp. 1207–1222 in this Compendium.

cooling cycle. It is short, however, a matter of 1 or 2 seconds. This instrument is complex and requires some cooling agent, such as dry ice, for the dew-point surface, but these problems are not fundamental difficulties and have been overcome for certain special studies. Dew-point recorders of this type have been successfully flown in both airplanes [14] and sounding balloons [3].

Another interesting device is the spectral hygrometer [13, 19, 36] operating on absorption of radiation by water vapor. Several absorption bands occur between the visible spectrum and a wave length of 4 μ. Most of these bands are too far into the infrared for good sensitivity with ordinary photocells so that an instrument with satisfactory sensitivity has not yet been developed, but there are newly developed photoconductive cells [6, 20, 43] now available which might make such an instrument more useful. Stronger and very broad-band absorption occurs at and around 6.5 μ where thermal detectors must be used. Here again there are newly developed infrared detectors [20, 55] which might make the instrument practical. Possibly an optical instrument of this type could be arranged to measure attenuation of its light beam due to liquid and solid particles when present in the air, as well as absorption due to water vapor, thus serving a double function. This would require comparison of attenuation over three paths: one restricted to a water-vapor-absorption band, another where gaseous absorption was a minimum, and the third a standard reference path.

It was pointed out when mentioning the sonic thermometer that the speed of sound is a function of humidity as well as temperature, so that the instrument can be used as a hygrometer if the temperature is accurately known. At low humidities, however, its moisture dependence is very small. At present the dew-point hygrometer leads the field in accuracy at low humidities.

MEASUREMENT OF PRESSURE, ALTITUDE, AND WIND VELOCITY

These three measurements are grouped together because they are made with standard aircraft instruments available on any long-range aircraft [47]. Measurement of barometric pressure is the simplest of all instrumental problems. A standard precision aircraft altimeter, when attached to a properly designed static fitting, is accurate to ± 1.5 mb (50 ft) [29]. It is most important, however, that the static pressure fitting be carefully installed and calibrated on each individual aircraft.

Of course, for the pressure reading to have significance as such, the altitude must be known, and accurate measurement of altitude is relatively difficult. Fortunately, constant pressure surfaces have little slope in the atmosphere, so that it is sufficient to fly along such a surface by pressure altimeter, making periodic measurement of true altitude. Knowledge of the barometric pressure at the surface, and the density of the air at all levels from the surface to flight level, permits accurate calculation of true altitude. This information could be obtained by parachute radiosonde, but with poor accuracy.[1] However, altitude above terrain can be accurately measured (± 50 ft) by a radio altimeter [47]. Such an altimeter should be carried by any reconnaissance aircraft for navigational purposes as well as for meteorological purposes. Consequently, the measurement of altitude and pressure offers no problem over ocean surfaces or flat terrain of known altitude. Over mountainous terrain, certain established check points might extend the usefulness of the radio altimeter. On a regular route or airway, radio responders or beacons could be used.

Wind velocity and direction at flight level are determined by drift meters or other navigational aids. These methods are well established and are a routine part of aircraft navigation [47]. The accuracy is generally inferior to the best balloon-tracking techniques, but satisfactory for many purposes.

MEASUREMENT OF LIQUID-WATER CONTENT

Liquid-water content M is the equivalent aloft of precipitation rate at the ground, and just as important a parameter. Its most common values are below 1 g m^{-3}, but values in excess of 5 g m^{-3} have been observed. M should be measured to an accuracy which will match the accuracy of the humidity measurement q, so that the contribution of liquid water to the total water content will have significance. This requires about 10 per cent accuracy, which is not difficult to attain under good conditions. In practice, the observation of water contents below 0.1 g m^{-3} is most difficult, and in that range plus or minus 0.02 g m^{-3} is a common error. Even at -30C saturated air holds 0.34 g m^{-3} of water vapor, so that M measured to plus or minus 0.02 would be close to the required accuracy limits.

Most M-measuring instruments present a trapping surface or opening of some kind to air flowing past the aircraft and measure the water caught per unit time. From the true air speed and the area of the surface, one can calculate the volume swept per unit time and hence M, if one knows the efficiency of collection. No surface is 100 per cent efficient for all drop sizes. The smaller drops are deflected around the surface by the air stream. Because the problem of collection efficiency, in its various phases, is of first importance in studies of aircraft icing, it has received much attention [31]. The first curve in Fig. 2 gives the collection efficiency of a sphere 1 cm in diameter as a function of drop size. It shows that a collector must be at least that small to catch a representative sample of cloud drops. This results in a secondary problem because the low rates of collection are difficult to measure. Also a small collector which is efficient for cloud drops is inadequate for light rain because the wide spacing of the drops may cause a serious sampling error. A satisfactory solution may be a long narrow collecting surface. It is sometimes convenient to use two collectors of different sizes as a means of separating the M due to rain

as compared with the total M. The second curve in Fig. 2 gives the collection efficiency of a large area, in this case the nose of a B-17 airplane. A collector in the

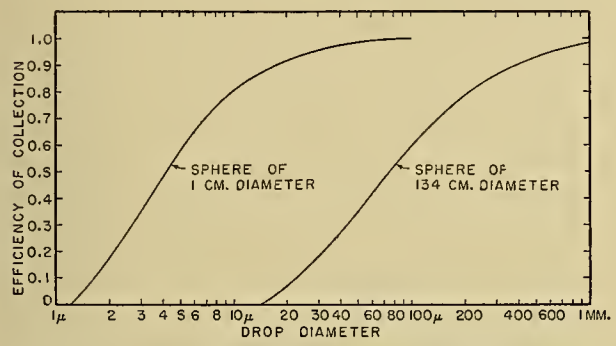


FIG. 2.—Collection efficiencies for cloud and raindrops at the stagnation points of two spheres. The calculations were made by R. M. Cunningham for a 1-cm sphere (showing the performance of a small capillary collector) and for a 134-cm sphere (indicating the collection efficiency for a collector centered at the nose of a B-17 airplane). Pressure was assumed to be 800 mb, temperature +5C, and air speed 200 mph.

center of the nose would be less than 60 per cent efficient for drops of less than 100 μ in diameter.

Having removed the water from the air, we must measure the rate of collection. The most obvious means is to pipe it to a flow meter, but unfilled tubes of any length cause lag. The most widely used M meter is called a "capillary collector" [50] because it employs a capillary porous surface as a collector so as to keep its plumbing full of water at all times. The lag of such a system is limited only by the method of measuring flow rate. Many such methods have been devised, none of them entirely satisfactory [12, 52].

Evaporation from the collecting surface is a serious limitation of the capillary collector and similar types of M meters. Evaporation losses are assumed to be negligible in dense cloud, but may be appreciable in rain; and evaporation makes it nearly impossible to use deicing heat for supercooled water particles. Because of evaporation in clear air, the plumbing system should permit reverse flow or correct for the loss of water in some way.

Another method is to collect both water and air by means of a scoop and determine the amount of water by measuring the amount of heat required to evaporate it, or by measuring the dew points of the air before and after all the water has been evaporated. The latter procedure is inaccurate because it determines a small quantity as a difference between two large quantities [27]. Other types using heat evaporate the water as it impinges on the collecting surface. One can either supply heating power at a constant rate and measure the temperature, or maintain a constant temperature and measure the heating power. Another variation is a heated surface whose electrical conductivity is a function of its wetness. For all of these, the major problems are constancy of calibration and dependence on factors other than M, for example, air temperature, waterdrop temperature, and humidity. The heated types, however, can be designed for use under icing conditions.

An icing-rate meter is a useful instrument for measuring M under icing conditions in supercooled clouds [32]. Such meters are not readily available, but several types have been built, the most successful being some variation of the rotating-disc type [52], in which ice accretion is continuously measured on the edge of a disc rotated at a uniform rate. This type, however, is usually built with a thin disc which has good collection efficiency for cloud drops but does not measure icing rate in rain due to splash and "run-back" effects. Another method is to expose a rotating cylinder or series of rotating cylinders of different diameters to the air stream and weigh the ice accumulation after a known time at a known air speed [39, 52]. This process also gives a measure of the drop-size distribution but is awkward and discontinuous.

For reconnaissance aircraft a simple recording instrument must be chosen. The N.A.C.A. reports success with one of the heated types just mentioned [32]. A small heated cylinder is exposed at right angles to the air stream with a thermocouple measuring and recording its surface temperature at the stagnation point. A sharp drop in recorded temperature indicates entry into a cloud, and subsequent variations give a rough measure of M. The instrument reacts less and somewhat differently to snow, so that snow can be differentiated from rain under some conditions. By using two such heated cylinders, one centered at the nose of the aircraft and one exposed so as to collect small drops efficiently, the records should give a rough idea of drop size as well as of M.

Electromagnetic-radiation devices may prove to be useful instruments for measuring liquid-water content if the drop-size distribution is known [22]. A light-transmission device working in the visible spectrum (a type of visibility meter) seems the simplest type. Assuming a uniform drop diameter d, M is proportional to $d \log_e I_0/I$, where I_0/I is the ratio of transmitted light intensity in clear air to that in the cloud. Flight models of such instruments have been successful [28, 35]. An instrument operating on back-scattered light may also be possible [33]. Microwave radar may be used as well as light, since the intensity of back-scattered radiation at those frequencies is proportional to Md^3 and since radar detection of clouds is possible over short ranges.[2] The strong dependence on drop size is unfortunate for this application.

Radar, of course, is a very complex instrument; but if it is to be carried on an aircraft for other purposes, it should be considered for M measurement. It would have the advantage of indicating the value of the quantity Md^3 over any selected neighboring region instead of only at the aircraft.

In summary, it is clear that rather complex and specialized instruments are necessary for accurate measurement of liquid-water content, but simple and useful instruments such as the N.A.C.A. "cloud indicator" [32] have been devised which can dependably record

2. See "Radar Storm Observation" by M. G. H. Ligda, pp. 1265–1282 in this Compendium.

some quantity related to M. Such instruments should be more widely used.

SPECIAL RESEARCH INSTRUMENTS

Obviously, there are meteorological quantities other than temperature, humidity, pressure, wind velocity, and liquid-water content which are important for particular studies, and in some cases for general meteorological use. Some research programs have required very extensive aircraft instrumentation, and many ingenious instruments have been devised [10]. Figure 3

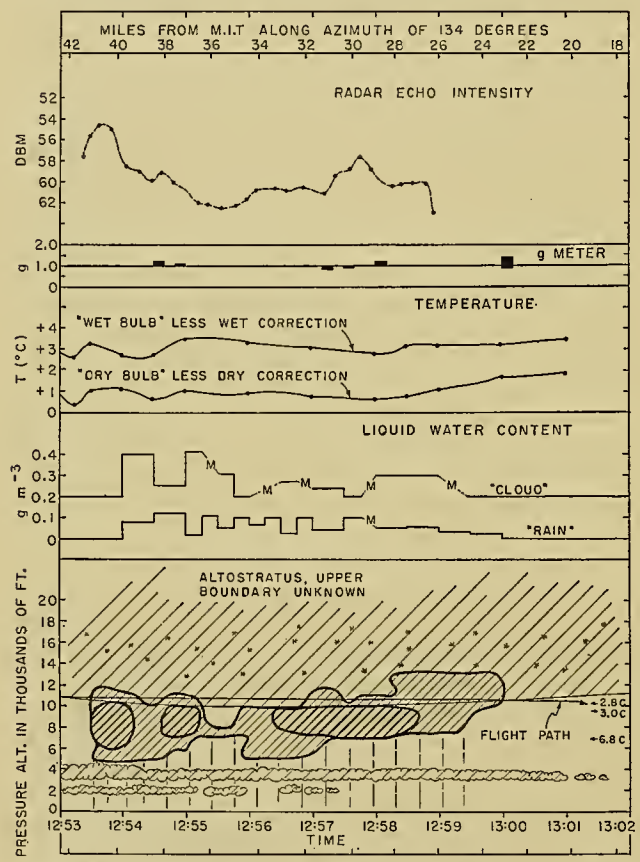


FIG. 3—An atmospheric cross section and plot of instrumental measurements and observations, most of which were obtained from a B-17. At the top is a plot of radar echo intensity (DBM = decibels below one milliwatt) from the storm; next are recorded vertical accelerations; then temperature records, liquid water contributed by cloud and by rain, and a vertical cross section along the flight path constructed from radar information as well as direct observations. The flight was at a level where snow was melting to rain, creating a stratified radar echo as shown by the heavy slant hatching. (*Courtesy of Signal Corps Weather Radar Research at M.I.T.*)

presents in chart form some of the measurements obtained with equipment designed for measurements associated with the uses of radar in meteorology [7]. Aircraft icing studies have also required many specialized instruments. Only a few can be mentioned here.

Cloud and Raindrop Size. A small cloud drop may be only 5 μ in diameter, a large raindrop 5000 μ. So wide a range of sizes requires quite different techniques of measurement. In both cloud and rain there always exists a range of sizes, so that a size distribution is what should be measured. To obtain a representative result, an instrument must sample or measure upwards of a thousand drops.

Cloud drops have been measured at mountain stations by catching samples in oil or on greased slides and measuring them with microscope techniques [23, 25]. Some success with these techniques in airplanes has been reported [8], but a simpler airborne method is to expose a sooted slide [40, 52] to the air stream for a brief known interval and measure afterwards the splash traces remaining. As in the case of liquid-water collectors, the sampling device must be kept small. The rotating cylinder method [39], which makes use of selective collection efficiencies, has been mentioned as a liquid-water measuring instrument. All these methods are seriously limited by being essentially discontinuous and far from automatic.

The measurement of associated optical phenomena [26, 56] shows much promise, as it provides a means of sampling a large number of drops simultaneously, and yields a size distribution instead of individual size measurements. The corona has been used extensively [30], and the rainbow [33] and forward-scattering [24] studied as size measuring methods. Transmission measurements [28, 35] and back-scattering for either light or radar energy[2] might also be used if the water content were known, but such measurements would determine only an effective drop size, not a drop-size distribution. In designing such equipment, it is important to remember that cloud density and general illumination may change rapidly, making it difficult, for example, to locate the angle of a corona or rainbow by means of angular scanning.

Direct photography of the droplets has been attempted [11], but so few drops are found in focus per photograph that a representative sample is not feasible within the short time available. Electrical methods are interesting. One, similar to measuring naturally charged particles [16], calls for inducing and measuring the maximum charge on individual drops as they pass through the instrument. Another measures the electrical effect of drops impinging on a charged probe [18].

Most of these instruments for airborne use have been developed in connection with research programs in cloud physics and aircraft icing, and have resulted in a general knowledge of drop-size distributions for different types of clouds. For more exacting studies, an accurate, short-time-constant instrument with automatic recording is needed.

No satisfactory instrument exists for measuring raindrop size from aircraft. Because the mass of raindrops is more than a thousand times greater than that of cloud drops, most methods which involve physically catching and measuring individual drops are impractical. Soot-coated screens have been tried in place of the sooted slides or solid surfaces used for cloud drops [2]. To obtain a satisfactory sample on one screen would require large unwieldy sizes. Some study has been made of techniques of measuring momentum transfer from individual drops to small collecting plates

[15], but a practical instrument has not resulted. Another approach is to allow the drops to enter a box of quiet air (carried with the airplane) and observe in some way the different rates of deceleration.

A raindrop-size meter which has received much attention to date operates on the attenuation of a small-area light beam by individual drops passing through the beam [34]. This method avoids some of the uncertainties of those which require catching the drops on any sort of collector. For example, the drops are unaffected by the instrument as they are measured, its "collection efficiency" can be 100 per cent for all sizes, and its calibration is only a second-order function of air speed. Attenuation of the light beam causes a phototube to emit an electrical pulse for each drop. The size of the pulse is a measure of the size of the drop, so that sorting and counting circuits can be used to give a nearly continuous measure of drop-size distribution. Calibration can be accomplished with accurately sized solid spheres such as glass beads. The method, however, has a poor signal-to-noise ratio, and is much too complex and expensive at present for more than research use.

Snowflake and Ice-Crystal Size. Observation of snowflake and ice-crystal size and type is obviously of parallel importance to the measurement of rain and cloud particles. It is also obvious that it is a more complex assignment, due to the intricate and varied shapes of snowflakes. An ingenious plastic replica technique has been developed [37] which can be used at the ground or in the air. For aircraft observations, a decelerating tube [38] is mounted forward from the nose of the aircraft to catch the samples with minimum damage by allowing them to decelerate with respect to the aircraft in a cushion of still air. Large crystals and agglomerations usually break apart. Plastic replicas are then made and can be examined at any later time. However, the method is discontinuous and requires the operator's full attention. A count of the number of particles caught in a given time is an indication of the number per cubic meter originally in the air. A better count might be made by one of the raindrop-counting devices, which, for snow, would give little idea of the size but might still count the number per cubic meter.

Other Research Instruments. Many other measurements now made for research purposes may soon become of more general importance. For example, a nuclei count, particularly of nuclei of crystallization, may become a significant observation [42]. Such an instrument would require a representative sample of air, which, in this case, should be easy to obtain on an aircraft since particles as small as nuclei follow the air stream closely.

The Thunderstorm Project [5] made extensive use of vertical gust and draft measurements. Accurate measure of these quantities is extremely difficult because of their very nature, and because the characteristics of the aircraft and the manner in which it is flown enter the measurements. Fortunately, for many purposes only a general indication of turbulence and vertical drafts is required. To obtain such information, the aircraft should be trimmed for level flight and flown with as little control as possible. Vertical accelerations and air-speed fluctuations are then a rough measure of turbulence, and altitude changes are a rough measure of vertical drafts. Standard, relatively simple instruments are available for recording accelerations, air speed, and altitude; but conversion of the measurements to meteorological parameters is difficult [9]. Changes of the angles of pitch and yaw, that is, relative direction of the air stream past the aircraft, may also indicate turbulence, and instrumentation for measuring these angles also exists [48].

A project investigating atmospheric electricity has developed electric field strength meters [17, 53] for airborne use, and instruments for measuring and recording the electric charges on individual precipitation particles. These instruments appear to be entirely satisfactory, and could be engineered for routine use.

Certain other research instruments have already been described. For example, instruments for measuring light transmission and hence visibility were discussed as means of determining liquid-water content. As visibility meters, they would naturally be restricted to low visibilities (less than 1500 m) due to short path lengths [28, 35]. Velocity-of-sound instruments were first mentioned as thermometers. They also have application as hygrometers, anemometers, and possibly other purposes [1].

A rather specialized but interesting instrument has been suggested [36] which would measure directly the refractive index of air at microwave radio frequencies. The index is a function of temperature and humidity.

Some meteorological quantities at a distance from the aircraft can be measured, others observed. Conditions below are measured by "dropsonde," a radiosonde released from the aircraft to parachute to the surface.[1] When precipitation, rain or snow, is present within a hundred miles, or perhaps more, its location can be accurately determined by airborne radar and its intensity estimated [46]. Active research proceeds in this field, so that the usefulness of radar may be expected to increase in the near future. It is already possible to detect clouds over short ranges, permitting the altitude and thickness of cloud layers above and below flight level to be observed. The freezing level can often be located by radar, turbulence within storms indicated, and, as mentioned earlier, some indication of liquid-water content given. The effectiveness of hurricane reconnaissance is tremendously increased by airborne radar. Because of the continuing development of weather radar, and because of its navigational use to aircraft flying in bad weather, the meteorological applications of airborne radar should receive serious study. Further space is not devoted to radar here because of detailed discussion elsewhere.[2]

SUMMARY

It is apparent from the foregoing discussion, first, that meteorological measurements from aircraft are of increasing importance; second, that techniques, however complex, have been devised for a wide variety of

measurements; and third, that simplified dependable instruments are seriously needed both for routine use and for specialized research programs. Improvement in this field of endeavor might be accelerated by a more integrated approach, closer coordination among those working in the field, and more unified objectives. It is time, for example, to consider a "one-piece" instrument which would record a number of the quantities we have discussed above. The military aerograph AN/AMQ-2 [29, 47] was a step in that direction, but it records only temperature and humidity (with serious lag), pressure altitude, and air speed. In a unified instrument many of the quantities could be reduced to the same type of measurement—a temperature or a pressure measurement, for example—and all quantities might be recorded coincidently on a single chart or magnetic tape. It is clear that advancement requires both sound engineering applied to known techniques, and able scientific attack where techniques are now inadequate.

In addition to much helpful advice from many coworkers in this general field, I wish especially to acknowledge the constructive criticism of Robert M. Cunningham who has been actively associated with the development of aircraft meteorological instruments since 1942, and to recognize the able assistance of Mrs. B. A. Wallace in the preparation of the manuscript.

REFERENCES

1. Barrett, E. W., and Suomi, V. E., "Preliminary Report on Temperature Measurement by Sonic Means." *J. Meteor.*, 6: 273–276 (1949).
2. Blanchard, D. C., *The Use of Sooted Screens for Determining Raindrop Size and Distribution.* General Electric Res. Lab. Occasional Rep. No. 16, Project Cirrus, Contract No. W-36-039-sc-38141, Schenectady, N. Y., Nov. 1949.
3. Brewer, A. W., Cwilong, B., and Dobson, G. M. B., "Measurement of Absolute Humidity in Extremely Dry Air." *Proc. phys. Soc. Lond.*, 60: 52–70 (1948).
4. Brock, G. W., "Liquid-Water Content and Droplet Size in Clouds of the Atmosphere." *Trans. Amer. Soc. mech. Engrs.*, 69: 769–770 (1947).
5. Byers, H. R., and Braham, R. R., Jr., *The Thunderstorm.* Washington, D. C., Supt. of Documents, 1949.
6. Cashman, R. J., "New Photo-Conductive Cells." *J. opt. Soc. Amer.*, 36: 356 (1946).
7. Cunningham, R. M., and Miller, R. W., *Five Weather Radar Flights: Measurement and Analysis.* Mass. Inst. Tech. Signal Corps Contract No. W-36-039-sc-32038, Tech. Rep. No. 7, Cambridge, Dec. 1, 1948.
8. Diem, M., "Messung der Grösse von Wolken-Elementen." *Ann. Hydrogr., Berl.*, 70:142–150 (1942).
9. Donely, P., "Summary of Information Relating to Gust Loads on Airplanes." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1976 (1949).
10. Dresios, H. C., *Instrumentation of a B-17 Aircraft for Weather Radar Flights.* Mass. Inst. Tech. Signal Corps Contract No. W-36-039-sc-32038, Tech. Rep. No. 10, Cambridge, March 1, 1950.
11. Elliott, H. W., *Cloud Droplet Camera.* National Res. Lab. Rep. No. M.I.-701, National Research Council, Ottawa, Canada, Dec. 1947.
12. Falconer, R. E., and Schaefer, V. J., *A New Plane Model Cloud Meter.* General Electric Res. Lab. Occasional Rep. No. 2, Project Cirrus, Contract No. W-36-039-sc-32427, Schenectady, N. Y., May 1948.
13. Foster, N. B., and Fosketт, L. W., "A Spectrophotometer for the Determination of the Water Vapor in a Vertical Column of the Atmosphere." *J. opt. Soc. Amer.*, 35: 601–610 (1945).
14. Friswold, F. A., Lewis, R. D., and Wheeler, R. C., "An Improved Continuous-Indicating Dew-Point Meter." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1215 (1947).
15. Gunn, R., *Preliminary Report of the Measurement of the Size and Distribution of Raindrops.* Naval Res. Lab. Tech. Rep. No. A767A, Washington, D. C., Oct. 4, 1944.
16. —— "The Electrical Charge on Precipitation at Various Altitudes and Its Relation to Thunderstorms." *Phys. Rev.*, 71: 181–186 (1947).
17. —— "Electric Field Intensity inside of Natural Clouds." *J. appl. Phys.*, 19: 481–484 (1948).
18. Guyton, A. C., "Electronic Counting and Size Determination of Particles in Aerosols." *J. industr. Hyg. Tox.*, 28: 133–141 (1946).
19. Hamermesh, B., Reines, F., and Korff, S. A., "A Photoelectric Hygrometer." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 980 (1945).
20. Harrison, G. R., Lord, R. C., and Loofbourow, J. R., *Practical Spectroscopy.* New York, Prentice-Hall, Inc., 1948. (See pp. 300–325)
21. Hilton, W. F., "Thermal Effects on Bodies in an Air Stream." *Proc. roy. Soc.*, (A) 168: 43–56 (1938).
22. Houghton, H. G., "On the Relation Between Visibility and the Constitution of Clouds and Fog." *J. aero. Sci.*, 6: 408–411 (1939).
23. —— and Bemis, A. C., "Cloud Particle Studies on Mt. Washington." *Bull. Amer. meteor. Soc.*, 20: 400–401 (1939).
24. Houghton, H. G., and Chalker, W. R., "The Scattering Cross Section of Water Drops in Air for Visible Light." *J. opt. Soc. Amer.*, 39: 955–957 (1949).
25. Houghton, H. G., and Radford, W. H., "On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 6, No. 4 (1938).
26. Humphreys, W. J., *Physics of the Air.* New York, McGraw, 1929. (See pp. 433–555)
27. Jones, A. R., and Lewis, W., *A Review of Instruments Developed for the Measurement of the Meteorological Factors Conducive to Aircraft Icing.* N.A.C.A. Res. Mem. No. A9CO9, Washington, D. C., April 1949.
28. Kampe, H. J. aufm, "Visibility and Liquid Water Content in Clouds in the Free Atmosphere." *J. Meteor.*, 7: 54–57 (1950).
29. Keily, D. P., *Meteorological Instrumentation for All-Weather Operations Section.* Air Tech. Service Command, U. S. A. F., Wright Field, Dayton, Ohio, Oct. 1945.
30. Köhler, H., "Wolkenuntersuchungen auf dem Sonnblick im Herbst 1928." *Meteor. Z.*, 46: 409–420 (1929).
31. Langmuir, I., and Blodgett, K. B., *A Mathematical Investigation of Water Droplet Trajectories.* General Electric Res. Lab. Rep., Contract W-33-038-ac-9151, Schenectady, N. Y., July 1945.
32. Lewis, W., and Hoecker, W. H., "Observations of Icing Conditions Encountered in Flight during 1948." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1904 (1949).
33. Malkus, W. V. R., Bishop, R. H., and Briggs, R. O., "Analysis and Preliminary Design of an Optical Instrument for the Measurement of Drop Size and Free-Water Content of Clouds." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1622 (1948).

34. Massachusetts Institute of Technology, Weather Radar Research, *First Technical Report under Signal Corps Project.* Contract No. W-36-039-sc-32038, Dec. 31, 1946.
35. Orthel, J. C., *Visibility Indicator.* Aeronautical Ice Res. Lab. Rep. AIRL 6034 48-14-7, Air Materiel Command, Dayton, Ohio, March 1948.
36. Randall, D. L., and Schulkin, M., *Survey of Meteorological Instruments Used in Tropospheric Propagation Investigations.* Rep. CRPL-2-1, National Bureau of Standards, Washington, D. C., July 21, 1947. (See pp. 27–28)
37. Schaefer, V. J., "A Method for Making Snowflake Replicas." *Science,* 93: 239–240 (1941).
38. —— *An Air Decelerator for Use on De-Icing, Precipitation Static and Weather Reconnaissance Planes.* General Electric Res. Lab., Contract No. W-33-038-ac-9151, Schenectady, N. Y., Jan. 1945.
39. —— *Demountable Rotating Multicylinders for Measuring Liquid Water Content and Particle Size of Clouds in above and below Freezing Temperatures.* General Electric Res. Lab. Rep., Contract No. W-33-038-ac-9151, Schenectady, N. Y., Oct. 1945.
40. —— *The Preparation and Use of Water-Sensitive Coatings for Sampling Cloud Particles.* General Electric Res. Lab., Contract No. W-33-038-ac-9151, Schenectady, N. Y., April 1946.
41. —— *Final Report on Icing Research.* General Electric Res. Lab., Contract No. W-33-038-ac-9151, Schenectady, N. Y., Aug. 8, 1946.
42. —— "The Detection of Ice Nuclei in the Free Atmosphere." *J. Meteor.,* 6: 283–285 (1949).
43. Starkiewicz, J., "Lead Selenide Photo-Conductive Cells." *J. opt. Soc. Amer.,* 38: 481 (1948).
44. Terada, K., and Yamamoto, G., "Method for Measuring Air Temperature on a High-Speed Airplane." *J. Meteor.,* 4: 201–202 (1947).
45. Thornthwaite, C. W., and Owen, J. C., "A Dew-Point Recorder for Measuring Atmospheric Moisture." *Mon. Wea. Rev. Wash.,* 68: 315–318 (1940).
46. U. S. Army Air Forces, *Radar Weather Reconnaissance.* AAF Manual 105–101–1, Air Force Headquarters, Washington, D. C., May 1945.
47. —— *Aircraft Weather Reconnaissance.* AAF Weather Service Manual 105–128–1, Air Weather Service Headquarters, Washington, D. C., Sept. 1945.
48. —— *Modification of B-17 Being Used in Research and Development of Sferics Equipment.* TI-2017, Addendum No. 91, Air Materiel Command, Wright Field, Dayton, Ohio, Jan. 31, 1947.
49. U. S. Army Signal Corps, *Status of Development of the Sonic Velocimeter.* Rep. No. 21, Evans Signal Laboratory, Belmar, N. J., Oct. 6, 1944.
50. Vonnegut, B., "A Capillary Collector for Measuring the Deposition of Water Drops on a Surface Moving through Clouds." *Rev. sci. Instrum.,* 20: 110–114 (1949).
51. —— "Vortex Thermometer for Measuring True Air Temperature and True Air Speeds in Flight." *Rev. sci. Instrum.,* 21: 136–141 (1950).
52. —— Cunningham, R. M., and Katz, R. E., *Instruments for Measuring Atmospheric Factors Related to Ice Formation on Airplanes.* Mass. Inst. Tech. De-Icing Res. Lab. Rep., Part I, April 1946—Contract No. W-33-038-ac-5443, reprinted as AAF Tech. Rep. No. 5519, 1946; Part II, March 1948—Contract No. W-33-038-ac-14165, reprinted as AF Tech. Rep. No. 5727, 1948, Air Materiel Command, Wright Field, Dayton, Ohio.
53. Waddel, R. C., Drutowski, R. C., and Blatt, W. N., "Army-Navy Precipitation-Static Project, II, Aircraft Instrumentation for Precipitation-Static Research." *Proc. Inst. Radio Engrs., N. Y.,* 34: 161–166 (1946).
54. Warfield, C. N., and Kenimer, R. L., *An Infrared Cloud Indicator.* N.A.C.A. Wartime Rep. No. L18 (ACR No. L 5104), Washington, D. C., Nov. 1945.
55. Williams, V. Z., "Infra-Red Instrumentation and Techniques." *Rev. sci. Instrum.,* 19: 135–178 (1948).
56. Wilson, J. G., "Optical Methods of Measuring the Size of Small Water Drops." *Proc. Camb. phil. Soc.,* 32: 493–498 (1936).

LABORATORY INVESTIGATIONS

Experimental Analogies to Atmospheric Motions *by Dave Fultz* 1235

Model Techniques in Meteorological Research *by Hunter Rouse* 1249

Experimental Cloud Formation *by Sir David Brunt* 1255

EXPERIMENTAL ANALOGIES TO ATMOSPHERIC MOTIONS

By DAVE FULTZ

University of Chicago

Introduction

One of the very old dreams of meteorologists and other scientific observers who have concerned themselves with the phenomena of the atmosphere has been that of solving some problems (as many as possible, of course) by means of experimental work on a small scale. In view of the difficulties of purely theoretical approaches and of interpreting the uncontrollable variations of the actual atmosphere, many have had hopes of obtaining valuable results and insights from such experiments as a supplement to, and a source of, theoretical ideas. Typical statements of such views are those of Schmidt [56, p. 1135] and Richardson [89]. In recent times the increasingly far-reaching successes of model experimentation in aerodynamics, hydraulics, oceanography, and other fields have given renewed impetus to efforts at serious work on meteorological questions by this means.

There are many serious difficulties in the application of model techniques, as they have been developed in other fields, to the study of such a complex system as the atmosphere.[1] Particularly in the medium- and large-scale aspects of atmospheric motions, so many factors are involved that a straightforward dynamic-similarity analysis of the usual kind leads inevitably to the conclusion that, within practical limitations, the model must be identical with the prototype. The choice of real fluids available for experimentation, the serious inconveniences connected with the presence of gravity, and many other factors made this conclusion unavoidable. Basically, therefore, any model experimentation that aims at real resemblance to the atmosphere of a planet and thence to the discovery of principles governing phenomena in such atmospheres, will have to accept deliberate distortions of certain kinds and will have to intercompare many partial types of experiments and analyses in the effort to arrive at valid conclusions concerning the operation of these principles. In this respect the situation is similar to, though more complicated than, that which is encountered in sedimentation studies on river models.[1]

I hope in the succeeding parts of this paper to review briefly some of the rather considerable amount of experimental work of various types which has been done with the atmospheric problem more or less in mind, to mention some of the reasons for expecting a renewed and vigorous attack on this area to be more profitable in the future than it was in the past, and to suggest some immediate directions along which such an attack might develop. With some exceptions, the topics to be discussed will be restricted to problems concerned with relatively large-scale motions, such as cyclones or the general circulation. Experimental work on certain small-scale phenomena, such as convectional layers and flow in the friction layer, are discussed by D. Brunt[2] and H. Rouse.[1]

General Considerations

In the problem of the atmosphere as a hydrodynamic fluid we are concerned essentially with these factors:

1. Strong effects of rotation.
2. Thermodynamic activity or "convection" as an important ultimate driving mechanism and probably the only one of real significance.
3. Primary effects of "friction" in one sense or another.
4. Important alterations brought about in the operation of all the preceding factors by the great horizontal extent of the atmosphere relative to its vertical extent.

The last factor, perhaps more than any of the others except rotation, fixes the specific character of an atmospheric motion and must somehow be dealt with in the interpretation of experiments. Since these factors are inextricably combined with other dynamic factors which are more fully understood and more completely analyzed, the task of disentangling them and assigning each to its proper place is formidable.

In an experimental approach, which must of course be guided as closely as possible by theory and observation, a more than ordinary amount of groping will be inevitable. Similarity analyses need to be made but, for example, even the significance of Π-terms such as $g/a\Omega^2$, where g is the acceleration of gravity, a is the radius of the experimental sphere or a planet, and Ω is the angular velocity of the sphere, is not the same in the heating experiments in a thin spherical shell reported by Fultz [22] (and discussed below) as for the planet. In the planetary case, g is the intensity of a central force, and the solid surface is an equipotential surface for gravity plus the centrifugal force. In the experiments mentioned above, g refers to the intensity of a uniform field of force, and the spherical surface is not an equipotential surface. In spite of this and other differences, kinematic similarities at least did appear in these experiments; furthermore, there are qualitative theoretical reasons for expecting this to be so. Thus the fact that kinematic similarities were found is not likely to be mere coincidence.

It would be foolish to expect these similarities to extend to every aspect of the dynamics of these two systems. But it is not unreasonable to expect similarity in some predominant, basic mechanism. As suggested

1. Consult "Model Techniques in Meteorological Research" by H. Rouse, pp. 1249–1254 in this Compendium.

2. Consult "Experimental Cloud Formation" by D. Brunt, pp. 1255–1262 in this Compendium.

by the theory from which these experiments developed, this mechanism, if substantiated by further checks, would principally be "frictional" (turbulent) mixing processes that are similar in the two cases.

The obvious step beyond the purely kinematic check of average velocity fields is to proceed to more refined measurements of important dynamical quantities over a more extensive range of parameter values. Once an accurate picture is obtained of how far qualitative and quantitative similarity extends, many avenues of reasoning and experiment will be opened for attempts to interpret both the agreements and the disagreements in terms of unifying principles.

An important step in this process will be to attempt to produce experimental arrangements which have close resemblances in one or two selected Π-parameters at the expense, in general, of similarity in others. Perhaps the most important example of this possibility is the case of convection in thin fluid shells on a rotating paraboloid as compared with the spherical shell experiments mentioned above. If the angular velocity of the paraboloid is such that the resultant of the gravitational and centrifugal forces is exactly perpendicular to the surface at every point, the solid surface will have the planetary property of being an equipotential surface. More realistic thermodynamic actions should be possible in this arrangement than in the above-cited experiments with a rotating spherical shell. A free upper surface can be attained and the Coriolis parameter would still be variable, but all these advantages are obtained, of course, at the expense of abandoning any approach to geometrical similarity to planetary shapes and of tolerating rather considerable horizontal variations of the apparent gravity acceleration. Various other practical advantages, such as the ability to produce vigorous convectional mixing with practicable temperature differences within the fluid, may possibly also be lost. By appropriate modifications of this or other types it should be possible so to vary the contributions of the three major factors in these experiments that each one can be more or less isolated for study.

In developing an experimental program, a certain backlog of experience will have to be built up on the purely practical side. Since the requirements of importance to meteorological questions involve somewhat different emphases than, for example, in aerodynamic experimentation, trials of satisfactory techniques in themselves will occupy a considerable period. However, the present development of experimental technique, both in general and in specific types of model experimentation, appears very favorable. Many alternatives are available which will require only slight further development to become adaptable to such a program.

In matters both of interpretation and of technique, any experimental problems which can be analyzed theoretically with relative completeness will be of great value in bringing difficulties out into the open. For example, the work concerned with the stability of a polar vortex, which will be referred to later, has been particularly illuminating in this direction [23, 42]. In this work, a relatively complete theoretical analysis could be carried out for at least the basic motion and compared directly with the experimental results. By quantitative comparisons, it was found that surface tension forces were making an appreciable dynamical contribution in the phenomenon being studied. Such a contribution, having no counterpart in the major atmospheric problem in view, can be eliminated either experimentally or analytically before drawing conclusions concerning the prototype from the model. It is quite doubtful that this effect would have been noticed if a quantitative theory had not been available. On the other hand, with this experience in mind, it will be easier to make qualitative estimates of the magnitude of such effects in any subsequent experiments of a similar nature.

Experiments on Rotatory Phenomena—Cyclones and Tornadoes

Almost from the beginning of modern meteorological thought, and after the discovery of the existence of barometric storms, attempts have been made to construct experimental models of cyclones. In fact, many of the cyclone theories of the early nineteenth century seem to have drawn their ideas from such experiments and from comparison with tornadoes. These experiments have usually shown a more distinct resemblance to tornadoes than to present-day ideas of cyclones. An early piece of work of this kind, which is in fact very similar to many later experiments, is that of Wilcke [33, 71] in 1780 at Stockholm. In all probability there were many others of the same general nature during the eighteenth century.

In Wilcke's experiments (Fig. 1), a thick steel wire

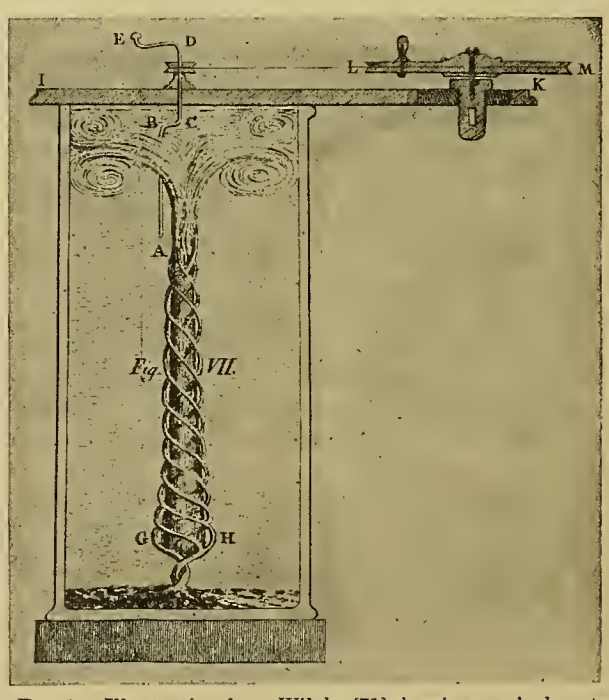


FIG. 1.—Illustration from Wilcke [71] showing a spiral vortex generated in water by mechanical rotation at the top of the cylinder. The cylinder is approximately 14 in. high and 7 in. in diameter. The rates of rotation are not specified but presumably are of the order of a hundred or so revolutions per minute. Wilcke also drove the vortex from the bottom, investigated an alcohol-oil system, and worked with vortices generated by outflow from a hole in the bottom of the cylinder.

a few inches in length was mounted eccentrically and rotated about the central axis of a large cylindrical container. A vortex was generated along the central axis with outflow at the wire end and inflow at the opposite end of the container. The motions were followed by means of burnt lime in the water.

All sorts of modifications of this experiment have since been made. Some investigators have used mechanical means, such as fans or other devices, for generating the vortex in both air and liquids. In this group were Andries [5], Hirn [82], Dechevrens [13], Weyher [45] (Fig. 2), and Dines [15], all in the last two decades of the 19th century, and Hale and Luckey [30], Exner [17], Lunelund [43], and Letzmann [41] (an extensive work on air vortices) since 1900. Others have investigated the vortices produced by mechanical removal of fluid in a field of gentle, initial rotation (the bath-tub drain experiment), for example, Aitken [3, 4] between 1900 and 1918, Letzmann [40] who worked with water in 1927, and also some others of the previous group. Among those who have investigated the vortex motions induced by the "secondary flows" associated with friction, especially where angular accelerations of the entire system are involved, are von Bezold [8] and Belopolsky [6, 7] in the latter part of the nineteenth century and Ahlborn [1] in 1924. The final group are those who have worked with thermally generated motions producing either isolated vortices or more complex fields of rotational motion. This group includes Vettin [67–69] (Fig. 5) and Czermak [12] (Fig. 3) before 1900, and Aitken [3], Exner [17], Sipinen [59], and Terada and Hattori [65], between 1915 and 1926. (Summaries of most of the above-mentioned investigations and other similar ones are given in [11] and [70].)

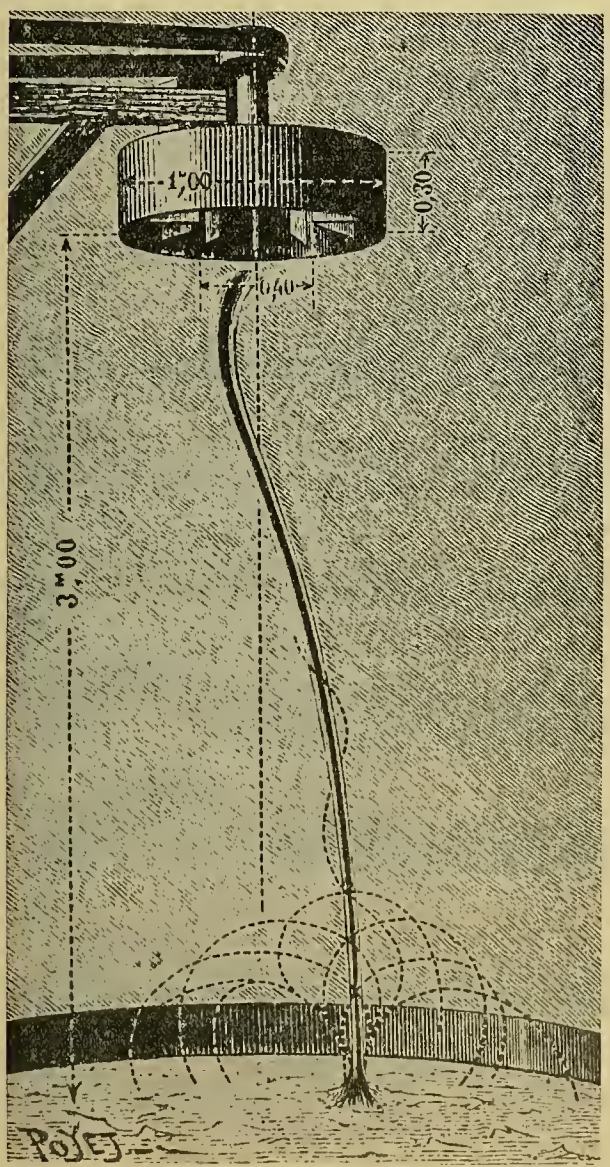


FIG. 2.—Engraving of a tornadolike vortex produced by Weyher in the open above a large basin of water. The fan at the top produces the motion. (*After Mascart* [45].)

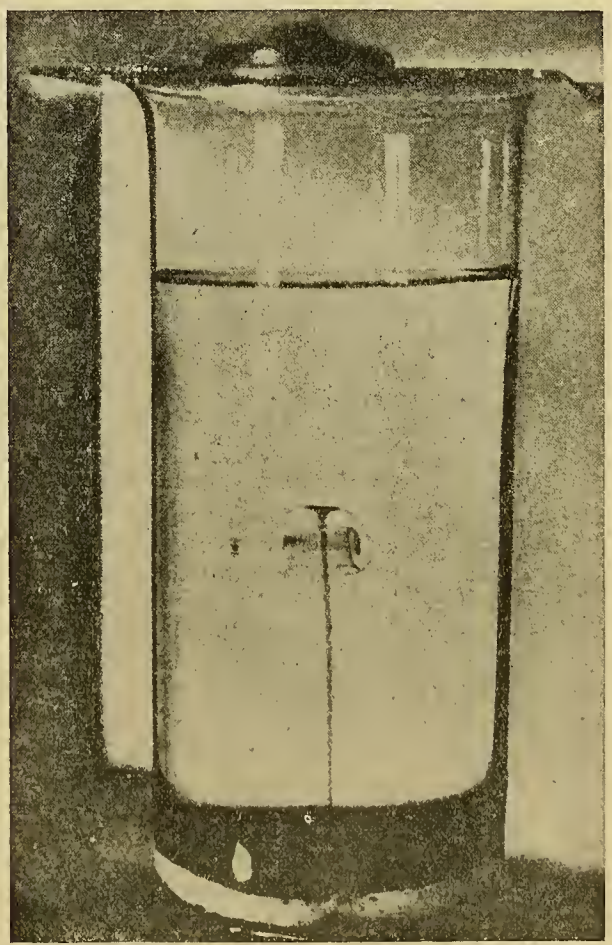


FIG. 3.—Photograph by Czermak of a convective column in a motion similar to those described by Oberbeck [47]. The water in the beaker is gassed-out and allowed to come to complete rest. Heating at the bottom of the beaker is produced electrically by a spiral of fine platinum wire 6 mm in diameter. Note the delicacy of the vortex motion at the cap. Czermak also investigated the effect of stratification on these motions. (*After Czermak* [12].)

It is difficult to sum up the net significance of all this work with respect to the problem of the cyclone or even the tornado. Many curious and beautifully well-defined phenomena are discussed in these studies, but their actual implications for the questions which, in most cases, they were specifically intended to illuminate

are still uncertain. The two atmospheric phenomena to which these studies show the most resemblance are tropical cyclones and tornadoes. Generally speaking, aside from qualitative details, the experiments lead only to conclusions that result also from elementary theory and go back to the seventeenth and eighteenth century beginnings of the discussion. An example of this is the conclusion that necessary conditions for vortices of these types are concentrated regions of ascent or descent (thermally produced or due to the initial field of mechanical motion) combined with initial absolute circulations in fluid circuits enclosing the regions. One fairly general result obtained by experiment, that as far as I know was not anticipated in theory, concerns a general tendency of concentrated vortices to arrange themselves so as to end at a boundary, for example at the earth's surface (Fujiwhara [21], Hale and Luckey [30]). Thus one might generate a horizontal vortex and have it break and arrange itself so as to meet the surface vertically. (See Wegener [70] for tornado developments of this kind, or Fujiwhara for possible synoptic applications.) This result is, of course, probably related to Helmholtz' proposition that a vortex filament cannot have a free end in the interior of a fluid in continuous motion.

However, one gets the impression that, at least with respect to the tornado problem, there is something worth studying in much more detail in some of these experiments. The phenomena in Letzmann's work, or in that of Dines and Weyher in air with water vapor as the tracing element, show almost complete qualitative similarity to actual tornadoes, even to the hollow sheath apparent along the tornado tube in many photographs, or to the swellings traveling along the tube as noted in a few descriptions of the natural phenomenon [70]. It is a curious fact in this connection that hardly anywhere in these investigations, or for that matter in most of those to be mentioned later, has any real attempt been made to make numerical measurements of even the simplest and most obvious field quantities. The principal exception among the investigations mentioned above is that of Lunelund [43], who generated vortices in a liquid with a free surface by rotating a simple shaft below the surface. He then measured the geometry of the free surface on photographs, mainly for comparison with the form to be expected for a Rankine combined vortex ("*vr*" vortex in the outer part and solid rotation in the inner part). Another exception is an investigation carried out by Stümke [62] at Göttingen in 1940 on the anticyclonic motions induced by a slow expansion motion from the center of a rotating pan. The measured quantity, the slope of the free surface, showed good agreement with Stümke's theory of the phenomenon.

Experiments Involving Discontinuities—Cold Fronts and Cyclone Systems

The study of discontinuous fluid motions begun by Helmholtz in the 1850's and illustrated in the demonstrations of Oberbeck [47] in 1877 has had a special importance in meteorology since it culminated in the work of Margules and the polar-front studies by the Norwegians. Here also the climate of opinion on the meteorological side was influenced, though probably not essentially suggested, by a number of experimental investigations. Those of special interest to us fall into two groups: an early one which stemmed from studies of squall lines (*Böen*) and later, explicitly, of the cold front; and the other one, much later (after the idea of cyclone families along the polar front had been constructed by V. and J. Bjerknes), which was concerned with systems of vortices developing along similar discontinuity surfaces.

In the first group we have the cold-front propagation studies with liquids by Schmidt [56, 57, 92] around 1910 and by Ghatage [27] in 1936, the interna' wave studies in 1923 by Defant [14] which included some internal hydraulic jumps, the more recent work on jets by Spilhaus [60], and the demonstration possibilities indicated by Haynes [32] and Dinkelacker [76]. These studies have had some historical significance in the development of ideas on cold-front and squall-line phenomena. For example, the nose of the front (*Böenkopf*) (Fig. 4) as discussed by, say, Koschmieder [38], seems

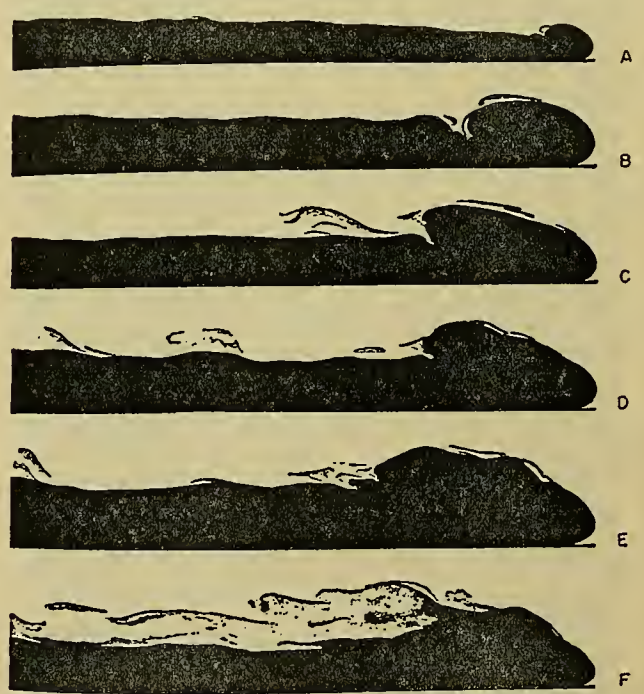


FIG. 4.—Cold-front forms produced by Schmidt in a glass-walled trough 181 cm long, 31 cm high, and 4 cm wide by allowing a salt solution to flow into clear water. The density differences increase toward the bottom of the figure. The pictures were obtained by exposing sensitized paper behind the tank to a magnesium flash. (*After Schmidt* [56].)

to have been suggested primarily by the experimental work. Not very much has actually been done, except by indirect reasoning, to verify observationally the occurrence of this characteristic. (See, however, Weickmann [94].) Similarly the velocity formulas for the front evaluated from the work of Schmidt, Ghatage, and the more recent work of Yih [72] have not been carefully

checked. These all are in qualitative agreement in giving a relatively constant velocity c for a fluid layer of density ρ_1, advancing under its own excess weight through another of density ρ_2. The velocity is given by

$$c = K\sqrt{gh\frac{\rho_1 - \rho_2}{\bar{\rho}}},$$

where h is approximately the height of the nose of the front, and where K has a numerical value from 0.6 to 0.9 for the total depths (two or more times h) used by these experimenters. The fact that the experimental case, during the steady part of the motion, represents a balance between a supply of gravitational energy and frictional dissipation alone, without any influence of tendencies toward Margules-type geostrophic equilibria, suggests that little is to be expected from the result on a large scale. It may, however, be significant in smaller occurrences such as the pseudo-fronts beneath thunderstorms or in very unsteady motions where Coriolis effects have little chance to develop.

The second group of investigations consists of work in air by Sipinen [59] in 1924 and Harwood [31] in 1945. Sipinen's work is particularly interesting because of the use he makes of the very thin, warm, smoky layers which can be put on a smooth surface by puffing cigarette smoke gently over it. Slight currents then induce vortices at the boundaries between clear (cool) and smoky (warm) air layers. Or they can easily be developed and seen via the smoke by slight warming (*e.g.*, by the hand) of a spot on the surface.

Harwood's work, however, is the more significant because it is capable of immediate quantitative extensions. He utilized a wind channel whose base was divided longitudinally into two equal parts. One half was cooled while the other was heated. Titanium tetrachloride smoke served as a tracer in either the cool or the warm layers. In certain ranges of air velocity, channel depths, and rates of heating and cooling, he was able to obtain either stable wave corrugations on the interface or families of vortices with members spaced at regular intervals along the deformed discontinuity surface. Some extensions which would be simple in principle, although perhaps difficult in detail, should make it possible to study wave propagation and instability rates for measured thermal and velocity conditions. Such studies, if sufficiently successful in yielding the required type of measurements, would have very obvious utility in checking and suggesting alterations in the many existing theoretical calculations for such wave properties.

The practical problems are such that the nonrotating case would obviously have to be tackled first. But there is no reason, provided the scale of the apparatus is sufficiently small (Harwood's working section was 1 yd. long by 16 in. wide), why the whole equipment could not be put on a rotating table of the moderate size being constructed at present at the University of Chicago, or certainly in a rotating room on the scale of that at Göttingen [49]. An arrangement, in fact, could be achieved here in which it certainly ought to be possible to investigate systems of light and heavy fluid arranged side by side, similar to those of Margules [44] and Starr [61] which arose from a discussion of the energy transformations in the atmosphere. The principal obstacles preventing a close approximation to Starr's system are the infinite extent, up- and downstream, of his east-west currents, and the assumption of passage through equilibrium states at every stage. The first could probably be obviated to a large extent by arranging the initially adjacent regions of heavy and light fluid along a narrow rectangular container sufficiently long to make the end effects rather small, and the second by using very small density differences (0.001 in specific gravity is quite feasible).

Thermal Convection Studies

In the field of thermal convection two groups of investigations of meteorological interest from our present viewpoint may be recognized. The first includes a number in which rotational effects were not considered in the experimental plan, and the second includes some in which they were incorporated. The latter group will be considered in connection with the general circulation.

The first class includes work by Aitken [2] between 1870 and 1880, by Terada and Hattori [65] in 1926, the investigation by Kobayasi and Sasaki [84] in 1932, and a long series of experiments described in papers by Sandström [53–55] between 1908 and 1919, V. Bjerknes [10] in 1916, Jeffreys [35] in 1925, and Godske [28] in 1936. The experiments considered in this latter series and in the convectional part of Terada's work consisted of various arrangements of heat and cold sources in rectangular containers of water. The important meteorological idea discussed in the Bjerknes-Sandström-Jeffreys controversy was that, on an atmospheric scale, a distribution of heat sources at levels higher than the corresponding cold sources would lead to a much less vigorous atmospheric circulation than the reverse case of, for example, cold sources on high plateaus versus heat sources at the ocean surface. This result was apparently verified by Sandström's experiments which showed much more vigorous and different types of circulations in the latter case than in the former. The theoretical arguments on this point depend on whether one adopts the assumptions of Bjerknes or those of Jeffreys or takes some intermediate position. Experimentally, Sandström obtained velocities which eventually became hardly perceptible when the heat source was above the cold source. However, unpublished work done by J. G. Phillips at the University of Chicago from 1942 to 1944 has shown definite back-and-forth currents in the zone between the hot and cold sources at quite low total temperature differences when the hot source is higher. The experimental conclusion depends on what is taken as the proper similarity criterion in this case (*e.g.*, on the relative importance of various types of horizontal and vertical heat transfer) and on whether Phillips' velocities correspond to small or great velocities in the atmosphere. Offhand one would expect that the relative vertical positions of the predominantly side-by-side thermodynamic sources affecting the atmosphere would not produce great effects, because of

the large horizontal extents involved and the importance of vertical turbulence. The experimental results here need to be extended so that some numerical similarity comparisons can be made.

Experiments on the General Circulation

Historically, the principal investigations which have proceeded with the general circulation problem in mind are those of Vettin [67–69] between 1857 and 1884, Exner [17] in 1923, and Ahlborn [1] in 1924 (see also the work of Rossby [90, 50] in 1926 and 1928). Vettin performed an extremely interesting series of experiments on convectional flows in rotating and nonrotating systems. The most important from our present standpoint utilized a rotating disc of air. In one case he placed ice at the center of the disc (Fig. 5) and obtained relative

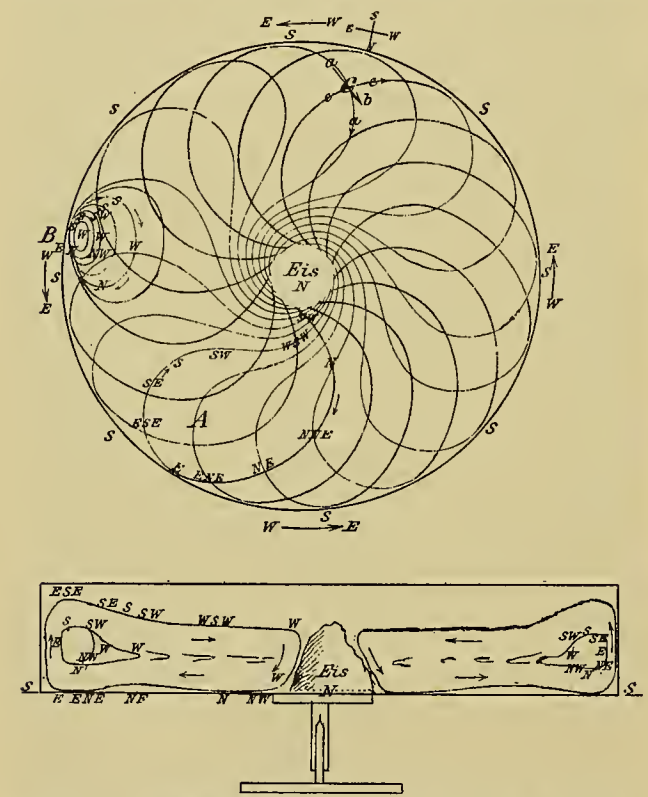


Fig. 5.—Idealized average relative circulation of air in a slowly rotating disc with a cold source at the center. Diameter 1 ft, height 2 in. approximately. The sharp disagreement with results in a disc of water of about the same height-diameter ratio (see Fig. 11) shows that extreme care is necessary in interpreting the mechanisms responsible for such motions. (*After Vettin* [69].)

circulations corresponding with what has become the traditional picture of the trade-wind cell. With the same apparatus, using air again as the medium, he provided a small gas flame arranged so as to heat the same point on the rotating base plate. This generated well-defined vortices within the air, accompanied by average relative circulations.

Vettin had the disadvantage in 1857 of writing before any clearly developed theory of dimensional or model analysis had been formulated. He attempted to draw sweeping meteorological conclusions from his work and was consequently criticised very severely by Dove [16], who was one of the prominent meteorologists of the time. Certain of the experimental phenomena which Vettin attempted to interpret meteorologically were trivially, if at all, related to the atmospheric case, but Dove's criticisms, on the other hand, were oversevere. In fact certain of the ideas which Vettin derived from his experiments, for example, those regarding the steering of storms, are more nearly in accord with the facts than the objections which Dove raised against them.[3]

Exner's experiments [17] nearly three-quarters of a century later were almost identical in one phase with Vettin's except that he used water. (However, Exner included the air vortex study mentioned earlier.) He placed a thin cylinder of ice, dyed with an ink, at the center of the rotating pan of water. The cooled portions of the water were consequently colored and could be seen spreading out along the bottom in tongues, as seen in Fig. 6. Systems of vortices, which Exner stated

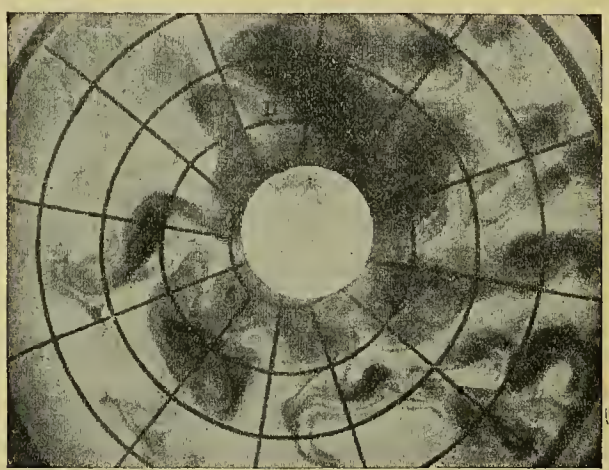


Fig. 6.—Photograph from above of Exner's experiment with tongues of ice water (colored) spreading along the bottom from a center cylinder of ice and developing vortices along their edges. Exner states the circulation in the cold water to be easterly on the average. (Rotation counterclockwise, periods 3–7 sec; diameter of pan 1 m, height of water 4–6 cm; gas flames at edge; camera rotating with pan.) (*After Exner* [17].)

to be not unlike a polar-front cyclone family in appearance, developed along their boundaries. The mean rela-

3. The whole affair ended in a rather curious and interesting way. Vettin did not publish anything further for 27 years—until persuaded to do so, apparently by friends. In a note to the article which Vettin finally wrote in 1884 [69], in the first volume of the *Meteorologische Zeitschrift*, W. Köppen says the following, after mentioning that Dove's "Law of Storms" also appeared in 1857: "Dove, der sich durch äusserst umfassende Literaturstudien und mit Hülfe einer lebhaften Phantasie, aber mit wenig eigenen Beobachtungen, ein scheinbar in sich abgeschlossenes Bild um den grossen Bewegungen der Atmosphäre entworfen hatte, wandte sich gegen den kühnen Eindringling in das von ihm beherrschte Gebiet . . . mit einer Abweisung von schwer verständlicher Heftigkeit . . . Wir dürfen hoffen das die Zeit . . . günstiger . . . ist, und das die Einführung des Experiments in die Meteorologie nicht mehr als Kinderspiel sondern als ein Gegenstand von höchster Bedeutung anerkannt werden wird . . . ."

tive motions were not observed, but should be the same as those of Fig. 5 if Vettin's observations were accurate on this point (see discussion of Fig. 11 below).

Ahlborn's experiment [1] in 1924 consisted principally of rotating a small sphere, 10 cm in diameter, in a cubic container of water. The container was 50 cm on a side. The angular velocity used was about 140 rpm, which is a relatively high speed. The motions observed in the water resembled two ring vortices centered in planes at about twenty degrees of latitude on either side of the equatorial plane. Again the relative motions in these vortices were presumed to resemble trade-wind cells. But since the action is simply that of a centrifugal pump with frictionally induced secondary flow and depends essentially on having fluid brought to rest at the outer boundary by the box, it is quite difficult to see the relevance of this particular physical mechanism in this special form to the atmospheric problem. It is, of course, conceivable that frictional effects might, for different reasons, add up in similar ways.

At this point some work of J. Thomson [66], which was outlined in his Bakerian Lecture to the Royal Society in 1892, may be mentioned. Thomson constructed a semitheoretical picture of the general circulation of the atmosphere in 1857, one of the earliest of the modern series of such attempts. This involved a meridional picture of a trade-wind type of cell extending from equator to pole aloft with a low-level cell in mid-latitudes to give the presumed average poleward drift in the westerlies at the surface. His interpretation of this cell as a secondary flow resulting from friction was somewhat similar to the ideas of Ahlborn but envisaged a more reasonable mechanism. The poleward drift in mid-latitudes he considered to be due to frictional reduction of the speed of the low-level westerlies below that corresponding to balance with the pressure field imposed from above by the very rapid westerlies of the antitrades. This effect he illustrated at the lecture by vigorously stirring the water in a round pan and then observing the inward (poleward) transport at the bottom during the decay of the motion by means of particles slightly heavier than the water. Evidently he was unaware of the work of Vettin, but he suggested a systematic series of experiments with a rotating cylindrical pan, a heat source at the bottom along the outer rim, and a cold source near the center as one means of demonstrating his theory. (This is the experiment illustrated in Fig. 11 except for the lack of a strong cold source at the center.) So far as I have been able to discover, no one acted on Thomson's suggestion until Exner began his work. Apparently Exner was unaware of Thomson's ideas on the subject.

In recent years, results have been obtained at the University of Chicago which tend to indicate that more intensive work of this type is worth carrying out. This work was begun in 1946 by the writer at the suggestion of Professors Rossby and Starr. Theoretical work [51] had suggested that the effects of lateral mixing in a *thin* spherical shell might be predominant in determining the gross character of the mean zonal circulations of a planetary atmosphere. Apparatus was constructed for rotating an inverted hemispherical shell of liquid of relatively small thickness (contained between two concentric glass bowls) and for causing convective interchange of the liquid between pole and equator by means of a heat source at the lower pole.

These experiments have shown conclusively that systematic relative zonal motions do occur in such a situation (Figs. 7 and 8). Moreover, for the particular sets

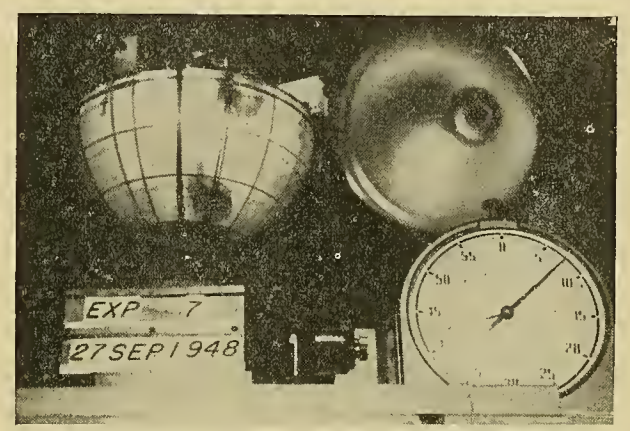


FIG. 7.—Photograph of water in the hemispherical shell apparatus taken shortly after injection of two ink clouds, at approximately latitudes 5° and 50°, from a long hypodermic needle and ink tank fixed to the glass shell. An electric heating element is located at the pole. (Rate of rotation is 8.18 rpm, heating at 20 v, mean radius of shell 10 cm, thickness of space containing water 1.6 cm, rotation is from right to left as seen.) (*After Fultz* [22].)

of conditions so far investigated, if these relative motions are represented nondimensionally as ratios of relative speed u to the absolute speed C_E of a point

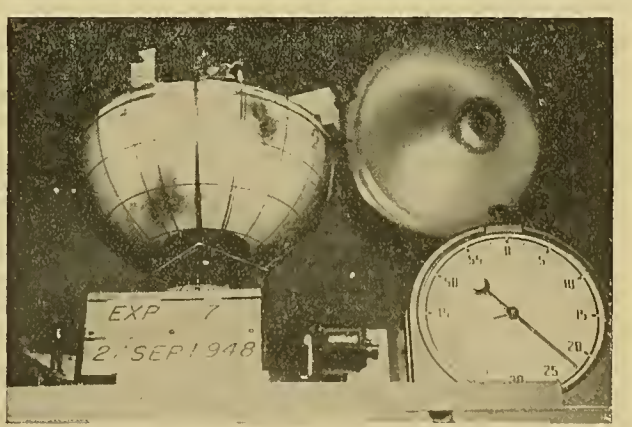


FIG. 8.—Photograph taken 14.6 sec after Fig. 7. For 10° latitude u/C_E is close to +0.016, while for 50° latitude it is close to −0.036 (*cf.* Fig. 7). (*After Fultz* [22].)

fixed on the equator of the bowls, they have values ranging up to 0.15.[4] This is just the order of magnitude of such ratios in the earth's atmosphere where C_E is about 1000 mph. Also, as shown in Fig. 9, even the

4. This ratio $u/C_E = u/r\Omega$ is apparently an even more important similarity parameter than is indicated by its kinematic significance as a velocity ratio. In the form $\Omega u/r\Omega^2$ it is a ratio of a characteristic Coriolis force to a characteristic centrifugal force (except for the unimportant numerical factor). In the form $(u^2/r)(1/\Omega u)$ it is a ratio of a characteristic rela-

latitudinal distribution of the zonal motion, under some heating conditions, shows considerable resemblance to atmospheric zonal wind profiles at moderately high levels. This and other resemblances mentioned in the original paper [22] suggest that a study of the mechanics of these systems in as great quantitative detail as possible will have real significance in suggesting working principles which would be applicable to the earth's atmosphere. The opportunity will then exist to establish areas in the conditions of the experiment that give the closest resemblances to planetary cases. More or less data for comparison exist at least for atmospheric layers on the earth, the sun, Mars, and Jupiter. With the controllability conferred by the experiment, one can then compare these areas with conditions which do not give similarity and have some hope of separating out the way in which the essential factors are operating. For example, it would be of the greatest interest, if it should turn out to be possible to generate such motions, to compare conditions required for producing motions divided sharply into zones, like those of Jupiter, with the conditions mentioned above that give motions resembling to some extent those of the earth's atmosphere.

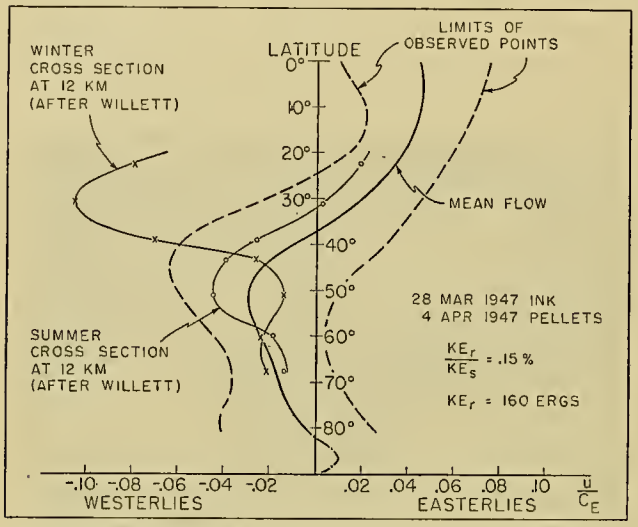


FIG. 9.—Mean estimated curve of u/C_E in the hemispherical shell of Figs. 7 and 8 for 16.27 rpm and 30 v indicated heating compared with the mean zonal winds for winter and summer at 12 km along a cross section (from Willett; see [22]) running from Havana to Aklavik, N. W. T. (*After Fultz* [22].)

In many respects the possibility of making measurements in an experiment is restricted by purely instrumental or observational difficulties. Any practicable experimental size is so small, for instance, that measurements of any velocity field to a density comparable with meteorological synoptic measurements is possible only by indirect means. In this case the utilization of the double refraction property of certain viscous fluids in deformation flow is one of the most promising possibilities [22]. Other devices which we have found to be of such great utility as to allow measurements to be made which would otherwise be almost impossible to secure include such an instrument as the "rotoscope" shown in Fig. 10. This instrument is modeled on one used by D. Thoma [19, 23] for observing the relative flow in centrifugal pumps. Provided observation is carried out on a line of sight coinciding with the axis of rotation of an object, the image of the object can be reduced to apparent rest by rotating the Dove reversing prism in the rotoscope at the proper rate. Effectively, therefore, observations can be carried out in a relative coordinate system rotating at any desired rate. The advantages of such a system are difficult to appreciate without one's having had experience with attempts to make reasonably precise measurements on a rotating object, but these advantages have been demonstrated again and again in some of our more recent investigations. These experiments have concerned a mechanically driven vortex in a two-fluid layer occupying the hemispherical shell in a slightly modified version of the original apparatus. A large amount of data on wave velocities and various types of instability points in this system has been collected very rapidly by this means. Such data could hardly have been obtained so well in any other way.[5]

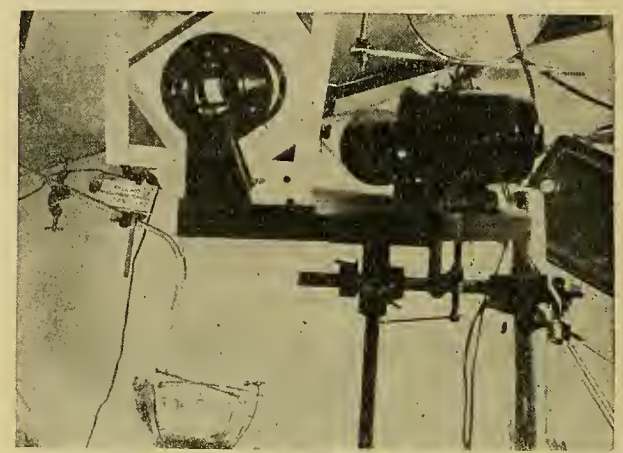


FIG. 10.—General view of rotoscope setup for observing the rotating hemispherical shell apparatus. The hemisphere is at the bottom and is viewed in a first-surface mirror (partly concealed by the rotoscope) so as to obtain a convenient direction for the line of sight. Part of the image of the hemisphere can be seen in the Dove reversing prism which is rotated in its barrel by the variable speed motor at right.

Suggestions for Future Research

The impression one receives in reviewing this long series of experimental work is that to make it really

tive inertia force to a characteristic Coriolis force. In addition, as pointed out to me by R. R. Long, it is a factor in a ratio of characteristic fluid velocity to the velocity of Rossby long waves relative to the basic current. The remaining factor in this ratio depends on geometrical quantities only. It would thus appear that this parameter measures the approach of a given system of currents to several properties peculiar to large-scale meteorological motions. For example, as $u^2/r(\Omega u)$, it indicates the degree of approximation to quasi-geostrophic motions.

5. The problem in this investigation was suggested by Rossby's pole-flight result for anticyclones [52]. At the present time it seems almost certain that an exactly analogous phenomenon has been demonstrated in the experiments. The results will be presented at length elsewhere [23, 42].

fruitful there must be a *systematic and sustained effort, guided by theory, to obtain numerical measurements of as great a variety as is practically possible.* The rather sterile, though often remarkably interesting, results of the large majority of these investigations can probably be traced to the qualitative character of the observations more than to any other factor. Another basic difficulty with applications of these experiments to the atmosphere has been the lack of simple, theoretical, guiding principles which take into account the thinness of the atmosphere. Experiments need to be designed around principles, even if only semi-quantitatively. In the experiments at the University of Chicago, for example, various applications of the vorticity theorem due to Rossby (*e.g.*, [51]) have provided the theoretical suggestions for much of the experimental work. It is not going to be easy, nor even possible in some cases, to measure adequately the field quantities which may appear to be most immediately required in comparing possible theoretical explanations. Nevertheless, certain types of measurements are quite possible. As experimental work progresses, techniques and altered concepts of the quantities to be measured should develop to a far higher point than is the case at present.

In connection with general circulation problems, the following experimental steps appear to me to be the next ones to take.

1. The simple types of techniques used in hydraulic or aerodynamic experiments (or in the studies cited above) to measure velocity and temperature should be extended to the cylindrical discs studied by Vettin, Thomson, and Exner. Several modifications would be necessary and height-diameter ratios should be investigated to as low values as are possible.[6] Besides the general possibility of investigating the motions near a pole for thin shells, there are indications that evidence could be obtained bearing on a comment by Jeffreys [36] that a reversed distribution of thermodynamic sources might reverse the sense of the average relative zonal motions of the atmosphere as functions of latitude (see also below). Preliminary observations on a rotating dishpan apparatus constructed by F. Hall for a similar type of study indicated that with a Bunsen burner under the center (instead of a cold source there) the more rapid outward radial currents show relative westerlies instead of easterlies as should be the case to correspond with Vettin's result. The particular equipment is subject to vibration so that some reservations must be entered, but very recent (March, 1950) rotoscope observations and photographs with the dishpan (radius 7 in.) filled with water to a depth of from $2\frac{1}{2}$ to 3 in. and heated at the outer rim show some extremely striking patterns. Figure 11 shows an aluminum-powder

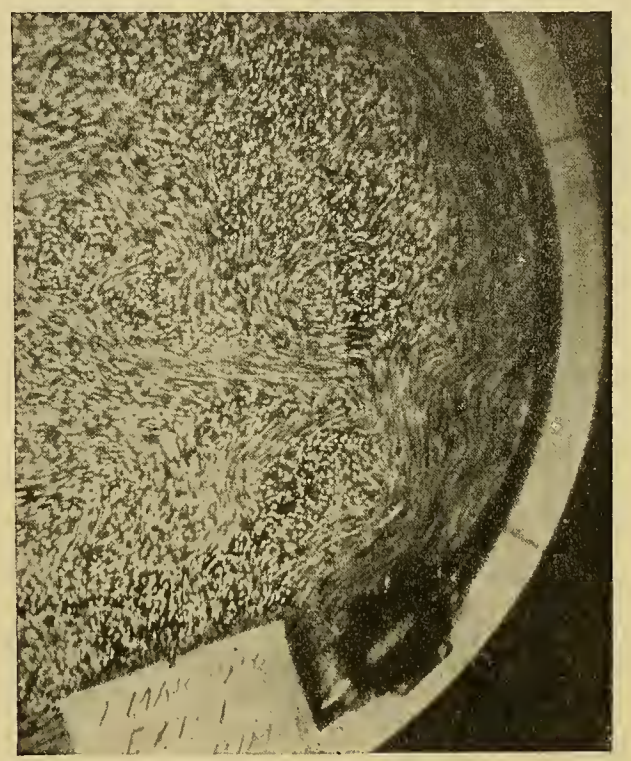


FIG. 11.—Aluminum-powder streak photograph of the relative circulation at the top surface of water in a pan rotating clockwise (12 rpm), which is being heated at the bottom near the rim. The narrow current near the rim is westerly (*i.e.*, moving more rapidly than the pan) while that near the middle is predominantly easterly. The second current radially in from the counter, for example, is almost certainly easterly, on the basis of visual observations. (Photograph 8 min after heating began, exposure $\frac{1}{2}$ sec, westerly current $u/C_E = -0.05$; at 12 min after heating began, temperature at top center was 30.7C, at bottom rim 31.9C taken with a mercury thermometer.)

streak photograph taken through the rotoscope with an exposure of $\frac{1}{2}$ sec so as to show the motions relative to the pan. The narrow band of motion near the rim (moving clockwise) is a *westerly* current with u/C_E values running to about -0.06 or more, when the angular velocity of the pan is 12 rpm. The rest of the top surface is occupied by a number of eddies of various sizes which, for example, give average values of $+0.02$ for u/C_E at a distance of $0.6r_0$ from the center (r_0 is the rim radius). In this condition the water temperature away from the pan surface is hotter at the rim than at the center by 1–2C. The burner was also moved in to $0.5r_0$. The average temperature gradient in the water is then reversed with the temperature at the pole perhaps $\frac{1}{2}$C warmer than at the rim because of the rapid heat conduction in the metal of the pan. The relative

6. It has recently come to my attention that in 1902 C. A. Bjerknes carried out experiments with a rotating cylinder of water (12 cm high, 36 cm wide, 7 rpm) to verify qualitatively Ekman's theory of the surface drift current in the ocean. A jet of air 10 cm broad was blown along a diameter relative to the rotating cylinder. Measurements of the currents with floating balls and a sensitive directional vane showed the proper deflection to the right (Northern Hemisphere rotation) at the surface and increasing rightward deflection in the top centimeter where the Ekman spiral would be expected. This suggests the possibility of studying wind-driven currents in oceanic basins of suitable shapes on an experimental basis, so long as variations of the Coriolis parameter are not important, and of utilizing paraboloidal shapes where such variation is important. (See Ekman, V. W., "On the Influence of the Earth's Rotation on Ocean Currents." *Ark. Mat. Astr. Fys.*, Vol. 2, No. 11, pp. 51–52 (1905) and also [85] and [86].)

velocity distribution, on the other hand, is not strongly altered. The westerly current at the rim is weaker and narrower and the easterly current, at say $0.7r_0$, is stronger and better defined with u/C_E as much as $+0.04$ or more. In this case, narrow, strong, easterly currents are set up *inward* toward the center as observed earlier by F. Hall. One of the important differences between this and Vettin's experiment is the high conductivity of the pan (his base plate was of glass). This leads to static instability over the entire bottom that is of course relatively stronger near the burner position. Thus, ink observations suggest that systematic vertical motions, if present at all, are very vaguely defined and that the rising motions are rather irregularly distributed over the pan from a Bénard-like hot layer at the very bottom. At any rate, considerable additional doubt (see also below) is thrown on any simple reasoning of the usual textbook variety concerning the effects of simple heat-source distributions and the conservation of angular momentum in the general circulation. Such reasoning certainly ought to apply to a simple disc if anywhere, but it is very hard to conceive of an axially symmetrical circulation which would account for both the westerlies and the easterlies at the top surface in this experiment without introducing various epicycles that have no obvious relation to the heat sources.

Figure 12 illustrates another striking effect which was accidentally observed in the rotating pan. This effect is a result of the rotational property stated by Taylor [63, 64] (see also a more recent discussion by Görtler [29]) that slow relative motions in a fluid initially rotating as a solid should tend to be two-dimensional in planes perpendicular to the axis of rotation. In Fig. 12, ink is seen a minute or so after being poured into the rotating pan in a quite arbitrary manner. The mass of ink, instead of spreading out in the usual eddying fashion, rapidly forms vertical walls and the cylindrical surfaces gradually wind around one another, retaining their parallelism to the axis.[7] They diffuse much more slowly than in a nonrotating pan. Such an effect of rotation would obviously be of the greatest importance in diffusion or convection problems in establishing preferred directions with radically different coefficients of transfer. In the apparatus the effect is still strongly present when the rotational instability parameter $g/a\Omega^2$ is as high as 140 (6 rpm) and is even evident to a lesser extent when heating is going on with a metal bottom (*e.g.*, whole vertical streamers of ink transfer laterally without tilting much). The rotational *stability* is obviously much higher and the hydrostatic stability probably much less in this case than for the earth. However, certain other preliminary experiments with the same dishpan apparatus show that this effect has some very important implications from a meteorological standpoint. The bottom of the rotating dishpan was insulated with a layer of plaster and heating carried out as before at the rim [25]. The top surface then shows a wide meandering band of strong westerlies (u/C_E of order 0.1) with rather larger vortices than before. Toward the small central area the average motion is uncertain but is at least very weak westerly or perhaps easterly. If the Taylor ink columns are produced before heating is started and then allowed to remain, they *retain* their cylindrical character while moving with the thermally induced currents. In other words, the motions appear to remain quasi-barotropic and are certainly in addition quasi-geostrophic. On the other hand, the temperature field shows strong static stability and is strongly baroclinic with a dome of cold water occupying the center bottom. The rim-center temperature differences run to the order of 5C. Tentatively then, one can conclude that it is quite possible to have a strongly baroclinic and reasonable pole-equator temperature distribution in this experiment and yet also have quasi-barotropic currents and disturbances.[8]

FIG. 12.—Ink initially of slightly greater density, a couple of minutes after being poured into a pan of water in solid rotation at 15 rpm. Note the vertical walls. As in Taylor [63] the walls remain vertical for many minutes and gradually draw out into surfaces which are sharply divided by clear areas when viewed end-on. The same effect is very marked to at least as low as 5 rpm and occurs under the heating conditions of Fig. 11 when the pan bottom is insulated by a layer of plaster.

7. One is reminded of the Margules-Helmholtz slope condition for a discontinuity surface of wind which reduces to parallelism to the earth's axis when the density difference is zero.

8. There are a number of observational difficulties concerning the representativeness of the top surface motions and other matters that make it necessary to regard any detailed

2. The work of Fultz [22] should be extended to spherical shells incorporating variable depth-radius ratios which Rossby's theory would imply to be important. The data from these experiments need to include much more complete temperature values and also more extensive velocity observations. By use of the rotoscope and appropriate mirror or lens systems it will be possible to obtain velocity values much more easily and accurately. With enough data to be statistically significant, it will be possible to calculate such properties of the motion as the horizontal Reynolds' stresses (momentum transfers). The data taken in 1947 have been found insufficient to give reliable estimates of such secondary quantities.

3. A most important companion set of experiments to these spherical shell studies needs to be carried out in paraboloidal shells rotated at rates such that the free surface figure coincides with that of the paraboloid.[9] All the previous types of work should be carried out here. An even better check on the Jeffreys comment above would be obtained by comparing a cold-pole–warm-rim setup with a warm-pole–cold-rim. With a moderately deep layer it should be possible to compare the efficacy of heat sources at high altitudes and cold sources at low altitudes with the opposite case in a way that would have relatively conclusive application to the controversy mentioned earlier.

There is one caution here that a rough calculation brings up. If we consider the thermal expansion of water (or most other practicable liquids) and ordinarily available temperature gradients, it appears that the thermally developed horizontal pressure-gradient forces are likely to be of a very small order relative to gravity. If these forces are not to be masked by inaccuracies of figure (which would give rise to components of the external action along the surface), the paraboloid may need to have a slope at each radius accurate to 10^{-3} parts or better. Corresponding requirements would subsist for the constancy of angular velocity of the basic rotation. Such accuracy will be very difficult to obtain practically. There may be means of handling the situation without proceeding to this degree of precision, but some experience and trial and error will be needed to work them out.

4. Shearing waves and instability at density discontinuity surfaces in general deserve extensive investigation, in both the rotating and nonrotating cases with various geometrical forms (see, for example, Keulegan [37]). Here the problem of obtaining velocity measurements of the required density and accuracy for meteorological application is very serious. But, at least in the case of liquids, modifications of the standard aerodynamic aluminum-powder techniques using sparser distributions (or possibly liquid bubble tracers) with rapid-sequence flash photography should be valuable.

In all of the cases mentioned above there are many obvious extensions such as (*a*) varying the contour and roughness of the underlying surfaces, for example. mountain barriers or other forced depth changes, (*b*) studying the motions of thermally induced currents around obstacles of various shapes or moving the obstacles themselves through the fluid with and without concomitant convective motions (see [26] and [24]), and (*c*) studying the effect of a varying horizontal conductivity of the underlying surface on convective motions, for example, ocean versus land differences of this kind.

Concerning smaller-scale problems it appears that important work could be done on vortices (tornadoes and hurricanes) by careful measurements of field quantities around one or the other of the experimental cases. Probably the simplest would be one of the steady mechanically driven air vortices such as Weyher's. Velocities could be obtained by a series of probes with a microsized Pitot tube or hot-wire anemometer. The vexing problem of just how much weight to attach to all the reasoning concerning "*vr*" vortices and momentum conservation in these circumstances should receive some elucidation. Projected experiments of other types have been referred to earlier.

There exist at least two fields in which the possibilities of significant experimentation seem very large and which, from the meteorological angle, are just opening up at the present time. One of these is the field of hydraulic and supersonic gas flow analogies, opened for meteorology by Freeman [20]. Current work by Rossby indicates that there may exist even larger-scale analogies to the hydraulic jump than those considered by Freeman and later workers. Quite aside from the desirability of modifying the extensive experiments which have been carried out in this field so as to lay special stress on the meteorologically important quantities, there is every prospect that significant amounts of rotation can be introduced experimentally. For example, a small version of the complete table described by Matthews [46] could rather easily be arranged so as to be rotatable. Similar open-channel work with subcritical flow has in fact already been done in the rotating room at Göttingen by Fette [18].

The second field is one in which hardly anything of meteorological significance has yet been done experimentally, but in which many possibilities must exist. This area is that of electromagnetic analogies. Such analogies have quite often been used in potential flow problems where a direct electrical analogy in conducting fluids is very easy to apply in practical problems for which the boundaries are complicated [39, p. 65].

An extensive discussion of another sort of identity between electromagnetic and hydrodynamic systems (with experimental illustrations) has been given by V. Bjerknes [9]. Another has been interestingly pointed out and used by Ising [34]. In this paper, advantage is taken of the fact that there is complete identity in form between the equations for an incompressible fluid in the usual meteorological relative coordinate system rotating at a constant angular velocity and those for a corresponding system of electric currents in a uniform

conclusions concerning these experiments as quite provisional. They will be discussed as comprehensively as possible in an early publication on this aspect of our work.

9. This was first brought to my attention in this connection by the group at the University of California at Los Angeles.

magnetic field. For example, there is a phenomenon of "hydraulic induction" exactly corresponding to electromagnetic induction, in which fluid currents develop in any closed circuit whose orientation changes in such a way as to change the flux of the earth's rotation vector through the circuit.[10] Ising was able to demonstrate this effect experimentally in an apparatus exactly corresponding to a single-turn electric generator.

One type of experiment suggested by the analogy would be to place a disc of fluid in a uniform magnetic field whose direction is perpendicular to the plane of the disc. The fluid would then be electrically charged and convection currents generated in it by appropriate heat sources or by other means. The Coriolis forces would be replaced by electromagnetic forces on the moving fluid, and the experiment would have the advantage over rotating disc experiments of not involving a centrifugal force. A numerical check on this experiment was carried out by S. L. Hess who found that the charges, electric potentials, and magnetic fields required to produce mechanically significant forces from such slow motions are so large as to make the experiment thoroughly impractical. Nevertheless, there are many other possibilities, and it is altogether likely that useful ones will eventually be developed.

REFERENCES

1. Ahlborn, Fr., "Die drei grossen Zirkulationen der Atmosphäre." *Beitr. Phys. frei. Atmos.*, 11:117–153 (1924).
2. Aitken, J., "On the Distribution of Temperature under the Ice in Frozen Lakes" (in *Collected Scientific Papers*, edited by C. G. Knott, pp. 30–33. Cambridge, University Press, 1923) *Proc. roy. Soc. Edinb.*, 10:409–415 (1879–80).
3. —— "Notes on the Dynamics of Cyclones and Anticyclones," Parts I and II, in *Collected Scientific Papers*, edited by C. G. Knott, pp. 438–458 (1900–01). Part III: *ibid.*, pp. 459–467 (1915). Cambridge, University Press, 1923.
4. —— "Revolving Fluid in the Atmosphere." *Proc. roy. Soc.*, (A) 94:250–259 (1917–18).
5. Andries, P., "Die Entstehung der Cyklonen." *Z. öst. Ges. Meteor.*, 17:307–317 (1882).
6. Belopolsky, A., "Einige Gedanken über die Bewegungen auf der Sonnenoberfläche." *Astr. Nachr.*, 114:153–154 (1886).
7. —— "Über die Bewegungen auf der Sonnenoberfläche." *Astr. Nachr.*, 124:17–22 (1890).
8. Bezold, W. von, "Experimentaluntersuchungen über rotirende Flüssigkeiten." *Ann. Phys., Lpz.*, (3) 32:171–187 (1887).
9. Bjerknes, V., *Die Kraftfelder.* (*Die Wissenschaft*, Bd. 28.) Braunschweig, F. Vieweg & Sohn, 1909.
10. —— "Über thermodynamische Maschinen die unter Mitwirkung der Schwerkraft arbeiten." *Abh. sächs. Ges. (Akad.) Wiss.*, Bd. 35, Nr. 1, 33 SS. (1916).
11. Bouasse, H., *Tourbillons, forces acoustiques, circulations diverses*, Tome I. Paris, Delagrave, 1931.
12. Czermak, P., "Ueber warme Luft- und Flüssigkeitsströmungen." *Ann. Phys., Lpz.*, (3) 50:329–334 (1893).
13. Dechevrens, P. M., "Sur la reproduction expérimentale des trombes." *C. R. Acad. Sci., Paris*, 105:1286–1289 (1887).
14. Defant, A., "Theoretische Überlegungen und experimentelle Untersuchungen zum Aufbau hoher Zyklonen und Antizyklonen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 132: 81–103 (1923).
15. Dines, W. H., "Experiments Illustrating the Formation of the Tornado Cloud." *Quart. J. R. meteor. Soc.*, 22:71–73 (1896).
16. Dove, H. W., "Einige Bemerkungen über die meteorologischen Aufsätze des Hrn. Vettin." *Ann. Phys., Lpz.*, (2) 102:607–613 (1857).
17. Exner, F. M., "Über die Bildung von Windhosen und Zyklonen." *S. B. Akad. Wiss. Wien*, Abt. IIa, 132:1–16 (1923).
18. Fette, H., "Strömungsversuche im rotierenden Laboratorium." *Z. tech. Phys.*, 14:257–266 (1933).
19. Fischer, K., "Untersuchung der Strömung in einer Zentrifugalpumpe." *Mitt. hydraul. Inst. München*, 4:7–27 (1931).
20. Freeman, J. C., Jr., "An Analogy between the Equatorial Easterlies and Supersonic Gas Flows." *J. Meteor.*, 5:138–146 (1948).
21. Fujiwhara, S., "On the Growth and Decay of Vortical Systems." *Quart. J. R. meteor. Soc.*, 49:75–104 (1923).
22. Fultz, D., "A Preliminary Report on Experiments with Thermally Produced Lateral Mixing in a Rotating Hemispherical Shell of Liquid." *J. Meteor.*, 6:17–33 (1949).
23. —— "Experimental Studies of a Polar Vortex I." *Tellus*, 2:137–149 (1950).
24. —— "Experimental Studies Related to Atmospheric Flow around Obstacles." *Geofys. pura e applicata*, 17:89–93 (1950).
25. —— "Experiments Combining Convection and Rotation and Some of Their Possible Implications." *Proc. Midwest Conf. Flu. Dynam.* (May, 1950).
26. —— and Long, R. R., "Two-Dimensional Flow around a Circular Barrier in a Rotating Spherical Shell." *Tellus*, (1951) (in press).
27. Ghatage, V. M., *Modellversuche über die gegenseitige Bewegung von Luftmassen verschiedenen Temperaturen.* Inaug. Diss., Göttingen, 35 SS., 1936.
28. Godske, C. L., "Note on an Apparent Discrepancy between Bjerknes-Sandström and H. Jeffreys." *Quart. J. R. meteor. Soc.*, 62:446–450 (1936).
29. Görtler, H., "Einige Bemerkungen über Strömungen in rotierenden Flüssigkeiten." *Z. angew. Math. Mech.*, 24: 210–214 (1944).
30. Hale, G. E., and Luckey, G. P., "Some Vortex Experiments Bearing on the Nature of Sunspots and Flocculi." *Proc. nat. Acad. Sci., Wash.*, 1:385–389 (1915).
31. Harwood, W. A., "Experiments on Frontal Waves." *Quart. J. R. meteor. Soc.*, 71:377–383 (1945).
32. Haynes, B. C., "A Density Channel for Illustrating Fronts and Occlusions." *Bull. Amer. meteor. Soc.*, 20:37–38 (1939).
33. Hildebrandsson, H. H., "Wilcke's Experiment zur Darstellung der Wettersäulen." *Z. öst. Ges. Metcor.*, 11:221–222 (1876).
34. Ising, G., "Die physikalische Möglichkeit eines tierischen Orientierungssinnes auf Basis der Erdrotation." *Ark. Mat. Astr. Fys.*, Vol. 32A, No. 18, pp. 1–23 (1945).
35. Jeffreys, H., "On Fluid Motions Produced by Differences of Temperature and Humidity." *Quart. J. R. meteor. Soc.*, 51:347–356 (1925).
36. —— "The Function of Cyclones in the General Circula-

10. In more usual meteorological terms, this is a simple consequence of the Bjerknes circulation theorem, and the experiment may be regarded as a demonstration of one aspect of the theorem, or, alternatively, of the rotation of the earth.

tion." *P. V. Météor. Un. géod. géophys. int.*, Lisbonne, 1933. II, pp. 219–230 (1935).

37. Keulegan, G. H., "Interfacial Instability and Mixing in Stratified Flows." *J. Res. nat. Bur. Stand.*, 43:487–500 (1949).
38. Koschmieder, H., "Über das Zerfliessen einer Kaltluftmasse." *Meteor. Z.*, 59:303–306 (1942).
39. Lamb, H., *Hydrodynamics*, 6th ed. Cambridge, University Press, 1932.
40. Letzmann, J., "Experimentelle Untersuchungen an Wasserwirbeln." *Beitr. Geophys.*, 17:40–85 (1927).
41. —— "Experimentelle Untersuchungen an Luftwirbeln." *Beitr. Geophys.*, (Köppen-Band II) 33:130–172 (1931).
42. Long, R. R., "A Theoretical and Experimental Study of the Motion and Stability of Certain Atmospheric Vortices." *J. Meteor.*, 8:207–221 (1951).
43. Lunelund, H., "Über Rankines kombinierte Wirbel." *Acta Soc. Sci. fenn., Phys. Math.*, Bd. 2, Nr. 24, 21 SS. (1921).
44. Margules, M., "On the Energy of Storms" in Abbé, C., "The Mechanics of the Earth's Atmosphere." *Smithson. misc. Coll.*, 51:533–595 (1903).
45. Mascart, M., "Expériences de M. Weyher sur les tourbillons, trombes, tempêtes et sphères tournantes." *J. Phys.*, (2) 8:557–572 (1889).
46. Matthews, C. W., "The Design, Operation, and Uses of the Water Channel as an Instrument for the Investigation of Compressible-Flow Phenomena." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 2008 (1950).
47. Oberbeck, A., "Ueber discontinuirliche Flüssigkeitsbewegungen." *Ann. Phys., Lpz.*, (3) 2:1–16 (1877).
48. —— "Strömungen von Flüssigkeiten infolge ungleicher Temperatur innerhalb derselben." *Ann. Phys., Lpz.*, (3) 11:489–495 (1880).
49. Prandtl, L., "Erste Erfahrungen mit dem rotierenden Laboratorium." *Naturwissenschaften*, 14:425–427 (1926).
50. Rossby, C.-G., "Studies in the Dynamics of the Stratosphere." *Beitr. Phys. frei. Atmos.*, 14:240–265 (1928). (See p. 261)
51. —— "On the Distribution of Angular Velocity in Gaseous Envelopes under the Influence of Large-Scale Horizontal Mixing Processes." *Bull. Amer. meteor. Soc.*, 28:53–68 (1947).
52. —— "On a Mechanism for the Release of Potential Energy in the Atmosphere." *J. Meteor.*, 6:163–180 (1949).
53. Sandström, J. W., "Dynamische Versuche mit Meerwasser." *Ann. Hydrogr., Berl.*, 36:6–23 (1908).
54. —— "Meteorologische Studien im schwedischen Hochgebirge." *Göteborg. VetenskSamh. Handl.*, (4) Vol. 17, No. 2, 48 pp. (1914).
55. —— "The Hydrodynamics of Canadian Atlantic Waters" in *Canadian Fisheries Expedition, 1914–1915*, pp. 221–343. Ottawa, Dept. of Naval Service, 1919.
56. Schmidt, W., "Gewitter und Böen, rasche Druckanstiege." *S. B. Akad. Wiss. Wien*, Abt. IIa, 119:1101–1214 (1910).
57. —— "Zur Mechanik der Böen." *Meteor. Z.*, 28:355–362 (1911).
58. —— "Modellversuche zur Wirkung der Erddrehung auf Fluszläufe." *Festschr. Zent-Anst. Meteor. Geodyn., Akad. Wiss. Wien*, SS. 187–195 (1926).
59. Sipinen, U., "Einige Versuche mit kleinen Luftwirbeln." *Ann. Acad. Sci. fenn.*, (A) Bd. 23, Nr. 1, 23 SS. (1924).
60. Spilhaus, A. F., "Note on the Flow of Streams in a Rotating System." *J. mar. Res.*, 1:29–33 (1937).
61. Starr, V. P., "The Readjustment of Certain Unstable Atmospheric Systems under Conservation of Vorticity." *Mon. Wea. Rev. Wash.*, 67:125–134 (1939).
62. Stümke, H., "Rotationssymmetrische Gleichgewichtsstörungen in einer isothermen Atmosphäre nebst einem Modellversuch mit rotierender Flüssigkeit." *Z. Geophys., Meteor., Geod.*, 16:127–149 (1940).
63. Taylor, G. I., "Experiments with Rotating Fluids." *Proc. roy. Soc.*, (A) 100:114–121 (1921).
64. —— "Experiments with Rotating Fluids." *Proc. First Int. Congr. appl. Mech.*, (Delft), C. B. Biezeno and J. M. Burgers, ed., 1924. (See pp. 89–96)
65. Terada, T., and Hattori, K., "Some Experiments on Motion of Fluids." Parts I, II, III: *Rep. aero. Res. Inst. Tokyo*, 2:87–112 (1926); Part IV: *ibid.*, pp. 287–326.
66. Thomson, James, "On the Grand Currents of Atmospheric Circulation." *Phil. Trans. roy. Soc. London*, (A) 183:653–684 (1892).
67. Vettin, F., "Meteorologische Untersuchungen." *Ann. Phys., Lpz.*, (2) 100:99–110 (1857).
68. —— "Über den aufsteigenden Luftstrom, die Entstehung des Hagels und der Wirbel-Stürme." *Ann. Phys., Lpz.*, (2) 102:246–255 (1857).
69. —— "Experimentelle Darstellung von Luftbewegungen unter dem Einfluss von Temperatur-Unterschieden und Rotations-Impulsen." *Meteor. Z.*, 1:227–230, 271–276 (1884).
70. Wegener, A., *Wind- und Wasserhosen in Europa.* (*Die Wissenschaft*, Bd. 60.) Braunschweig, F. Vieweg & Sohn, 1917.
71. Wilcke, J. C., "Forsok til Uplysning om Luft-hvirflar och Sky-drag." *K. Vet. Acad. nya Hand.*, (2) 1:1–18, 83–102 (1780); 3:3–35 (1782); 6:290–307 (1785).
72. Yih, C. S., *A Study of the Characteristics of Gravity Waves at a Liquid Interface.* S. M. Thesis, Iowa Inst. Hydraul. Res., 1947.

SUPPLEMENTARY REFERENCES

(Added in press)

73. Abbé, C., "Comprehensive Maps and Models of the Globe for Special Meteorological Studies." *Mon. Wea. Rev. Wash.*, 35:559–564 (1907).
74. —— "Projections of the Globe Appropriate for Laboratory Methods of Studying the General Circulation of the Atmosphere." *Bull. Amer. math. Soc.*, 13 (2):502–506 (1907).
75. Czermak, P., "Experimente zum Föhn." *Denkschr. Akad. Wiss. Wien*, 73:63–66 (1901).
76. Dinkelacker, O., "Darstellung einer Inversion und eines Kaltlufteinbruchs in der Höhe." *Meteor. Z.*, 54:477–479 (1937).
77. Exner, F. M., "Zur Wirkung der Erddrehung auf Flussläufe." *Gcogr. Ann., Stockh.*, 9:173–180 (1927).
78. —— "Bemerkungen zu Ahlborn 'Die drei grossen Zirkulationen der Atmosphäre.'" *Meteor. Z.*, 42:161–162 (1925).
79. Fultz, D., "Non-dimensional Equations and Modeling Criteria for the Atmosphere." *J. Meteor.*, 8:262–267 (1951).
80. Helmholtz, H. v., "Wirbelstürme und Gewitter." *Dtsch. Rdsch. Geogr.*, 6:363–380 (1876). (See pp. 376 ff.)
81. Hensen, W., "Der Einfluss der Erdumdrehung auf Tideflüsse in der Natur und im Modell." *Bautechnik*, 17:285–288 (1939).
82. Hirn, G. A., "Étude sur un classe particulière de tourbillons, etc." *Bull. Sci. d'Hist. Natur. Colmar*, 40 pp. (1878).
83. Hobbs, W. H., "The Mechanics of the Glacial Anticyclone Illustrated by Experiment." *Nature*, 105:644–645 (1920).
84. Kobayasi, T., und Sasaki, T., "Über Land- und Seewinde." *Beitr. Phys. frei. Atmos.*, 19:17–21 (1932).
85. Krümmel, O., *Handbuch der Ozeanographie*, Bd. II, 766 SS. Stuttgart, 1911. (See pp. 471–479 and *c.*)
86. Lasareff, P., "Sur une méthode permettant de démontrer

la dépendance des courants océaniques des vents alizés et sur le rôle des courants océaniques dans le changement du climat aux époques géologiques." *Beitr. Geophys.*, 21: 215–233 (1929).

87. Prandtl, L., "Beiträge zur Mechanik der Atmosphare." *P. V. Météor. Un. géod. géophys. int.*, Edimbourg, 1936. II (1939).
88. Riabouchinsky, D. P., "Sur les phénomènes hydrodynamiques intermittents." *Bull. Inst. aérodyn. Koutchino*, 5:98–110 (1914).
89. Richardson, L. F., *Weather Prediction by Numerical Process.* Cambridge, University Press, 1922. (See pp. 219–220)
90. Rossby, C.-G., "On the Solution of Problems of Atmospheric Motion by Means of Model Experiments." *Mon. Wea. Rev. Wash.*, 54:237–240 (1926).
91. Sandström, J. W., "Über die Ablenkung des Windes infolge der Erddrehung und der Reibung." *Ark. Mat. Astr. Fys.*, Bd. 9, No. 31, 8 SS. (1914).
92. Schmidt, W., "Weitere Versuche über den Böenvorgang und das Wegschaffen der Bodeninversion." *Meteor. Z.*, 30: 441–447 (1913).
93. Sprung, A., "Zur Anwendung des Princips der Flächen in der Meteorologie." *Z. öst. Ges. Meteor.*, 16:57–63 (1881).
94. Weickmann, L., "Die Gewitter vom 11 Mai 1910." *Jb. Meteor. bayer.*, Bd. 33 (1911).
95. —— und Haurwitz, B., "Mechanik und Thermodynamik der Atmosphäre," Gutenberg, B., *Lehrbuch der Geophysik*, Lief. 5, SS. 797–961. Berlin, Gebr. Borntträger, 1929 (See pp. 917–935)

MODEL TECHNIQUES IN METEOROLOGICAL RESEARCH

By HUNTER ROUSE

Iowa Institute of Hydraulic Research, State University of Iowa

Introduction

Scale-model investigations of fluid motion have long since proved their worth in such widely varied fields as aeronautics, ballistics, hydraulics, and naval architecture. New aircraft, projectiles, turbines, spillways, and ships are almost invariably tested at reduced scale before final acceptance of design, and prototype behavior is being predicted with ever-increasing accuracy. Particularly in the field of flood control and river regulation, economies of construction or improvements in design which model studies generally reveal have offset the costs of the studies many times over. It is therefore hardly surprising that meteorologists frequently wonder whether model techniques would not be useful in meteorological research as well.

Wind-tunnel studies of meteorological phenomena have, in fact, already been described in the literature on at least a dozen occasions, a card file of such papers being maintained by the U. S. Weather Bureau. During World War II, moreover, a considerable number of model investigations were made of particular atmospheric occurrences, although many of the results have not yet been published. Each project of this nature was undertaken for the specific reasons behind all laboratory research at small scale: on the one hand, the essential variables can be controlled at will; on the other hand, small-scale experimentation generally reduces to an extreme degree the time and expense involved. It is true, of course, that few large-scale occurrences can be reproduced to perfection in a laboratory model, while some—such as those of the atmosphere as a whole—are so complex that their duplication in miniature would be utterly impossible. Nevertheless, many flow phenomena become subject to analysis only when reduced to their bare essentials, and hence even a very rough approximation of actual conditions is often fully justified.

If model studies are to provide useful prototype indications, the following steps are generally necessary in addition to the actual tests. First of all, a preliminary analysis of the problem should be made, based upon dimensional considerations and such physical principles as are readily applicable, so that the subsequent experiments may be efficiently conducted. Secondly, the experiments should be planned to make optimum use of available equipment and instruments, or suitable apparatus must be devised. Finally, the results of the experiments should be reduced to their most general form for utilization at prototype scale. In the following pages each of these aspects of model technique is discussed in detail with particular application to meteorological research, and typical problems already investigated by such means are described.

Similitude Requirements

A very powerful tool of laboratory research is the procedure known as dimensional analysis [3]. Frequently maligned by some because of its limitations and frequently misused by others who do not recognize its limitations, the method is actually of great worth as a systematic guide and check in planning new experimental studies. Based upon the physical requirement of dimensional homogeneity in any functional relationship, such analysis permits the variables involved to be reduced in number, generalized in form, and brought into an arrangement suitable for experimental investigation.

The Π-theorem [4] of dimensional analysis states that if the n variables involved in a function require m dimensional categories for their expression, then the function can be rewritten in terms of $n - m$ dimensionless groups of these variables. Moreover, this theorem provides a systematic algebraic procedure whereby any series of variables involved in a phenomenon, regardless of their number or the number of the dimensional categories, may be reduced to a significant dimensionless form. Here it is sufficient to say that any problem of geometry alone will thus be expressible in terms of ratios of lengths, and that model and prototype boundaries and flow patterns will be geometrically similar if all corresponding length ratios have the same numerical value in both cases. Likewise, a problem of kinematics will be expressible in terms of ratios of lengths and times —or combinations thereof, such as velocities—and hence model and prototype occurrences will also be kinematically similar if all corresponding kinematic ratios are identical. Finally (at least so far as phenomena of mechanics are concerned), a problem of dynamics will be expressible in terms of ratios of quantities involving length, time, and either force or mass; accordingly, model and prototype occurrences will be dynamically similar as well if all corresponding dynamic ratios are the same in both.

The various quantities required to describe any phenomenon of fluid mechanics include a series of lengths a, b, c, . . . descriptive of the boundary geometry; a series of flow characteristics such as velocity V and differential pressure Δp; and a series of fluid properties —the density ρ, the differential specific weight $\Delta\gamma$, the dynamic viscosity μ, and the bulk modulus of elasticity e. Although many different dimensionless arrangements of these variables may be obtained by the Π-theorem, the following is most significant for conditions of similarity [14]:

$$\frac{V}{\sqrt{2\Delta p/\rho}} = \varphi\left(\frac{b}{a}, \frac{c}{a}, \cdots \frac{V}{\sqrt{a\Delta\gamma/\rho}}, \frac{Va}{\mu/\rho}, \frac{V}{\sqrt{e/\rho}}\right).$$

The first term is a dependent parameter sometimes known as the Euler number $\mathbf{E} = V/\sqrt{2\Delta p/\rho}$, consisting primarily of the flow characteristics. The next two (or more) are length ratios. The last three are groups involving the fluid properties, known respectively as the Froude number $\mathbf{F} = V/\sqrt{a\Delta\gamma/\rho}$ (gravity), the Reynolds number $\mathbf{R} = \rho Va/\mu$ (viscosity), and the Mach number $\mathbf{M} = V/\sqrt{e/\rho}$ (compressibility). Evidently, even though the actual form of the function is unknown, the dependent parameter will be the same at model and prototype scale if the corresponding independent parameters have the same numerical values—in other words, the model and the prototype flows will then (but only then) be mechanically similar.

Since all problems in mechanics are expressible in terms of three-dimensional categories, the Π-theorem will be seen to reduce by three the number of terms in the function, each of the dimensionless parameters differing from the others by one variable only. It is thus possible in such problems to distribute the various fluid properties singly (except for the density) among the three independent parameters **F**, **R**, and **M**. A problem which combines mechanics and thermodynamics, on the contrary, requires one more dimensional category—temperature—for its expression, as well as such additional flow-variables as temperature rise and heat flux, and such fluid variables as thermal capacity and conductivity [20]. The functional relationships—and hence the similitude requirements—obviously become far more complex under these conditions. Finally, if electrical phenomena are also involved—as is conceivable in certain meteorological problems—any possibility of complete similitude is removed from practical consideration.

In view of the great difficulty of reproducing large-scale phenomena in accurate detail at a convenient model scale, recourse must generally be had to various laboratory expedients. Some occurrences, of course, are relatively simple in their basic aspects, and hence may be simulated—and even investigated functionally—with relative ease. Typical of these is the matter of the velocity distribution in turbulent flow near very rough boundaries without regard to gravitational, viscous, or thermal effects. In many phenomena, however, the secondary effects can be ignored only as a first approximation, and then must be restored to consideration by analytical or experimental means. Conditions illustrative of this situation are encountered in ship testing, in which the model hull is towed at such a rate as to produce similarity of the gravity-wave pattern only, that portion of the resistance due to viscous action being evaluated from supplementary tests on thin plates [9]. Likewise, in some problems of heat transfer the temperature differences are too small to influence the flow pattern appreciably—or at the worst can be assumed to cause gravitational effects but not the more complex thermodynamic effects. However, many phases of fluid motion are completely impossible to simulate in even approximate detail, and efforts must be directed toward reproduction of the over-all occurrence regardless of the discrepancies in minor effects. River models with movable beds [6] are outstanding examples of willful distortion of slope, roughness, and vertical and horizontal proportions to achieve this end; the model channels are exaggerated in depth to maintain turbulent flow, and the proper time and discharge scales for over-all similarity are then determined in preliminary tests through reproduction of known past occurrences before the desired future occurrences are investigated.

Laboratory Methods of Prototype Simulation

Since air is the fluid medium involved in nearly all meteorological phenomena, the familiar wind tunnel of aeronautical research [19] is the type of pertinent equipment which first suggests itself. Such equipment has, in fact, already been used, without modification, for the study of flow over particular boundary configurations. The fact remains, however, that aeronautical wind-tunnels are designed for specific purposes which are generally quite different from those of research in meteorology. Whereas aeronautical studies require high speeds, test sections only long enough to accommodate model planes, and dynamometers capable of measuring complex force systems, meteorological needs usually involve high capacities rather than speeds, test sections which can be adapted to a wide variety of boundary forms, and means of measuring the characteristics of the flow pattern rather than the forces exerted. In fact, various meteorological studies have been made in an open hall without any tunnel whatever, the air flow being produced by means of a series of blowers directed as desired.

Although a technique such as that last cited is often satisfactory for qualitative studies, the fact remains that the air stream in quantitative investigations must generally be uniform and of moderately low turbulence, a condition which cannot be achieved by unconfined jets from blowers except in their immediate vicinity. It is therefore preferable to house the model in a uniform duct with a properly formed bell inlet, and to draw the air through this duct by means of an exhaust fan or fans at the downstream end. Because of the large capacities rather than speeds (and hence the small pressure changes) which are involved, such ducts may be erected quite cheaply for the particular project in question, and then readily modified for the next problem to be studied. The cost of a closed circuit on a large scale is generally prohibitive, and only smaller tunnels need be made recirculating. However, complete enclosure of the duct in a large hall is desirable in order to eliminate the effect of atmospheric disturbances upon the inflow, provided the hall is of sufficient size to permit stilling of the exhaust from the blowers before the air re-enters the inlet. On the other hand, tunnel systems utilizing smoke, gas, or heated air sometimes draw directly from and discharge into the out-of-doors; such tunnels require a relatively large stilling room just upstream from the test section to eliminate the variable effect of winds. While blowers may be obtained commercially for a wide variety of flow conditions, because of the small pressure changes involved in such studies the use of surplus airplane propellers driven at relatively

low speed often provides a cheap solution for temporary installations. Although the air speed should be controllable over as wide a range as possible, peak speeds greater than 100 ft sec^{-1} are seldom necessary—a matter of considerable practical importance, since power requirements vary with only the first power of the capacity but with the cube of the speed.

Just as problems of liquid flow are sometimes studied with gas as the fluid medium in order to utilize particular measuring equipment, some phases of gaseous movement may be investigated more conveniently with liquids. This is particularly true in the observation of low-velocity convection currents in which the variation of the specific gravity of the atmosphere due to thermal effects is simulated in water either thermally or through admixture of ordinary salt. The use of different dyes to indicate the various specific gravities is far simpler than the use of smoke in air, and the greater density of water makes the measurement of low velocities more feasible. Under such circumstances the glass-walled channel of the hydraulics laboratory can become a useful observational tool in meteorological research.

Aside from such specific apparatus, perhaps the most important requisite for the meteorological model laboratory is available space in which temporary equipment of any desired nature may be constructed. Head room as well as floor area is essential, together with the possibility of erecting light walls wherever desired to eliminate extraneous currents. The frequent requests for space in university field houses for such projects during the war are indicative of research needs; a small airplane hanger would be ideal for a laboratory of this nature.

As for the specific instruments involved in model studies of atmospheric movement [10, Vol. I; 11], probably the most essential is a series of anemometers of various types for the measurement of velocity. These range from indicators of spatial mean flow of the air stream itself, through point indicators of temporal mean flow varying spatially in both magnitude and direction, to point indicators of turbulent fluctuations. If the air speeds are above some 10 ft sec^{-1}, the ordinary Pitot tube in connection with a sensitive differential gage of the Wahlen type is useful in making either point indications of mean velocity or velocity traverses for integration to yield the spatial mean. Directional Pitot tubes are also available. Duct or tunnel systems producing a pronounced acceleration of the flow into the test section permit the mean air speed to be indicated in terms of the accompanying pressure change by means of boundary piezometers at suitable points. Vaned anemometers with either counter or direct electrical indication are also useful in large passages, and midget anemometers of this nature have been made for either small-scale or low-velocity conditions. The only satisfactory instrument yet devised for the measurement of velocity fluctuations as well as mean values is the hot-wire anemometer [17], which requires for its proper operation a mean air speed in excess of about 5 ft sec^{-1} and essentially constant temperature of the ambient fluid. The hot-wire technique has progressed to the point that not only the intensity but also the scale of the turbulence and the intensity of the resulting shear can be measured. The mixing characteristics of the turbulence may also be determined by the diffusion of heat or gas from a point or line source. Such other flow characteristics as the variation of pressure and temperature require the use of standard piezometers and thermocouples, respectively.

Quite as important as the actual measurement of the model flow characteristics is the visual or photographic observation of the flow pattern [10, Vol. I; 12]. Because of the normal transparency of either air or water, it is necessary to observe the alignment of short threads or the motion of opaque material suspended in the flow, such material being so finely divided or of a specific gravity so nearly equal to that of the fluid that its motion will be essentially the same. The use of dye filaments, aluminum powder, or oil droplets is common in water. Very fine powders are sometimes used in air, but either an oil fog or a chemical smoke such as that from titanium tetrachloride is more generally satisfactory. In all cases a dark background with top lighting slightly to the rear is desirable. In both media the change in refractive index with a local change in density is sometimes utilized, rear lighting from a point source such as a carbon arc being required.

In addition to permitting visual observation or photographic recording of flow patterns, the foregoing methods are also directly useful in the preparation of motion-picture sequences for instructional purposes. Since many phenomena of atmospheric motion which are not subject to quantitative study in the laboratory can still be simulated qualitatively with relative ease, graphic training films in which the essential characteristics of these phenomena are emphasized may readily be prepared. One such film, assembled during the war by the Iowa Institute of Hydraulic Research for the Chemical Warfare Service, showed at model scale in a low-velocity air tunnel the diffusion of smoke and gas by wind under various conditions of thermal stratification, boundary roughness, forestation, and urban development. Another, prepared by the Hydrodynamics Laboratory of the California Institute of Technology, simulated by gravity currents in water the characteristics of a gas attack on a valley stronghold.

Typical Laboratory Investigations

One of the earliest scale-model investigations of atmospheric phenomena is the study by Abe [1] of wind structure over Fujiyama. According to the general criteria for similitude discussed in an earlier section, it would appear that the Reynolds number should have the same magnitude for the model and prototype flows if viscous similarity is to prevail. Abe reasoned, however, that the analogy between molecular motion on a very small scale and eddy motion on a very large scale permitted similarity to be obtained by using the molecular viscosity (a fluid property) in the model Reynolds number and the estimated eddy viscosity (a flow characteristic) in the prototype Reynolds number. As a result, his model air speeds resulted in flows which were nearly, if not wholly, in the viscous range.

Modern boundary-layer theory indicates that even the qualitative indications which Abe obtained cannot be considered trustworthy, since the separation pattern for a particular boundary configuration varies markedly with the usual Reynolds number [10, Vol. II]. In other words, as the quantity VL/ν changes from a very small to a very large magnitude, at least four distinct types of flow may be recognized, as follows: (1) at extremely low values the viscous boundary layer extends a great distance from the boundary, and the low-velocity flow near the boundary is quite stable; (2) at somewhat higher values the zone of viscous action decreases in thickness and stable eddies develop behind the major irregularities; (3) at moderate values the boundary layer becomes thin but remains laminar, separation eddies forming more irregularly and passing off into the flow; (4) at high values the boundary layer becomes turbulent and the tendency toward eddy formation is restricted to the more angular portions of the boundary. The relative Reynolds-number ranges of these several types of flow depend to a considerable degree upon the form of the boundary configuration; no general values can therefore be given, but for rough indications reference may be made to standard diagrams of surface and form resistance for simple plates and bodies of revolution [15].

While the complete simulation of flow over gently undular terrain cannot be accomplished at model scale because of the extremely high velocities which would be required to maintain constancy of the Reynolds number, it so happens that flow past angular boundary configurations at all but very small scales and air speeds is essentially independent of viscous effects. In other words, at moderate to high Reynolds numbers the separation pattern for such "unstreamlined" forms is determined almost solely by their geometry. Therefore, with geometric similarity of the boundary the normal air speed of a wind tunnel is sufficiently high for similitude to be obtained. Typical of such an investigation is the study at model scale of air currents in the vicinity of the Rock of Gibraltar [8], the laboratory indications being in close agreement with observations made at the actual site.

If the terrain is not sufficiently irregular to ensure freedom from viscous effects, one is tempted to distort the vertical scale of the model until the necessary degree of angularity is secured. The success of such a procedure depends largely upon the degree of similitude to be obtained, since the theory of both the potential flow around a body and the viscous effects upon the basic pattern point to a gradual departure from similarity as the boundary geometry is distorted. It is altogether probable that the gain in freedom from viscous effects is wholly offset by the changes in pattern which this distortion produces. In other words, as the terrain in question becomes less irregular, true similitude can be approached in the model only by increasing the velocity of flow accordingly. The obvious limit of such increase is the onset of marked compressibility effects as the Mach number of the model approaches unity.

In addition to the mean flow pattern and the pattern of the major eddies, which formed the basis of the Gibraltar study, the accompanying pattern of turbulence is often of importance. As in all turbulence studies, at least three characteristics of the turbulence may be involved: (1) the intensity, or root-mean-square velocity of fluctuation; (2) the scale, or mean eddy size; (3) the diffusivity, or eddy viscosity, which is proportional to the product of the intensity and the scale. For flow over relatively smooth boundaries, the development of the turbulent boundary layer is primarily a function of the Reynolds number, and model similarity would hence require approximately the same magnitude of the Reynolds number as in the prototype. While this is practically out of the question, enough is now known about the mechanics of the boundary layer to permit evaluation of prototype conditions by analytical rather than experimental means [5]. With uneven boundaries, however, particularly those which involve large-scale irregularities rather than small-scale roughness, not only is the boundary-layer theory incomplete but specific problems are more readily subject to model investigation. Fortunately, although true atmospheric turbulence contains eddies varying over an extremely wide range of intensity and scale, it is usually the larger magnitudes of each which are of major interest and are at the same time most nearly reproducible by models.

Laboratory investigations of the turbulent diffusion of heat, gas, and water vapor in the neighborhood of smooth boundaries have frequently been made—but as general studies of the mechanics of such diffusion rather than as specific model studies of particular prototype conditions. In connection with the previously mentioned training film on chemical warfare, on the other hand, evaluation of diffusion characteristics for various rough-boundary configurations studied at the Iowa Institute required the detailed measurement either of the turbulence characteristics or of the actual diffusion of some material introduced into the flow. Typical of such evaluation, for example, was the experimental determination of the diffusion of gas or smoke from point

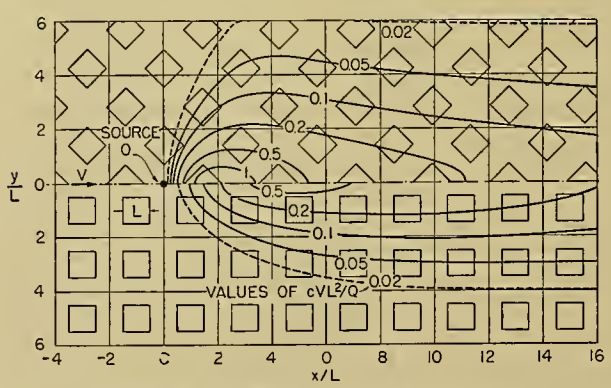


FIG. 1.—Generalized diffusion patterns at height L for two different block arrangements.

and line sources in typical urban districts. Owing to the pronounced angularity of the model buildings, the pattern of turbulence (or at least the large-scale portion

thereof) was essentially independent of the Reynolds number. Measurements of gas concentration c could therefore be reduced to generalized nondimensional contours of relative concentration cVL^2/Q, in which Q is the rate of gas release, V the mean air speed, and L a characteristic building dimension, as shown schematically in Fig. 1. Similar techniques are evidently applicable to studies of smoke abatement, the dispersion of radioactive by-products, or—at least qualitatively—the mixing of moisture-laden air in mountainous terrain.

Two somewhat allied phenomena of the atmosphere are impossible to reproduce quantitatively to scale in the laboratory, yet they involve complexities which are so great that even the qualitative laboratory simulation of their primary characteristics is often useful. One of these is the effect of the earth's rotation, and the other the effect of thermal stratification. Since the accelerative forces due to rotation of the flow system as a whole cannot well be included in a general model, it is usually necessary [13] to reproduce the effect in a rotating tank at very small scale, as is described elsewhere in this volume,[1] or—as is sometimes done—in a small rotating room. At such scale, of course, viscous action plays a role which is wholly out of proportion to that of any prototype condition, and the indications must be interpreted in the light of the mechanics of viscous flow. Although studies of this nature generally involve thermal (or density) stratification of the fluid [7], other stratification phenomena warrant detailed study in their own right. These, in fact, are of interest to a number of professions (oceanography, harbor hydraulics, and sedimentation), and have often been reproduced in the laboratory [2].

For purposes of convenience, studies of this nature are usually conducted in water, the density stratification being obtained by means of saline solutions of various concentrations. This technique has permitted the study of (1) wave motion at and mixing across a density interface, (2) the progress of sediment underflow (comparable to a dust cloud or even a cold front) through a reservoir, and even (3) the effect of a schematic mountain range upon successively higher thermal zones of the atmosphere. Under certain conditions, to be sure, viscosity plays a role which is quite out of keeping with prototype conditions, and thermal effects other than that of the density differential cannot be reproduced. However, so long as the interrelationship between gravitational action and mass reaction is the primary factor involved, model studies of this nature can be expected to yield useful indications if not reasonably accurate numerical results. Noteworthy is the fact that the relative motion of wind and water represents simply an extreme degree of stratified flow, moderately small-scale investigations of wind-driven waves having yielded useful indications of the forces involved [18].

Previous mention of turbulence has been restricted to that resulting from the surface resistance or form resistance of the flow boundary. The diffusion resulting from such turbulence is known as forced convection. Related in mechanism but distinct in origin is that type of turbulent mixing resulting from an unstable density-stratification of a fluid, the result being known as free convection. Typical of free convection is the pattern of mean flow and eddy motion produced by a boundary source of heat below an otherwise quiet fluid. The heated fluid becomes less dense than its surroundings and tends to rise, and at the same time inflow toward the rising column of eddies is induced through displacement and entrainment. Although other thermodynamic effects are involved in both the original heating and the subsequent mixing and cooling, except under extreme conditions these remain secondary to the effects of gravitational acceleration and turbulent diffusion. In other words, the essential aspects of the free convection could be reproduced in water by the introduction of heat or finely dispersed bubbles at a lower boundary or by the removal of heat or introduction of sediment at an upper boundary. In view of the complexities involved in giving proper consideration to an additional dimension and to the corresponding flow and fluid variables, it is therefore expedient as a first approximation to disregard the thermodynamic effects in such investigations and reduce the problem to one of mechanical (*i.e.*, dynamic) similarity alone.

Up to the present, laboratory studies of this nature have been restricted to the analysis of free convection under generalized conditions. There is no obvious reason, however, why similar techniques could not be applied to specific prototype conditions of boundary configuration and wind orientation; under any circumstances, existing data obtained at small scale should be applicable at least approximately to general prototype conditions. Results now at hand apply to convection from a point source and from a line source in stagnant fluid and from a line source in a uniform wind [16]. In each case the convection was attained in air by means of heat (either low gas flames or electric hotplates), the measurements of temperature differentials being reduced to specific-weight differentials by the simple ideal-gas equation. With reference to the defini-

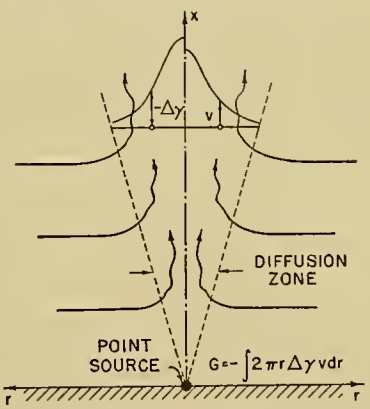


Fig. 2.—Definition sketch for free convection above a source of heat.

1. Consult "Experimental Analogies to Atmospheric Motion" by D. Fultz, pp. 1235–1248.

tion sketch of Fig. 2 for the point source, measurements of temperature and velocity made by Yih at the Iowa Institute over a considerable range of distance and heat output could thus be generalized in the composite dimensionless plots shown in Fig. 3. Noteworthy is the

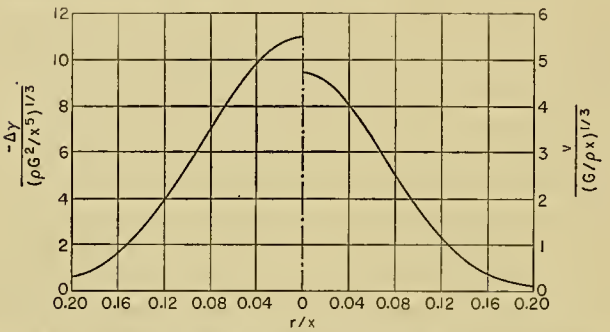


FIG. 3.—Dimensionless representation of experimental results for the conditions of Fig. 2.

fact that for any particular value of r/x a number of the Froude type—$F = v/\sqrt{x\Delta\gamma/\rho}$—is constant, indicating that the phenomenon is truly gravitational in its fundamental aspects.

Summary and Conclusions

Through the many professions dealing with fluid motion, there has been obtained a considerable amount of experience with scale-model investigations, which should be of use in developing comparable procedures for research in the field of meteorology. The general principles of dimensional analysis at once indicate the criteria for similarity of the various pertinent flow phenomena, and at the same time serve as a guide for effective research. Likewise, the closely related aspects of many problems in fluid motion make various techniques developed in other fields directly applicable to that in question. In other words, the scale model is now an established rather than an untried scientific tool, and the problem is one of adaptation to somewhat new requirements rather than of complete origination.

Nevertheless, too great emphasis cannot be given to the fact that each of the professions now utilizing model techniques has perfected them only through many years of painstaking labor. Dimensional analysis is an invaluable guide but in no sense an automatic control, and its proper use continues to depend upon sound experience both in selecting the essential variables and in reducing to a practicable form the complex relationship which usually results. Moreover, the extent to which special requirements of each particular field govern the experimental techniques can be thoroughly understood only through actual laboratory contact. The fact therefore remains that the foregoing presentation by one who is only indirectly familiar with the problems of this specific field can be but a rough indication of the situation which should exist after a number of years of direct accomplishment by future specialists.

At present writing it appears that meteorologists could profitably expand, according to their own requirements, many aspects of model investigations already proven useful in related fields. These include studies of both the mean flow pattern and the eddy structure produced by irregular boundary configurations; research in the diffusion of heat and vapor by both forced and free convection; experiments on stratified flows; and—at least in an exploratory manner—the simulation of conditions which combine two or more such effects. To what extent attainment of the Reynolds criterion of viscous similitude may become possible for other boundary conditions, and whether or not thermodynamic similarity may also become practicable, must be decided in the future. In any event, the profession can rest assured that the perfection of meteorological model techniques will open many an avenue of fruitful study hardly conceivable in advance.

REFERENCES

1. ABE, MASANAO, "Mountain Clouds, Their Forms and Connected Air Current." *Bull. cent. meteor. Obs., Tokyo*, Vol. 7, No. 3 (1929).
2. BELL, H. S., "Density Currents as Agents for Transporting Sediments." *J. Geol.*, Vol. 50, No. 5 (1942).
3. BRIDGMAN, P. W., *Dimensional Analysis*. New Haven, Yale University Press, 1937.
4. BUCKINGHAM, E., "Model Experiments and the Forms of Empirical Equations." *Trans. Amer. Soc. mech. Engrs.*, 37: 263–296 (1915).
5. DRYDEN, H. L., "Some Recent Contributions to the Study of Transition and Turbulent Boundary Layers." *Tech. Notes nat. adv. Comm. Aero., Wash.*, No. 1168 (1947).
6. *Engineering Hydraulics*, H. ROUSE, ed. New York, Wiley, 1950.
7. EXNER, F. M., *Dynamische Meteorolcg'e*, 2. Aufl. Vienna, J. Springer, 1925.
8. FIELD, J. H., and WARDEN, R., "A Survey of the Air Currents in the Bay of Gibraltar, 1929–30." *Geophys. Mem.*, Vol. 7, No. 59 (1933).
9. *Fifth International Conference of Ship Tank Superintendents*, G. HUGHES, ed. London, H.M. Stationery Office, 1949.
10. *Modern Developments in Fluid Dynamics*, S. GOLDSTEIN, ed., 2 Vols. London, Oxford, 1938.
11. PETERS, H., "Druckmessung," *Handbuch der Experimentalphysik*, L. SCHILLER, Hsgbr., Bd. IV, Teil 1, SS. 489–510. Leipzig, Akad. Verlagsges., 1931.
12. PRANDTL, L., and TIETJENS, O. G., *Applied Hydro- and Aeromechanics*. New York, McGraw, 1934.
13. ROSSBY C.-G., "On the Solution of Problems of Atmospheric Motion by Means of Model Experiments." *Mon. Wea. Rev. Wash.*, 54: 237–240 (1926).
14. ROUSE, H., *Fluid Mechanics for Hydraulic Engineers*. New York, McGraw, 1938.
15. —— *Elementary Mechanics of Fluids*. New York, Wiley, 1945.
16. —— "Gravitational Diffusion from a Boundary Source in Two-Dimensional Flow." *J. appl. Mech.*, 14: 225–228 (1947).
17. SCHUBAUER, G. B., and KLEBANOFF, P. S., *Theory and Application of Hot-Wire Instruments in the Investigation of Turbulent Boundary Layers*. N.A.C.A. ACR No. 5K27, 1946.
18. SVERDRUP, H. U., and MUNK, W. H., "Wind, Sea, and Swell: Theory of Relations for Forecasting." *U. S. Navy, Hydrogr. Off. Publ.* No. 601 (1947).
19. TOUSSAINT, A., "Experimental Methods—Wind Tunnels," in *Aerodynamic Theory*, W. F. DURAND, ed., Vol. III, pp. 252–319. Berlin, J. Springer, 1935.
20. WEBER, E., "Physical Units and Standards," Section 3, in *Handbook of Engineering Fundamentals*, O. W. ESHBACH, ed. New York, Wiley, 1936.

EXPERIMENTAL CLOUD FORMATION

By SIR DAVID BRUNT

Imperial College of Science and Technology

Historical Note

The motions which occur in unstable layers of fluid in which there is no general motion of the fluid have been described independently by a number of writers. The "cellular" convection which occurs in unstable fluids has been associated with the name of Professor Henri Bénard, on account of the thorough study which he made of the phenomena which he described in detail [1]. The present writer gave the name of "Bénard cell" to the typical convection cell which occurs in unstable layers initially at rest [3]. Bénard's observations were anticipated about twenty years earlier by James Thomson [17] (the brother of Lord Kelvin), who observed a "tessellated structure" in cooling soapy water in a tub casually seen in the yard of an inn. Thomson later reproduced the phenomenon in the laboratory. Weber [18] also observed a structure, similar to that described by Thomson, in a layer of alcohol and water on the slide of a microscope, but Weber gave a wrong interpretation of his observations. A correct interpretation was given later by Lehmann in *Molekularphysik*, Vol. I, p. 279. Many other subsequent writings by various authors followed the accidental discovery of convection in a cellular form, but only the earliest discoverers need be mentioned at this stage. Further brief accounts of independent discoveries of the same phenomena in a variety of liquids will be found in *Nature* (April and May, 1914).

In Fig. 1 are shown the lines of flow in a central vertical section of the typical Bénard convection cell. When the fluid has a large shearing motion, the axis of convection is replaced by a vertical plane of convection, and the fluid is divided into a series of parallel rolls, rotating in opposite directions, and Fig. 1 now

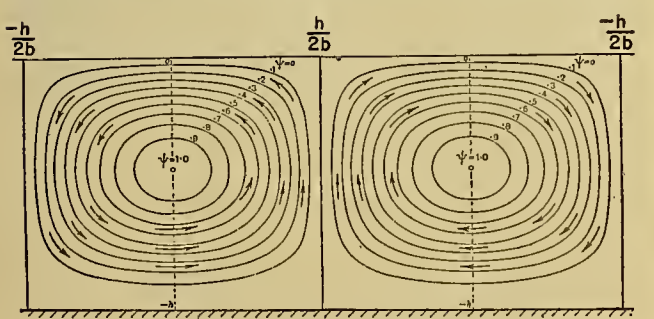


Fig. 1.—Circulation in the Bénard convection cell.

gives the general features of the circulation, transverse to the axis of shear, for a pair of adjacent rolls.

The first detailed investigation of the effects of shear was made by Terada [16], who carried out a series of investigations with a variety of liquids in which the rate of flow increased with height, but without change of direction. Terada's paper contains a number of striking photographs of these longitudinal (downwind) rolls, and the author draws attention to the opposite sense of rotation of adjacent rolls, stating that the rolls have a circular cross section.

Phillips and Walker [13] showed, by experiments in air, that rolls transverse to the direction of shear can be formed when the shear is small, and Graham [8], a pupil of Walker, later confirmed this result.

The Nature of the Simple Convection Cell

When instability in a fluid initially at rest breaks down, the fluid is seen to become divided into a number of polygonal cells. In a liquid, the motion normally consists of ascent in the centre of the cell, outward motion at the top, downward motion at the outer margin, and inward motion at the bottom of the cell.

The simplest illustration of such motion is obtained by pouring *cheap* gold paint into a small vessel to a depth of, say, a quarter of an inch. In cheap gold paint the liquid medium is usually benzene or some other highly volatile liquid, while the "gold" is in the form of thin flakes. The evaporation of the liquid leads to such rapid cooling at the top of the layer that marked instability is immediately produced. As the thin flakes tend to set themselves parallel to the motion, they are

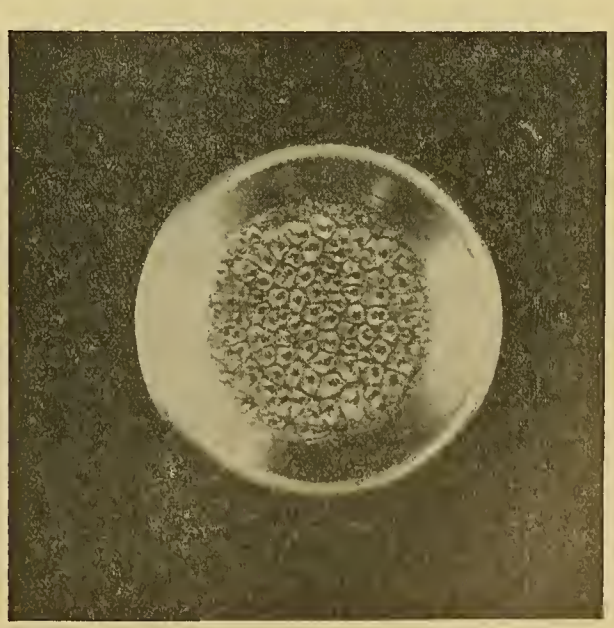


Fig. 2.—Convection cells formed in gold paint.

not visible at the centre or periphery of the cell, where they are seen edgewise from above. The cell will thus appear to have clear liquid at the centre and at the

outer boundary, as is shown in Fig. 2. Similar motions are produced in molten wax and in molten metals cooled from above, just as in any liquid cooled from above. Thus Bénard's discovery of convection cells was made, during experiments with liquid coherers containing metallic particles, in early experiments on wireless reception.

The clearly defined convection cells shown in Fig. 2 do not form immediately. When the volatile liquid is poured into a dish the liquid at first appears to be violently disturbed, showing no recognisable pattern. If the liquid is exposed to a light current of air, or if, even in still air, it is left uncovered, the violently disturbed state continues for a long time. When the vessel containing the volatile liquid is covered by a sheet of glass placed above the surface, but not in contact with the edges of the vessel, the free evaporation of the liquid is checked, and the cells as shown in Fig. 2 are quickly formed and retain their general form for an almost indefinite period.

Similarly, in the experiments on polygonal convection cells in air, described below, there is an initial stage in which the air is divided into polygonal cells of four, five, or six sides. These cells are unsteady, and eventually settle down to a fairly steady pattern with five or six sides.

The Theoretical Discussions of Lord Rayleigh and Others

Lord Rayleigh [14] was led to examine the physical conditions leading to the formation of the convection cell by seeing Bénard's description of his experiments, in some of which the cells were all truly hexagonal. Rayleigh started from the hypothesis that the velocities, and the departures of pressure and temperature from their initial values, were all of the form $e^{i\alpha x}\, e^{i\beta y}\, e^{i\gamma z}\, e^{ct}$, and that the motion with the highest increment would rapidly predominate over all others. He found a new criterion, showing that no motion will occur in a "statically" unstable liquid, unless

$$\frac{\rho_1 - \rho_0}{\bar{\rho}} > \frac{27\pi^4 \kappa \nu}{4gh^3},$$

where ρ_1, ρ_0, and $\bar{\rho}$ are the densities at the top and bottom of the layer and the mean density within the fluid, respectively; κ is the coefficient of thermometric conductivity; ν the kinematic coefficient of viscosity; and h the depth of the layer.

Thus for any given depth of layer there is a finite value of the difference of density at top and bottom of the layer which must be exceeded before motion can occur. Moreover, the shallower the layer, the greater is the difference of density required to produce motion, while for a given difference of density there is a lower limit to the depth in which motion can occur. This last feature is readily displayed in any suitable liquid. If, for example, a cheap gold paint is used as a medium, and the containing vessel is tilted so that at one side the depth decreases to zero, the diameter of the cells will be greatest at the deeper edge and will decrease with decreasing depth, but only to a finite size, beyond which there will be a shallow layer showing no cellular motion.

Bénard [2] showed that Rayleigh's theory gave, for a circular cell, the value of 3.285 for the ratio of diameter of cell to depth of fluid. Bénard's own experiments on spermaceti gave values of 3.27 to 3.34 for this ratio.

Rayleigh treated the case when the fluid is limited by two free surfaces. Two other cases require consideration, that of two rigid boundaries, and that of one free surface, the other surface being a rigid boundary. The constant A in the criterion

$$\frac{\rho_1 - \rho_0}{\bar{\rho}} > \frac{A\kappa\nu}{gh^3}$$

has been evaluated for different cases by Jeffreys [9] and more recently by Pellew and Southwell [12]. The latter writers give the following values of A:

	A
1. Two free surfaces	657.5
2. Two rigid boundaries	1708
3. One free, one rigid surface	1101

Rayleigh gave for case (1) the same value of A, equal to $\frac{27\pi^4}{4}$ or 657.5. Jeffreys [9] gave 1709.5 for case (2). The detailed methods of computation used by these authors differed somewhat, and so the values given above may be accepted as reasonably accurate.

The theoretical treatment by all the above-mentioned authors is based on the assumption that the components of velocity are sufficiently small to justify the neglect of their squares or products. This is a limitation which may lead to some uncertainties in the comparison of theory and experiment. Rayleigh further points out that in his treatment of the problem it is tacitly assumed that $(\rho_1 - \rho_0)/\bar{\rho}$ is small. If, then, the depth of the unstable fluid is small, the critical value of $(\rho_1 - \rho_0)/\bar{\rho}$ as defined by $\frac{A\kappa\nu}{gh^3}$ becomes large, and the interpretation of the criterion as stated by Rayleigh is open to doubt. When the fluid is a gas and the instability is produced by heating the gas from below, it becomes impossible to satisfy the criterion for small depths of the unstable layer, and polygonal cells of convection will not then occur.

We have therefore to embark upon the consideration of small-scale laboratory experiments bearing in mind that the theory of Rayleigh and of others is not likely to include all phenomena that may occur.

Motion in Unstable Layers of Air Having No General Motion

A convenient apparatus for studying the motions in unstable air can readily be constructed, having as base a smooth metal plate, fitting on top of a shallow box containing a series of parallel glass rods on which a thin wire is wound. The base plate is heated by the passage of an electric current through the wire. The top of the chamber consists of a plane sheet of glass which forms the base of a vessel that can be filled with water to a sufficient depth to act as a check

against any great rise of temperature at the top of the chamber. In a series of experiments carried out by Chandra [4] at the Imperial College of Science and Technology, South Kensington, London, a chamber so constructed was used, and the top of the chamber was supported on a series of brass cylinders, sets of such cylinders being made of lengths from 2 to 16 mm. The sides of the chamber were filled in by layers of felt, sufficient to prevent a rapid leakage of air into or out of the chamber.

In Chandra's experiments the temperature of the air was measured by means of platinum resistance thermometers, fixed as near as possible to the top and bottom plates respectively, while a third was fixed halfway between the plates. Observations of the temperature distribution within the chamber could thus be made at any time. The motion within the chamber was made visible by means of cigarette smoke, which was injected by means of a two-way pump. The motion could be watched and photographed through the top plate.

As might be inferred from Rayleigh's formula, no motion was observed until the excess of temperature at the base over that at the top exceeded a finite limit. Thus, with a chamber of depth 10 mm, the critical temperature difference was 11.4C. Observations were made with depths from 2 to 16 mm.

Figure 3 shows the structure found in a chamber of depth 8 mm, with a temperature difference of 28C

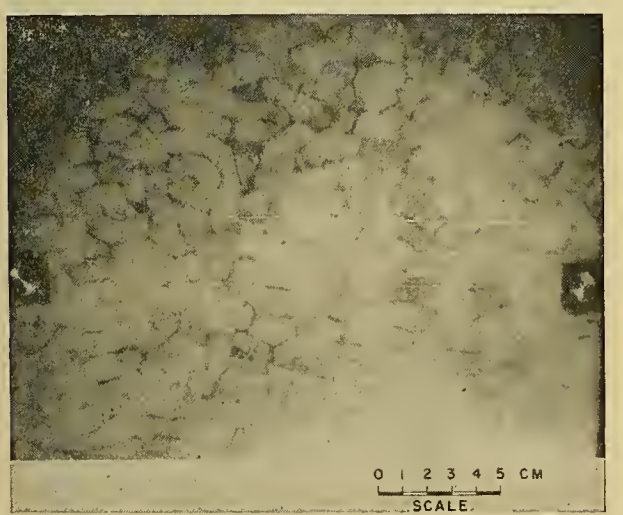


Fig. 3.—Convection pattern in a chamber of depth 8 mm; temperature difference between top and bottom of chamber, 28C.

between top and bottom of the chamber. This shows a series of polygonal cells, in each of which the motion was downward at the centre, upward at the outer margin, and inward at the top. In addition, there are some long rolls which fill the chamber when the smoke is first injected, and which gradually break up into separate polygonal cells. This feature was particularly noticeable in all chamber depths of 8 mm or more, but eventually the long rolls were replaced by polygonal cells.

Figure 4 shows the structure found in a chamber of depth 7 mm, the whole chamber being filled with polygonal cells almost instantaneously after the injection of the cigarette smoke. The rapidity with which the initial long rolls or strips of smoke would break up into the cellular pattern in a chamber of depth 7 mm was very striking. In Fig. 4 the diagonal of the hexagonal polygons was approximately three times the depth of the chamber, while in Fig. 3 the diagonals of the polygons were usually little greater than twice the depth, as the steady state had not been attained.

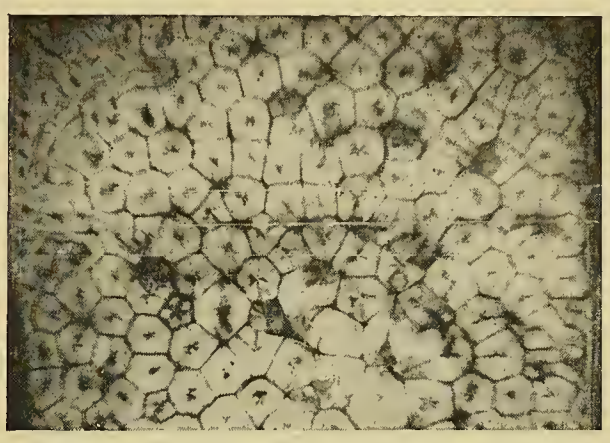


Fig. 4.—Convection pattern in a chamber of depth 7 mm.

With a chamber of depth 6 mm or less, polygons could not be formed with any vertical distribution of temperature. The change in the nature of the phenomena in going from 7 mm to 6 mm is best illustrated by Fig. 5, in which the base plate of the chamber was

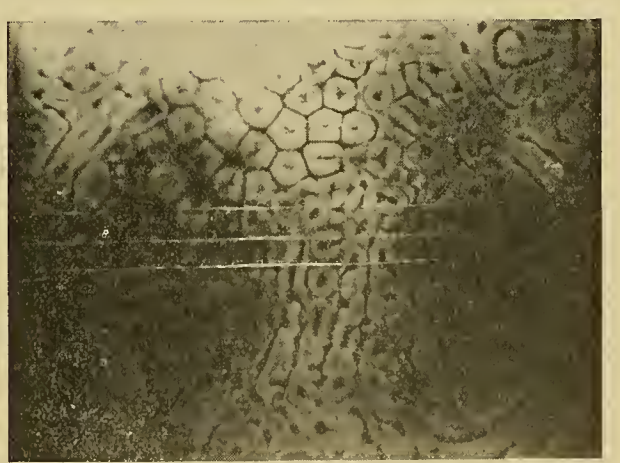


Fig. 5.—Convection pattern in a chamber of depth 6 mm. Where polygonal cells appear, a dent in the base plate yielded a depth of 7 mm.

dented downwards over the area covered by polygonal cells in the diagram. Over this area the depth of the chamber was about 7 mm, and above the remainder of the plate the depth was 6 mm, when the photograph was taken. When the smoke is first injected into the chamber it forms long rolls such as still appear in the photograph at one side. Each roll has a central white

line at first, but after a short time the pattern changes to a series of almost circular cells, each having a well-marked, clear centre. No motion is readily detected, but by the injection of small puffs of smoke through a capillary tube with its open end placed at the centre of the clear space, and in contact with the base plate, it was easily possible to detect ascent at the centre of these cells.

A structure similar to that found in layers of air of depth 6 mm was also obtained in deeper layers at very high temperatures. The reason for this is readily seen from a consideration of Rayleigh's criterion quoted above. If the difference of density, *i.e.*, of temperature, required to produce cellular motion is to be attained by heating the base of the chamber, it is clear that when there is a large difference of temperature between the top and bottom of the chamber, the mean temperature of the chamber will also be raised considerably. This leads to a rapid increase in the product $\kappa\nu$, and in shallow chambers it will be impossible to satisfy Rayleigh's criterion for motion.

The curious structure shown over the greater part of Fig. 5 can be formed with quite small differences of temperature in shallow chambers, differences which are much lower than are required by Rayleigh's criterion.

A simple expedient made it possible to test the view stated above. A chamber was constructed with a glass base and a metal top around which was fixed a narrow flange, so that liquid air could be poured into the vessel so formed. The evaporation of the liquid air cooled the top of the chamber, and gave polygonal cells readily in chambers of any depth. These could be easily seen from below, though the condensation produced inside the chamber quickly frosted the glass and shut off any view of the inside of the chamber.

A final experiment in the production of the simple convection cells merits mention. A depth of 12 mm was given to the chamber, and it was found possible to lay a dense but very shallow layer of smoke over the base plate. Almost immediately after the layer was formed a series of round holes were punched in the layer of smoke by the descent of colder air from above, giving a series of circular clear patches over the plate. Within a few seconds, convection patterns began to show within the chamber, placed centrally over the holes in the smoke carpet. This circulation carried small streamers of smoke upward and inward towards the centre, as shown in Fig. 6, and here and there it is possible to detect the streamers of smoke which have not yet reached the centre. After some minutes the whole chamber became filled with convection cells such as are normally obtained when smoke is injected into the chamber.

In the laboratory experiments described above, the convection cell has both centre and periphery clear of smoke. This is presumably to be explained by the formation of a "dust-free" space in immediate contact with the heated base plate. The air which fills the centre and periphery of the cells has been drawn from this dust-free space immediately above the base plate.

It should be emphasised that two types of motion are found in the course of these experiments. In the first type, the genuine Bénard convectional cell of polygonal form, the motion is downward in the centre.

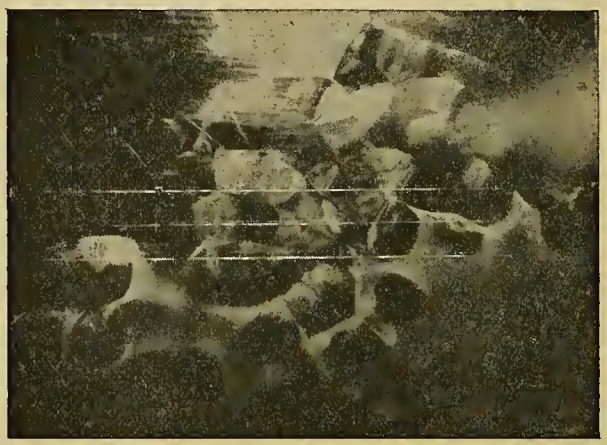


FIG. 6.—Dense layer of smoke over base plate in a chamber of depth 12 mm, showing clear holes formed in this smoke layer, and subsequent development of convection cells above the holes.

In the second type, shown in Fig. 5, alongside the typical polygonal cells over a restricted area, there is some difficulty in determining the nature of the motion by simply watching it with the naked eye, though it was found by injecting small quantities of smoke that there was ascent in the centre of the roughly circular structures. The motion in these cases will be referred to as "convection of the second type."

The results derived by Chandra are summarised in Fig. 7. The continuous line separating the two

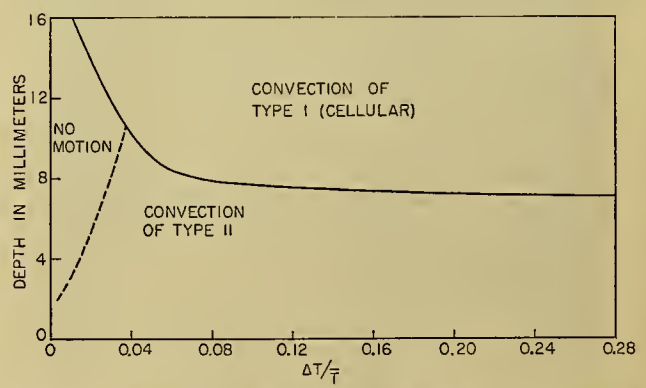


FIG. 7.—A summary of Chandra's observations of limiting conditions in air (*cf.* Fig. 12).

types of convection becomes asymptotic to a horizontal line a short distance below the 7-mm level, indicating that only the second type of convection can be obtained in chambers of depth appreciably less than 7 mm.

Motion in Unstable Layers Having a Shearing Motion.

Depth of Chamber 7 mm or Greater. A modification of the apparatus previously described was used to investigate the effect of shear on the form of the cir-

culations produced in unstable layers. In this modification the top of the chamber was a long glass plate, which could be moved across at any desired rate, being driven by a small electric motor. The top plate produces a shear, analogous to a wind varying with height.

It was found that unless the excess of temperature at the base over that at the top was above the limiting value found in the investigation with the top at rest, no structure was observable in the smoke, no matter what the rate of shear. In all the experiments with shear, care was therefore taken to ensure that the vertical gradient of temperature exceeded the limit appropriate to the depth of chamber.

In the first experiments made in these conditions, the shear was given a high value. Figure 8 represents

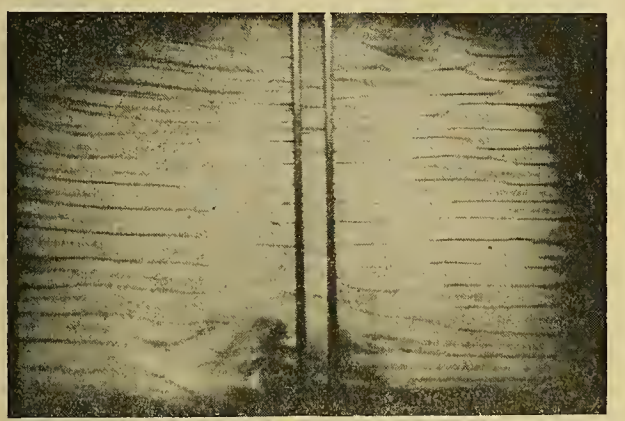


FIG. 8.—The effect of shear, by motion from left of right of the top plate, in a chamber of depth 8 mm, with temperature difference of 5°C.

the flow pattern in a layer 8 mm in depth, with a temperature difference of 5°C, and with the top plate moving at a rate of 10 cm sec^{-1}. The chamber is filled by a set of double rolls, with descent at the common boundary of the two rolls in a pair, and ascent at the

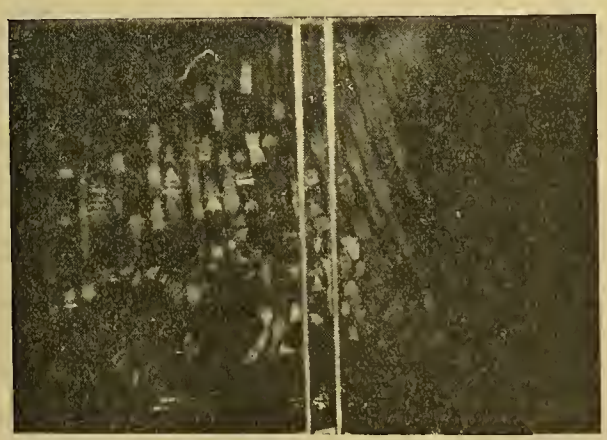


FIG. 9.—Transverse rolls formed by small shearing motion, with temperature difference of 9°C and rate of motion of top plate of 9 cm sec^{-1}.

other boundary. The central boundary of a pair, at which there is descent, is not so clearly defined as the boundary which separates adjacent pairs. This is clearly shown in Fig. 8. Rolls which are directed along the direction of shear are called "longitudinal" rolls.

When the rate of motion of the top plate is decreased to a considerably lower value, the longitudinal rolls are replaced by "transverse" rolls, or rolls perpendicular to the direction of shear. Careful observations were made to investigate whether these rolls had alternating sense of rotation, and it was established that adjacent transverse rolls do undoubtedly rotate in opposite directions. Figure 9 shows a system of transverse rolls.

With a still smaller rate of shear, and with the vertical gradient of temperature held constant, it was found that no rolls, either longitudinal or transverse, were formed, but that the original polygonal structure with no shear was distorted, as shown in Fig. 10. The

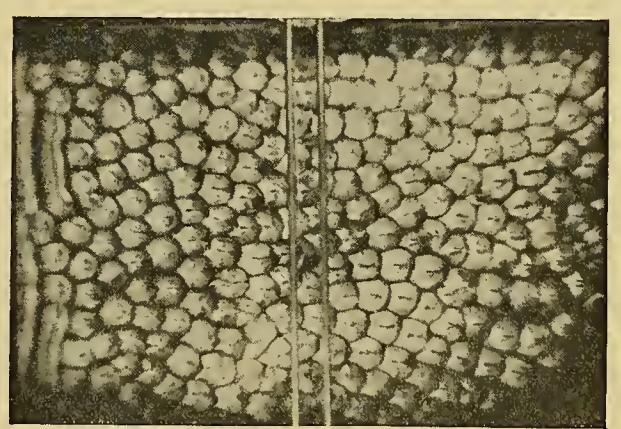


FIG. 10.—Distortion of the cell pattern by small shear.

observations made by Chandra [4] and by Dassanayake [5] indicate that the maximum rate of shear which will produce transverse rolls increases as the vertical gradient of temperature increases. Dassanayake made a special effort to maintain a constant vertical temperature difference of 4°C with a chamber 10 mm deep, and to approach as nearly as possible to the limiting shear above which the rolls are longitudinal, and below which the rolls are transverse. He succeeded in obtaining photographs showing both transverse and longitudinal rolls, occurring simultaneously, and superposed over a part of the chamber, when the top plate was moved at a rate of 0.5 cm sec^{-1}. The photographs were, however, not suitable for reproduction, being lacking in marked contrast.

With a difference of 4°C in temperatures at top and bottom of the 10-mm chamber, motion of the top plate will give transverse or longitudinal rolls, according as the rate of motion of the plate is less than, or greater than, 0.5 cm sec^{-1}. The mechanical arrangement used by Chandra for moving the top plate was insufficiently reliable to permit of accurate control of the shear, and his observations were limited to high values of the temperature difference and high rates of shear. We can therefore analyse the phenomena, in chambers of depths of 7 mm or more, as follows, for a given difference of temperature between top and bottom of the plate:

1. When there is no shearing motion, the chamber

is filled with polygonal convection cells, as described earlier.

2. When the top plate is moved at a rate not exceeding a certain low limiting value u_1, the original convection cells are distorted into a horseshoe pattern, as shown in Fig. 9.

3. When the top plate is moved at a rate exceeding u_1, but not exceeding another limiting value u_2, the chamber is filled with transverse rolls.

4. When the top plate is moved at a rate exceeding u_2, the chamber is filled with longitudinal rolls.

For a given depth of chamber, u_1 and u_2 will both increase as the difference of temperature between top and bottom of the chamber increases. It is not possible to state the values of u_1 and u_2 for different depths of chamber and for different temperature gradients. Only one value, u_2 for depth 10 mm and temperature difference 40C, is known, u_2 then being 0.5 cm sec^{-1}, approximately.

Depth of Chamber Less Than 7 mm. In the absence of shear the chamber is filled with structures such as are shown in Fig. 5. When a large shearing motion is imposed, the chamber is filled with spindlelike structures, and never shows the long continuous rolls which occur in deeper chambers. As the rate of shear is decreased the spindlelike structures become shorter and more irregular, but at no stage is the chamber filled with transverse rolls. Some idea of the completely different structures obtained in different depths of fluid may be gathered from Fig. 11, which was taken in a

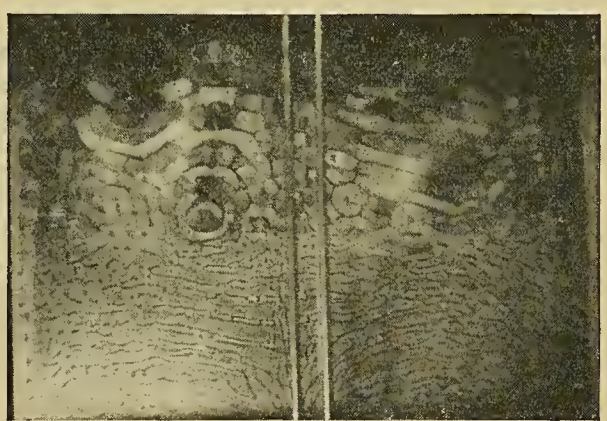


FIG. 11.—Change of cell pattern with depth. The picture represents a chamber of depth 8 mm at the top and 6 mm at the bottom of the picture. The temperature difference between top and bottom of the chamber was about 42C, and the rate of motion of top plate was 7 cm sec^{-1}.

chamber in which the depth varied from 6 mm at one side to 8 mm at the opposite side, with a temperature difference of 42C. The photograph in Fig. 11 was taken shortly after the top plate had been stopped.

A Comparison of the Phenomena in Air and in Carbon Dioxide

Dassanayake [5] carried out an investigation of the forms of convection in carbon dioxide, the gas being bubbled through hydrochloric acid and a solution of ammonium carbonate in Dreschel bottles, and finally through concentrated sulphuric acid to remove all water drops. The amount of fogging was kept as low as was consistent with ease of observation of the patterns of flow, in order to avoid seriously altering the character of the medium.

Carbon dioxide was chosen as a medium for this investigation because the values of κ and ν are considerably lower for carbon dioxide than for air. Thus at 15C $\kappa\nu$ is 0.030 for air, and 0.0076 for carbon dioxide. In Rayleigh's criterion quoted earlier, the critical difference of relative density is proportional to $\kappa\nu/h^3$. Thus for a given depth of chamber the critical condition established by the criterion is satisfied by a much lower gradient of relative density in carbon dioxide than in air. Thus it was found possible to form polygonal cells in carbon dioxide in a chamber of 4.5 mm depth, whereas in air no polygonal cells could be formed in a chamber of less than 7 mm depth. It should be noted that $(4.5/7)^3$ is equal to 0.265, which is in reasonable agreement with the theoretical value (0.25) of the ratio of $\kappa\nu$ for the two gases at 15C.

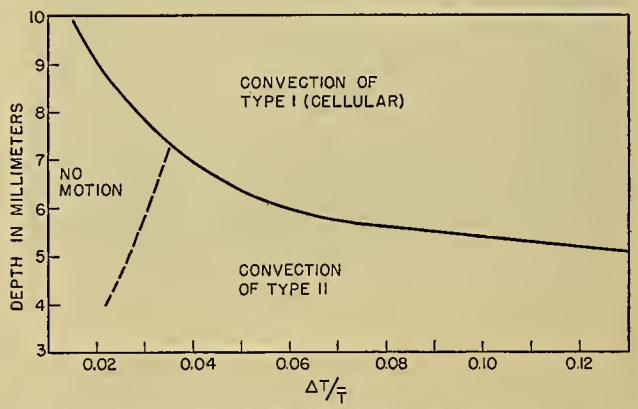


FIG. 12.—A summary of Dassanayake's observations of limiting conditions in carbon dioxide (*cf.* Fig. 7).

Figure 12 shows the limiting curves derived by Dassanayake for carbon dioxide. The results are similar in character to those of Chandra shown in Fig. 7.

Applications To the Atmosphere

In the atmosphere, motion in the convection cell is normally the reverse of that observed in the laboratory experiments, a central ascending current being surrounded by a descending current of much smaller speed. Clouds form within the ascending currents when these extend to a sufficient height to produce condensation. In considering the atmospheric analogues of the laboratory phenomena, we should therefore look for cloud formation within the regions corresponding to those in which descending motion is found in the laboratory experiments, bearing in mind that instability is an essential condition for the occurrence of these phenomena. The precise nature of the clouds will depend on the rate of shear, that is, on the rate of variation of wind with height. The main classes of convection cloud are classified in Table I. This table includes both cloud forms directly formed by condensation, and those

which are formed in the process of dissipation of clouds within which instability is set up, either by the joint effect of radiation and absorption, or by subsidence.

TABLE I. CLASSES OF CONVECTION CLOUD

Rate of variation of wind with height	Type of cloud	International Cloud Atlas
Zero	Isolated cumulus	N 10, 25, 55, 56
Very small	Cloudlets arranged in lines	N 16
Small	Short rolls, or long irregular wavy rolls, nearly across wind	N 11, 18, 29
Large	Long rolls roughly downwind	N 37, 51, 57 A 14, 16; C 23

The last column gives reference to pictures in the *International Cloud Atlas*.

Since the rate of change of wind with height is closely proportional to the horizontal gradient of temperature, we may say that the rate of change of wind is large or small according as the isotherms are closely clustered together or far apart. It follows that convectional cloud will be aligned along the isotherms when these are clustered closely together, but at right angles to the isotherms when these are widely separated.

Some attention was devoted by Deslandres [6] to the possibility of explaining the structure of the solar chromosphere as an aggregation of convection cells of the type we have called Bénard cells. This deserves passing mention as an early attempt to use the results of Bénard's small-scale experiments to explain phenomena on a much larger scale. There is, in addition to this, a considerable literature in French dealing with attempts to explain lunar craters by the same mechanism.

Some Further Notes On Convection

Durst, in Part III of a memoir by Giblett and others [7], has interpreted observed variations of wind in the horizontal as due to convective motions such as are shown in Fig. 10, in which the simple convection cell is distorted by a small shear of wind. Durst's results do not admit of brief summarising, and the reader is referred to the original memoir for details of the work.

The effect of entrainment of air by an ascending current, in diminishing the local variation of wind with height has been discussed by Malkus [10] in a paper of considerable interest, in which the slope of an ascending current to the vertical is derived theoretically, and in which it is shown that clouds may move with a speed different from that of their environment. The magnitude of this difference is shown to be a function of the vertical shear and the rate of entrainment of mass into the jet of rising air. If the entrainment of the ambient air is greatest on the upwind side, Dr. Malkus finds that the cloud will tend to grow on the upwind side, and to dissipate on the downwind side.

Malkus, Bunker, and McCasland, in a paper on observational studies of convection [11], have discussed clouds formed by forced convection due to the islands in the neighbourhood of Woods Hole, and have elaborated the discussion of the effects of wind shear previously given by Malkus.

The Theory of Bénard Cells by O. G. Sutton

In a recent paper Sutton [15] has shown that all the critical temperature differences (ΔT), as observed by Chandra and by Dassanayake, should satisfy a relationship:

$$\frac{\Delta T}{T_0} = f(\Delta T^{3/4}),$$

where T_0 is the temperature at the base of the chamber.

In the course of the development of his theory, Sutton shows that, when the curve and line whose equations are

$$y = 8.1 \times 10^{-3}(\Delta T)^{3/4} - 0.0117,$$
$$y = \Delta T/T_0$$

are plotted for any given T_0, their intersections will yield the critical temperature differences. The straight line intersects the curve in two points, is tangential to the curve, or has no intersections. The striking feature is the possible existence of two real roots, the higher root corresponding to the polygonal or Type I convection, and the lower root to the Type II convection, of the type shown in the greater part of Fig. 5.

While the values of ΔT varied from 3.5C to about 90C, the base absolute temperatures varied from 295K to 380K, and the depths of the chamber varied from 0.4 cm to 1.6 cm, Sutton's plot of $\Delta T/T_0$ against $(\Delta T)^{3/4}$ gives an amazingly accurate linear relationship between the plotted variables, the equation for which is

$$\frac{\Delta T}{T_0} = 8.1 \times 10^{-3}(\Delta T)^{3/4} - 0.0117.$$

Sutton's paper is by far the most important theoretical contribution to this subject since Rayleigh's original paper.

Desirable Extensions of the Investigations Mentioned Above

The most desirable investigations would be the collection of details of variation of temperature and wind with height within the types of cloud discussed above, the depth of the cloud layer being carefully measured at the same time.

The essential details of the laboratory phenomena have been collected, but the theoretical discussions referred to fail to lead to such results as are summarised in Figs. 7 and 12, for convection of Type II.

REFERENCES

1. BÉNARD, H., "Les tourbillons cellulaires dans une nappe liquide." *Rev. gén. Sci. pur. appl.*, 11: 1261–1271, 1309–1328 (1900).
2. —— "Sur les tourbillons cellulaires, les tourbillons en bandes, et la théorie de Rayleigh." *Bull. Soc. franç. phys.*, No. 266, pp. 112–115 (1928).
3. BRUNT, D., *Physical and Dynamical Meteorology*. Cambridge, University Press, 1939. (See pp. 219–220)

4. Chandra, K., "Instability of Fluids Heated from Below." *Proc. roy. Soc.*, (A) 164: 231–242 (1938).
5. Dassanayake, D. T. E., *Study of Some Special Cases of Instability of Fluid Layers, Particularly of Such Cases as Have Applications in the Atmosphere.* Unpublished Ph.D. Thesis, University of London, 1937.
6. Deslandres, H., "Tourbillons cellulaires et filaments polaires du soleil." *Ann. Obs. Astr. phys. Paris*, 4: 116–138 (1910).
7. Giblett, M. A., and others, "The Structure of Wind over Level Country." *Geophys. Mem.*, No. 54 (1932).
8. Graham, A., "Shear Patterns in an Unstable Layer of Air." *Phil. Trans. roy. Soc. London*, (A) 232: 285–296 (1933).
9. Jeffreys, H., "Some Cases of Instability in Fluid Motion." *Proc. roy. Soc.*, (A) 118: 195–208 (1928).
10. Malkus, J. S., "Effects of Wind Shear on Some Aspects of Convection." *Trans. Amer. geophys. Un.*, 30: 19–25 (1949).
11. —— Bunker, A. F., and McCasland, K., *Observational Studies of Convection.* Woods Hole Ocean. Instn., Tech. Rep. No. 3, 1949.
12. Pellew, A., and Southwell, R. V., "On Maintained Convective Motion in a Fluid Heated from Below." *Proc. roy. Soc.*, (A) 176: 312–343 (1940).
13. Phillips, A. C., and Walker, G. T., "The Forms of Stratified Clouds." *Quart. J. R. meteor. Soc.*, 58: 23–30 (1932).
14. Rayleigh, Lord (Strutt, J. W.), *Scientific Papers*, Vol. 6, pp. 432–446. Cambridge, University Press, 6 Vols., 1899–1920.
15. Sutton, O. G., "On the Stability of a Fluid Heated from Below." *Proc. roy. Soc.*, (A) 204:297–309 (1950).
16. Terada, T., "Some Experiments on Periodic Columnar Formation of Vortices Caused by Convection." *Rep. aero. Res. Inst. Tokyo*, 3: 1–46 (1928).
17. Thomson, J., "On a Changing Tessellated Structure in Certain Liquids." *Proc. R. phil. Soc. Glasg.*, 13: 464–468 (1882). Also, *Collected Papers in Physics and Engineering.* Cambridge, University Press, 1912. (See p. 136)
18. Weber, E. H., "Mikroskopische Beobachtungen sehr gesetzmässiger Bewegungen, welche die Bildung von Niederschlägenharziger Körper aus Weingeist begleiten." *Ann. Phys., Lpz.*, 94: 447–458 (1855).

RADIOMETEOROLOGY

Radar Storm Observation *by Myron G. H. Ligda* 1265

Theory and Observation of Radar Storm Detection *by Raymond Wexler* 1283

Meteorological Aspects of Propagation Problems *by H. G. Booker* 1290

Sferics *by R. C. Wanta* 1297

RADAR STORM OBSERVATION

By MYRON G. H. LIGDA

Massachusetts Institute of Technology

INTRODUCTION

Through the use of radar for precipitation detection the science of meteorology has acquired an entirely new and unique method of weather observation [14]. As a result of this use the meteorologist has been presented with graphic, dynamic, and up-to-the-second depictions of precipitation formations of all types, and these in several dimensions. Techniques for analysis of radar precipitation echo signals have not yet been completely developed, and for this reason radar is presently of minor (but rapidly increasing) meteorological importance. It appears to have vast potentialities both in the fields of physical meteorological research and weather observation and forecasting, as well as other closely allied activities.

Radar is of interest to the meteorologist in several different ways: (1) in the field of hydrometeor detection and the study of storm structure, (2) in the observation of winds aloft under adverse conditions [35], and (3) in certain regions of the world radars occasionally have their propagation radically modified by atmospheric conditions which meteorologists have been called upon to forecast [18, pp. 11–17]. Each of these subjects may be treated quite independently of the others, although one particular radar system may be involved in all three. For a discussion of the second and third of these subjects the reader is invited to refer to articles elsewhere in this Compendium.[1]

Radar operates on the principle that radio energy is scattered and reflected by the dielectric gradients which exist at the surface of various objects which, when they are detected by radar, are designated as *targets* [49, pp. 1–6]. A *pulse radar*, as distinguished from other types of radar systems, consists of a powerful radio transmitter which emits radio energy in short bursts or pulses, a highly sensitive radio receiver which detects the back-scattered energy, and indicators of various types (called *scopes*) which present the information visually for use by the operator. Radars discussed in this article are of the microwave pulsed type, *microwave* indicating that the operating wave length lies between 1 and 20 cm.

Radar indicates the relative position of a target in terms of polar coordinates: elevation angle, azimuth or bearing from a given direction, and range. Elevation and azimuth angle information is obtained from the orientation of the antenna at the instant of detection. Range determination is accomplished by timing the interval between the transmitted pulse and the received echo.

Subject to certain restrictions, the information presently available to the forecaster or physical meteorologist by means of radar observation is as follows:

1. "Instantaneous" location of all precipitation over several thousand square miles horizontally, and at all altitudes from a *single* observing station.
2. Direction and velocity of precipitation movement.
3. Qualitative information concerning intensity of precipitation.
4. Heights of cloud bases and tops.
5. Approximate height of the freezing level.
6. Information as to whether or not certain storms are thunderstorms.
7. Position, direction, and speed of hurricanes.
8. Upper-level wind data to high altitudes, with great accuracy and under adverse conditions of visibility and precipitation.
9. Wind shear when precipitation is falling through shear surfaces.
10. Turbulence within precipitation.
11. Distribution of fall-velocity for precipitation particles at any level above the radar.
12. Qualitative information regarding water-vapor and temperature distribution in the vertical.

Not all of the information contained in the list above is yet obtained to a useful degree of accuracy. For example, rainfall intensity can be measured only in the immediate vicinity of the radar and even then only approximately. Considerable effort is being made to improve the accuracy of such measurements and to remove as many restrictions and limitations from the above list as is physically possible.

THE BACKGROUND OF RADAR STORM DETECTION

About 1940 it became evident to those working on the development of radar that it was possible to construct radar systems with operating wave lengths of 20 cm (1500 mc sec^{-1}) and shorter. Before these systems were completely operational, Ryde [50] of the General Electric Laboratories Ltd. in England made a study to determine if the performance of such radar systems would be affected by precipitation. This information was necessary in order to determine whether aircraft could escape microwave radar detection by flying in clouds, rain, snow, or dust and sand storms. Ryde's calculations indicated that microwave radars *would* receive echoes from precipitation, and the results he obtained concerning the relative signal strengths of precipitation echoes and aircraft echoes were later

1. Consult "Instruments and Techniques for Meteorological Measurements" by Michael Ference, Jr., pp. 1207–1222; and "Meteorological Aspects of Propagation Problems" by H. G. Booker, pp. 1290–1296.

shown to be quite accurate. About three months after he had completed the essential parts of his calculations, the first radar-detected storm (a thunderstorm from which hail was observed to fall) was reported on February 20, 1941, in England. The operating wave length of the radar used was approximately 10 cm and the storm was followed out to sea to a distance of seven miles. At about the same time a like phenomenon was also observed at the Radiation Laboratories of the Massachusetts Institute of Technology in the United States.

Because of the secrecy enveloping the entire radar program, it was not until many months later that the potentialities of this discovery were realized by meteorologists; but even then, again because of security regulations, few were given detailed information. Consequently, exploitation of this new observational technique was greatly hindered. Also radars were generally inaccessible, those useful for storm-detection purposes were limited in number, and most of these were assigned to very high priority work for aircraft detection.

After World War II, several agencies in different countries initiated research programs in the field of radar storm detection. A large part of the information contained in this article is the result of this research. The rapid progress is due to the wealth of surplus radar and electronics equipment, aircraft, and the intense interest in the subject by persons skilled in the fields of meteorology, mathematics, physics, aviation, and electronics.

THEORY AND FACTORS OF RADAR STORM DETECTION

The theory of radar storm detection defines the process by which precipitation is detected by radar. A large number of factors must be taken into consideration, and determination of the value of some of them is an extremely difficult problem due to their nature or variability. A direct approach to the problem often employed is that of definition of the power received from a given or defined volume of the storm. It may be shown [8; 65, pp. 23–29] that the power received at the antenna due to the echo from precipitation is

$$P_r = \frac{\pi^2}{18(360)^2}\left(\frac{P_t A^2 \phi\theta h}{\lambda^6}\right)(\eta\lambda^4)\left(\frac{k}{R^2}\right), \qquad (1)$$

where P_r = power received (w),
P_t = power transmitted (w),
A = aperture of the parabola (m^2),
ϕ = beam width, vertical (degrees),
θ = beam width, horizontal (degrees),
h = pulse length (m),
λ = wave length (cm),
η = reflectivity of the storm region per unit volume (cm^{-1}),
k = attenuation factor,
R = distance of the storm (km).

The term $(P_t A^2 \phi\theta h/\lambda^6)$ is dependent upon the type of radar; the term $(\eta\lambda^4)$ concerns the nature of the precipitation; and the term (k/R^2) takes into account the effect of intervening space between the radar and the region of the storm for which the power of the echo is to be determined.

Equation (1) may be stated in slightly different form to facilitate computation:

$$P_r = 6.1 \times 10^{-16}\,\frac{P_t G^2 \lambda^2 \phi\theta h k \eta\psi}{R^2}, \qquad (2)$$

where G = gain of the antenna and parabola (pure number), and ψ = fraction of the beam filled by the storm.

Equation (2) states in convenient form all factors entering the problem, and these factors may be inserted in the equation in commonly used units. The new term G may be determined from A and λ. The term ψ is employed to remove the restriction that the beam must be entirely filled by the storm, thus making the equation general.

A clear understanding of the meaning and significance of these various terms is mandatory if correct interpretations are to be made of radar storm presentations. It is well beyond the scope of this article to give this subject the treatment it deserves, but brief discussions concerning each of the terms in equation (1) will be given, with references to enable more thorough study.

Factors Pertaining to the Radar System. *Transmitted and Received Power, P_t and P_r.* It should be readily apparent that the amount of power transmitted will affect the power of the echo signal; the stronger the former, the greater the latter. Maximum power output of a radar system is normally limited by the characteristics of its magnetron (a special type of vacuum tube used to convert direct-current power into radio energy of very high frequency). The great peak power of pulsed radar systems, which may be in terms of megawatts, is explained by the fact that the magnetron is operating only a few tenths of one per cent of the time [49, pp. 344–348]. This is determined by the ratio between the pulse repetition frequency and the duration of the radiated pulse of energy.

The lower limit of power which the receiver can detect is limited by the receiver's sensitivity. In microwave regions this sensitivity limit is not set by atmospheric static, but rather by the electron or thermal "noise" which originates in the various receiver components [49, pp. 28–41]. If the power for the signal which may just be distinguished from "noise" is used in equation (2), the equation may be solved for range, which will then become the *maximum* range at which a given storm may be detected.

Antenna Gain, G, and Aperture of the Parabola, A. The antenna aperture defines the projected area of the parabola (used to focus the radio energy) normal to the axis of the beam (see Fig. 1). Either A or the antenna gain G, whichever is more convenient, may be used in these formulas, since they are related by the semi-empirical formula

$$G = \frac{4\pi A f}{\lambda^2}, \qquad (3)$$

where f is a dimensionless factor which depends on the design and efficiency of the antenna and parabola (usually about 0.6 or 0.7).

FIG. 1.—Antenna parabolas of two of the radars used by the M.I.T. Weather Radar Research. The parabolic reflector to the left forms a conical beam of circular cross section. The one to the right (AN/TPS-10A) forms a "beaver-tail" beam of elliptical cross section 0.7° vertical and 2° horizontal. Both radars have an operating wave length of about 3 cm. (*M.I.T. Weather Radar Research.*)

The antenna gain is a term used to express the increase of power resulting from the focusing of the radiated energy into a narrow beam, in contrast to isotropic radiation [49, pp. 18–21]. Microwave radar operates at such high frequencies in the radio spectrum that the radio energy may be focused by parabolic reflectors in a manner analogous to visible radiation focused by a searchlight. Directivity is desirable from the standpoint of target direction determination; energy conservation is necessary for maximum range of detection. The gain factor is dimensionless and may be loosely regarded as the ratio between the power density in the beam, resulting from focusing, and the power density which would exist at the same range if the transmitter radiated isotropically.

Beam Width, ϕ and θ. The beam width is usually defined as the angle subtended at the antenna between points across the beam where the power density is one-half that along the axis. When using this term it must be understood that an appreciable amount of power is actually present beyond this angle. However, the power beyond the half-power point decreases rapidly with increasing angle, so for most purposes this definition is probably sufficient. If the beam is circular in cross section, it may be depicted as a cone with the apex at the radar antenna.

Cross-section figures of radar beams may be in a variety of shapes, depending upon the use for which the radar is designed [49, pp. 22–28]. Most common shapes are circular and elliptical, especially for radars useful for storm detection purposes. Elliptical sections are employed when high directional accuracy or resolution is desired in one dimension and good coverage in the other. For example, a radar used for height determination would require excellent elevation angle discrimination, but only fair azimuthal resolution. It is neither easy nor desirable to have a beam with no angular spread. It is not easy because the long wave lengths (compared to light) would require prohibitively large parabolas to bring them into exact focus. It is not desirable because the beam would have to be precisely pointed at a target in order for the radar to detect it, if the target were small in size. For storm detection this would be no particular drawback because of the extent of most storm cells. However, the difficulty and expense of construction and control of large-size parabolas rules out beam widths much less than about 1° for radars with wave length of operation of about 3 cm. For a 10-cm wave length radar this angle is at least doubled.

The subjects of beam width and pulse length (which follows) are worthy of careful study by radar meteorologists, for these two factors define the volume of space which is analyzed and strongly influence interpretation of the scope presentations of precipitation. There are distortions resulting from the finite beam width in the nature of azimuth and elevation angle errors, and from the pulse length in the nature of a range error [65, pp. 20–22].

Pulse Length, h. As beam width affects azimuthal and elevation angle resolution, so pulse length affects range resolution. Pulse length may be defined either in terms of time or length; a pulse one microsecond in duration has a wave-train of energy about 300 m long. Two targets in the beam of the radar will be resolved if—and only if—their range separation exceeds one-half the pulse length [65, pp. 17–20]. Extension of the analysis to storm detection results in the corollary that energy from the particles in only one-half the volume illuminated by the pulse can reach the receiver at the same time.

The volume illuminated by the pulse is determined by the pulse length and the linear distance across the beam. The greater the number of scattering particles lying within this volume the stronger will be the echo. By increasing this volume the signal power of the echo may be strengthened, provided this increase in volume does not result in a decrease in the transmitted power density per unit area normal to the beam. However, loss of resolution results both from widening the beam and from lengthening the pulse. Therefore, a designer of radar equipment for storm-detection purposes must balance these factors in order to achieve the best possible presentation of the storm.

Wave Length, λ. The evaluation of a radar as a storm detector depends to a major extent upon the operating wave length. Wave length and frequency are used interchangeably in radar discussion, one being related to the other by the simple formula:

$$f = \frac{c}{\lambda}, \tag{4}$$

where f = frequency (cycles sec^{-1}),
c = velocity of light (cm sec^{-1}),
λ = electromagnetic wave length (cm).

Roughly speaking, a wave length of 10 cm corre-

sponds to a frequency of 3000 mc sec^{-1}. During the war, radar wave length of operation was highly confidential information, since radar countermeasure development by the enemy was considerably facilitated if this was known to him. Accordingly, wave lengths on which radars operated were designated by letters: S-band for 10 cm; X-band for 3 cm; and K-band for about 1 cm. Thus, a radar with an operating wave length of 10 cm was called an "S-band" radar. These designations have persisted since the war, and are utilized in this article to familiarize the reader with terms widely encountered in the literature.

The radar energy back-scattered toward the receiver by a raindrop is inversely proportional to the fourth power of the wave length if Rayleigh scattering [48] applies. However, absorption and attenuation of the scattered energy may become great enough at very short wave lengths to restrict maximum range seriously, and a compromise value of wave length must be used. Speaking in general terms, we may say that for use in areas of very heavy rainfall a wave length of about 10 cm is best, while in regions where the average drop size is somewhat smaller, a wave length of about 3 cm is desirable [15,16]. From an analysis of this problem it seems that a storm-detection radar should be designed for the climate in which it is to be used [68]. Many authorities agree that the best "all-round" frequency for storm-detection radar equipment would probably be about 6000–7000 mc sec^{-1}, corresponding to a wave length near 5.5 cm.

Reflectivity Per Unit Volume of the Storm Region, η. This is a measure of the energy back-scattered by particles in the storm. It is a function of the number of particles per unit volume, their size distribution, the frequency of the energy scattered [65, pp. 30–34], the composition of the particles, and their shape and aspect. The process of electromagnetic scattering by hydrometeors is similar to that of the scattering of visible radiation by large molecules, by which the color of the sky is explained. Fundamental laws of scattering developed by Rayleigh [48] and Mie [45] were applied by Ryde [50] to the special case of electromagnetic scattering and diffraction by raindrops. It was found that a spherical particle which is small relative to the wave length of radiation falling upon it scatters the energy proportionally to the sixth power of its diameter [49, pp. 33–66]. This indicates that if the size of a drop is doubled, its scattering efficiency is increased by a factor of 64. Therefore it is evident that size is of much greater importance than number in the determination of reflectivity. Reflectivity per unit volume turns out to be a function of NR^6 where N is the number of drops and R is their radius. However, when the drops become large compared to the wave length ($2R > \lambda/10$), "Rayleigh scattering" no longer applies and the exact dependence of the scattering cross section upon λ is a complex function.

The composition of the particle affects its dielectric properties, and therefore its efficiency as a scatterer. For example, water has a greater dielectric constant than ice, hence for two spheres of equal diameter, one of water and the other of ice, the water sphere will scatter about five times as much electromagnetic radiation of 3- or 10-cm wave length. Shape and aspect are important for nonspherical particles, therefore the problem of evaluating the power of the echo from snow is difficult.

If the rainfall intensity is known and one assumes a drop-size distribution in accordance with Laws and Parsons' data [38], it is possible to compute the reflectivity per unit volume [6]. It is necessary, however, to assume that this reflectivity remains constant throughout the area covered by the radar beam (or fills a known fraction of it). The fact that reflectivity per unit volume is not a unique function of drop-size distribution greatly complicates the problem of determination of rainfall intensity by radar. At present, measurement of a given level of echo-signal power indicates only that the rainfall at that point lies within certain limits of intensity. This conclusion becomes evident when one realizes that a few large drops can give an echo signal equal to that from many small drops.

An additional factor which should be mentioned in connection with reflectivity is the variability of the back-scattered energy. This energy fluctuates rapidly with time because of interference between the electromagnetic waves scattered by a large number of moving drops. The received power P_r is the *average power* of a number of individual precipitation echo signals [49, pp. 81–85].

Attenuation Factor, k. During passage through the atmosphere between antenna and storm the electromagnetic energy radiated by the transmitting antenna and scattered by the precipitation particles suffers attenuation due to several causes: (1) water-vapor and oxygen absorption, (2) molecular scattering by atmospheric gases, and (3) hydrometeor scattering and absorption [51, 63, 65].

The problem of electromagnetic atmospheric attenuation, being a question not confined to the field of radar, has received the attention of a number of investigators [18]. Tables have been constructed giving attenuation in the free atmosphere and in rain in terms of decibels, as a function of path and wave length [49, pp. 58–62]. Precipitation attenuation is not a serious factor until drop sizes become appreciable in comparison to the wave length (about $\frac{1}{10}$ or larger). Because large drops are frequently present in thunderstorms and tropical rain, X-band (3-cm) radar may sometimes not be the best for observation of storms of these types [13]. Rain attenuation and atmospheric absorption are seldom observed with S-band (10-cm) radars.

Fraction of the Beam Filled by the Storm, ψ. Determination of the magnitude of this factor under certain conditions is difficult or impossible, but its significance cannot be minimized. The echo signal from a weak storm which fills the beam may equal that from a strong storm which only partially fills it, but at present it is not possible to tell from signal characteristics alone what the true condition is. The distances between half-power points at various ranges for beams of different angular widths are listed in Table I; it will be seen that

these distances are considerable even with respect to the heights of thunderstorms.

In the determination of rainfall intensity by radar, it must be assumed that the reflectivity per unit volume is constant in the volume of space contained by the limits of the beam and the pulse length. Because of the beam width of existing radar equipment, this condition can rarely be fulfilled except very close to the radar where the linear distance across the beam is small. Even then, reasonably uniform rainfall conditions must exist.

Table I. Beam Width Between Half-Power Points (in ft)

Range (miles)	Angular Width of Beam		
	0.7°	2°	4°
5	320	920	1,840
25	1,610	4,600	9,210
50	3,230	9,210	18,400
100	6,450	18,400	36,900

Distance of the Storm, R. This term is included in the equation to provide for range attenuation. If the storm completely fills the beam, range attenuation increases with the square of the distance. If the target consists of a point, such as an airplane, range attenuation increases with the fourth power of the range. Most investigators, in making rainfall intensity measurements by radar, endeavor to make their observations as close as possible to the radar for two basic reasons: (1) to insure complete beam filling by relatively uniform rain or snow, and (2) to reduce range attenuation, thus increasing the signal strength of the received echo [6].

OBSERVATIONAL VERIFICATION OF RADAR STORM-DETECTION THEORY

Exact verification of the theory (or of equation (1)) is very difficult because of several factors. The value of P_r must be measured by averaging the power of very weak signals (about 10^{-8} to 10^{-12} w) which are fluctuating rapidly. Moreover, the precipitation echo signal is not received continuously from a given point in the storm but at intervals equal to the transmitted pulse repetition frequency. Techniques have been developed to measure P_r, however, and they will be discussed in the section on signal analysis. A more fundamental difficulty lies in the problem of determining the number and size of the drops within the illuminated portion of the beam in order to calculate the reflectivity. This must necessarily be a sampling process which may be performed either on the surface or aloft [65, pp. 70–76]. Surface sampling, while easier to accomplish, may not be representative of conditions aloft in the atmosphere where the radar beam may be directed.

One greatly desired meteorological use of radar is the determination of rainfall intensity over a wide area, or at various points distant from the radar. It is evident from the theory that the only radar factors closely related to rainfall intensity are the reflectivity per unit volume of the storm and the attenuation [9]. Reflectivity is generally the more useful. Establishing the relationship, however, is a formidable undertaking, requiring assumed drop-size distribution, particle shapes and spatial distributions, as well as power measurements of weak, rapidly fluctuating signals. Several investigators [32, 34, 42] have reported empirical relationships between precipitation intensity and echo-signal power, but only at limited ranges and under certain specific conditions of precipitation.

OBSERVATION OF WEATHER PHENOMENA ON RADAR SCOPES

Types of Radar Scopes. Radar targets are usually presented visually by means of cathode ray tubes or *scopes*. There are about a dozen methods of presentation, depending upon the type and use of the radar, but only four or five of these are of more than casual interest to the meteorologist for storm-detection purposes [49, pp. 160–175]. A brief discussion of these will be given to assist the reader in understanding later descriptions of radar storm displays.

The A Scope. This scope is the simplest and most widely used display; its appearance is illustrated in Fig. 2. From it the operator learns several things about a target: its straight-line distance from the radar, the approximate intensity of its echo signal, and not infrequently its nature. Target echo signals are usually displayed as vertical deflections from a horizontal base line. The amount of vertical deflection is proportional to the power of the echo signal. The distance of the target from the radar is indicated by the relative position of the deflection from one end of the base line, usually the left. There are usually several strong echoes present at close range at all azimuths; these result from the presence of buildings or uneven terrain in the immediate vicinity of the radar. These echo signals are sometimes called "ground clutter."

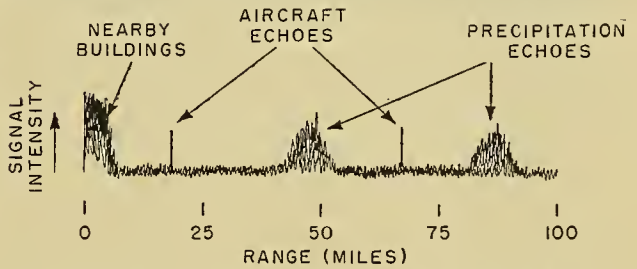


Fig. 2.—Diagram illustrating the appearance of A scope of SCR 615B (S-band) radar, showing characteristics of various types of echo signals. Range information is virtually useless without elevation and azimuth angle information, neither of which is shown on this type of presentation.

The A scope is also useful for target identification. An experienced operator can detect differences in echo-signal characteristics which are sufficient to inform him whether the target is an aircraft, ship, buildings or terrain, or precipitation. Precipitation gives a very distinctive echo signal because of its rapidly fluctuating character which is caused by the changing interference pattern established by the precipitation from pulse to pulse.

The R Scope. The R scope presents the same information as the A scope, but greatly expands the horizontal coordinate so that instead of some 100 miles being represented on the scope, only 5 or 10 miles of any desired portion of the A scope is displayed. Figure 3 illustrates the appearance of this presentation.

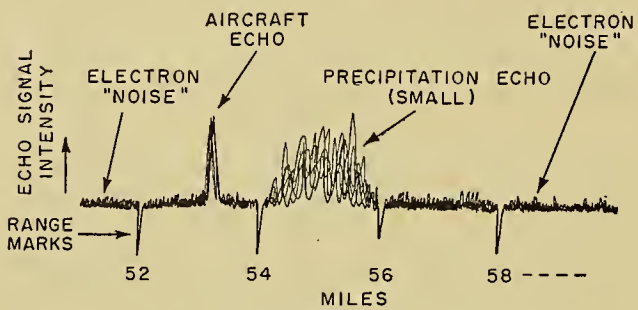


Fig. 3.—Diagram of R scope of SCR 615B radar (modified), illustrating difference in appearance of aircraft and precipitation echo signals.

The Plan Position Indicator or PPI Scope. The PPI scope, as it is popularly known, is one of the most favored scopes for storm detection purposes because of its graphic and easily interpreted display. The presentation is circular and therefore employs the entire area of the cathode ray tube face in contrast to the rather small portion utilized by the A and R scopes. When the radar beam is directed horizontally and rotated in azimuth, the PPI display takes the form of a circular map with the position of the radar represented at the center of the scope. Targets are displayed in position relative to the radar as if viewed from an infinite height directly above the radar. Range from the radar to any target may be determined from the relative distance from its echo signal to the center of the scope. To aid in range determination, an electronic circuit injects marks at proper intervals which correspond to range increments of 1, 10, or 20 miles or any other number and unit of length desired. As the sweep rotates, these marks draw concentric circles about the center of the scope. The appearance of the scope is illustrated in Fig. 4. The electron beam is so controlled that if the radar is not detecting a target at a given position, the face of the scope is not brightened at the corresponding point. When a target is detected, the electron beam is intensified so that it brightens the face of the scope at the proper position in azimuth and range from the radar. This process of target indication is called "intensity modulation" of the electron beam, as contrasted to the beam deflection method employed by A and R scopes, the beams of which are always on and are of constant intensity.

If the antenna is directed horizontally, the indicated range will correspond closely to the actual horizontal distance between target and radar, and its geographical position may be determined by reference to a map of the area. If the map is drawn on transparent material and is of proper scale, it may be laid directly over the face of the scope and geographical positions of targets can be determined directly.

Another version of the PPI scope may be controlled so that the indicated position of the radar may be displaced from the center in any direction to distances of

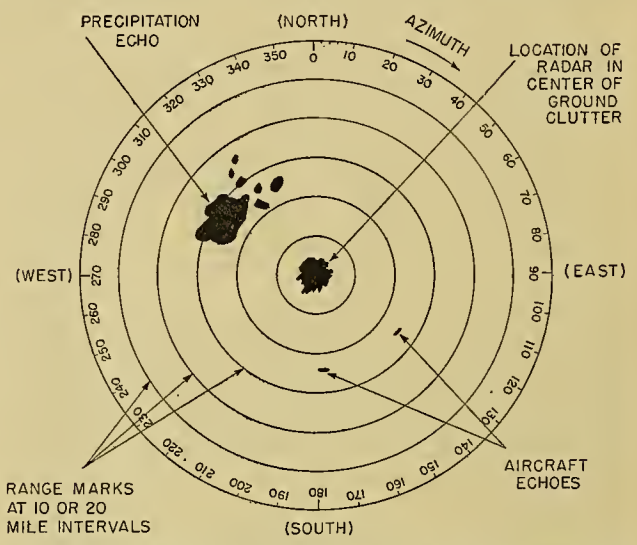


Fig. 4.—Diagram of features of a PPI scope. Some detail has been omitted to clarify the drawing. In all PPI-scope photographs which follow, north is at the top of the scope, east to the right, etc. Elevation angle in all cases is zero degrees.

about two diameters of the scope face. This scope is called an "off-center PPI" and enables close examination of selected areas by virtue of the great sweep expansion which can be employed.

The Range-Height Indicator or RHI Scope. This scope is particularly valuable for storm-detection purposes, but is found only on radars designed to scan a vertical plane. The electron beam is intensity-modulated like that of the PPI scope. The radar antenna scans cyclically from the horizon upwards to any angle, depending upon control settings or the design of the radar. In order for the display on the scope to be intelligible, the azimuth of the beam must remain nearly constant during the scanning process. Targets in the vertical cross section along a given azimuth are displayed on the face of the scope in coordinates of range versus altitude or elevation angle. Because practically all targets, whether storms or aircraft, will be found at altitudes less than 10 miles, and since the horizontal range of the radar is nearly always more than 50 miles, it is usually desirable to expand the vertical scale in order to utilize the maximum available area of the scope face. This results in more precise altitude determination due to the vertical scale increase. For this reason, RHI scopes are usually designed to have an altitude distortion of about 10:1. The appearance of the scope of a typical height-finding radar (AN/TPS-10A) is shown in Fig. 5.

It will be noted that the range and height of a target may be determined directly from this scope by means of properly calibrated scales. In order to fix the target in space, its bearing or azimuth must also be known. This information cannot be obtained from the RHI display alone; consequently it must be read from a dial

or other indicator operating in synchronism with the antenna azimuth control. This is also true for A- and R-scope azimuth readings.

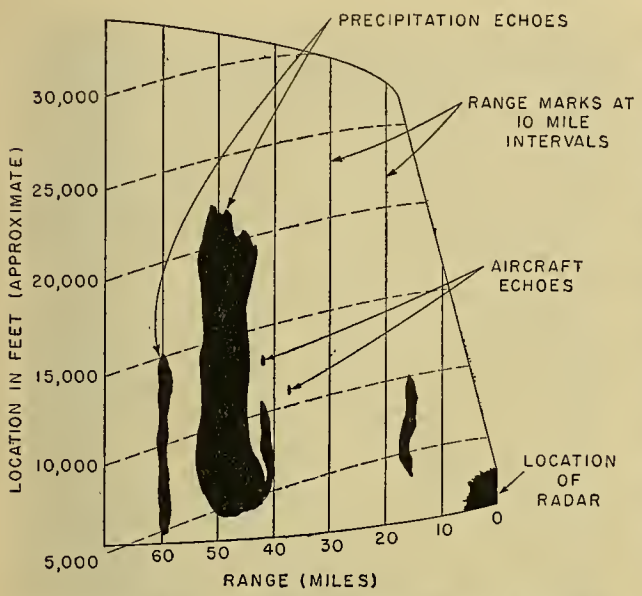


FIG. 5.—Diagram illustrating features of RHI scope of AN/TPS-10A radar (X-band). Photographic images (and this diagram) are reversed from normal presentation because of the design of camera used to obtain RHI-scope pictures which follow. Vertical distortion of scope about 10:1. Diagram illustrates appearance of mature thunderstorm at range of 40 to 55 miles.

Appearance of Scopes During Storms.[2] Precipitation may be divided into various types, using basic causes as the means of classification, as follows:

A. Frontal precipitation.
   1. Cold front.
   2. Warm front.
   3. Occluded front.
B. Orographic precipitation.
C. Hurricanes or typhoons.
D. Instability showers.
   1. Air mass.
   2. Thunderstorms.

This breakdown is convenient for the purposes of this discussion because the horizontal and vertical distribution of hydrometeors is indicative of the motivating cause. This distribution in space can easily be observed with radar.

A. *Frontal Precipitation.* 1. *Cold front.* Perhaps the most striking and easily understood of all radarscope displays is that of echo signals from the squalls associated with an active cold front. Photographs of PPI scopes similar to Fig. 6 have been widely published during the past few years and need little explanation, especially when accompanied by a synoptic map showing position and orientation of the surface front in the

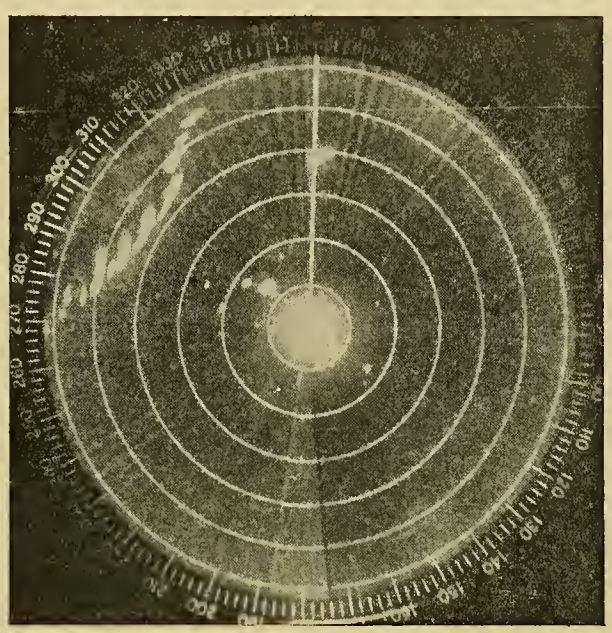


FIG. 6.—Photograph of PPI scope showing echo signals from showers accompanying cold front approaching Boston from the northwest. Isolated warm sector precipitation at azimuth 0°, range 79 miles, and at azimuth 310°, range 20–40 miles. (1730 EST, 3/27/48, S-band radar, range 120 miles, 20-mile markers.) (*M.I.T. Weather Radar Research.*)

conventional manner. Over fairly smooth terrain a well-sited radar with sufficient power will detect the precipitation accompanying an active cold front at distances in excess of 200 miles, depending upon the heights of the squalls.

Close examination of the precipitation pattern shown in Fig. 6 will reveal the distortion caused by finite beam width. This is a photograph of the PPI scope of an SCR 615B-type radar operating on a wave length of 10 cm and a beam width of 3° between half-power points. Notice that the individual cells have circumferential elongation [43]. This is not due to the meteorological situation, but is caused by the comparatively wide beam of this radar. Actually, the cells are usually almost as wide as they are long. This fact can be verified by observation with a radar which has a much narrower beam.

An important restriction on radar storm detection is the limited extent of the front which will pass within the range of this radar. As stated before, during favorable conditions the maximum range of detection may be more than 200 miles, but large portions of the average-sized cyclone will still escape detection by any single radar. This limitation is not due to the lack of power of radar systems, but rather to the curvature of the earth [49, pp. 53–55]. Under standard atmospheric conditions, the height h (in feet) of a horizontally directed beam above the surface of the earth at a range R (in miles) is approximately

$$h = \tfrac{1}{2}R^2, \qquad (6)$$

2. It should be noted that a single radar system cannot incorporate all the characteristics necessary for all types of radar weather observations. For example, the cloud-detection radar cannot be used for rain and snow observation because of severe attenuation of its energy when the size of the hydrometeors becomes appreciable with respect to its wave length of operation. In fact, this radar cannot even be used for cloud detection when the precipitation becomes moderate or heavy.

from which it may be seen that at a range of 200 miles the lowest portion of a horizontally directed beam of a radar located at sea level is about 20,000 ft above sea level [2].

One of the finest practical applications of radar storm detection is made when radar-equipped aircraft use it to avoid violent convective activity while flying through an active cold front [11, 46]. As Fig. 6 shows, the front is far from being a solid mass of storms of moderate altitudes. This application is especially useful at night or when the pilot's vision is restricted by clouds. A recent development has been shown to have value in selection of the least turbulent portions of storms when it is impossible to avoid them completely. This will be discussed in the section concerning echo-signal analysis.

Radar observations of the approach of a cold front show the following sequence of events [67]: Scattered storm-echo signals are first detected at maximum ranges (100–300 miles) depending upon the activity of the front. These echoes are caused by hydrometeors in the upper portions of the tallest cumulonimbus along the front, and generally lie in an arc which may be closely identified with the position of the front as reported by surface observation stations. As the front approaches, the radar detects precipitation at successively lower levels and the original cells appear to increase in size and intensity. When the nearest portion of the front is about 50 miles away, a large part of it appears to be a solid line of precipitation if the antenna elevation angle is kept at or near zero degrees. The inexperienced observer will sometimes conclude that the front is actually intensifying, whereas the radar is simply detecting rain at lower levels which is almost invariably more widespread. This trend continues until the front passes over the radar, at which time echoes from the more distant storms along the front may disappear from the scope entirely because of rain attenuation. At this time the precipitation may appear to be almost evenly distributed around the radar for a distance of many miles.

After frontal passage, the above sequence of events is reversed until the front passes beyond the maximum range of detection or dissipates.

While the line of storms taken as a whole appears to approach the radar, close examination reveals that the individual cells have a component of motion in the direction of the warm air movement ahead of and over the front. Thus, if the front appears to approach from the northwest and the warm air movement is from the southwest, the cells will show a movement just about due eastward, depending on the relative velocities of the cold and warm air masses.

2. *Warm front.* Interpretation of radar displays resulting from warm-front precipitation is considerably more difficult than the interpretation of echoes from cold-front precipitation. This is a result of the larger area covered by the precipitation and the possible variation of conditions. The precipitation causing the echo signals exhibits varying degrees of convective activity depending upon stability and convective stability conditions in the air masses involved. When conditions are stable in both air masses, the return appears as shown in Fig. 7. During more unstable warm-frontal conditions the PPI scope will show stronger echoes from individual storm cells. Absence of uniform or systematic

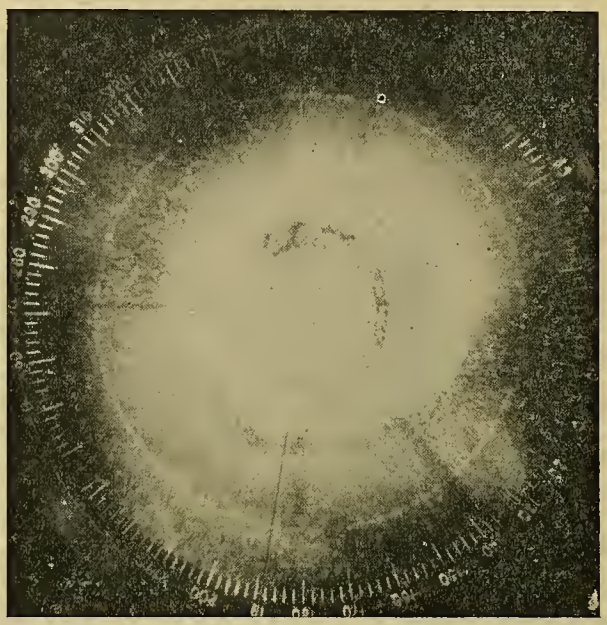


Fig. 7.—Stable warm-frontal precipitation echo signals on a PPI scope. Uniform snow echo signals are present at all azimuths. (1010 EST, 2/20/47, range 20 miles, 5-mile markers.) (*M.I.T. Weather Radar Research.*)

patterns is frequently observed. However, some sort of cellular structure is nearly always observed, and determination of the velocity and direction of cellular movement is usually straightforward. The movement of precipitation cells seems to be controlled by several factors, among which are the circulation within the system and the movement of the entire system itself. Attempts have been made to relate precipitation-cell movements to the winds aloft, but only rather general correlations have been found, and sometimes even these fail completely [29]. It is entirely probable that there is no simple relationship between the winds aloft and cell movement which can be applied to all storms. In order for a given precipitation cell to maintain itself it must have a continuous supply of moisture. This implies that air moves through the system at some speed different from that of the system itself.

At times it has been noted that the precipitation nearest to the radar is rain, while that detected at greater distances must be in the form of snow because of the increase in height of the beam with range (see equation (6)). As yet, no technique has been developed for positive determination of the nature of the precipitation, except in special cases, for example, when the height of the freezing level is known. The appearance of rain and snow echo signals on the PPI scope may be identical when both are present. However, differentiation is sometimes possible by R-scope inspection, as shown in Fig. 11.

The approximate height of the frontal surface may sometimes be determined by the presence of shear con-

ditions as shown by an RHI scope. This spectacular phenomenon is best detected when the precipitation is due to convective causes as was the case when the photograph shown in Fig. 8 was obtained. Where the

FIG. 8.—Photograph of RHI scope showing tall (snow) shower with abrupt shear below the altitude of 9000 ft. Lack of shear in showers more distant from the radar can best be explained by nonuniformity of the wind field. Storms are approaching the radar; shear therefore indicates decreasing wind velocity with altitude (in the plane of the picture) up to the point where echo signal edges become nearly vertical. Actual slope of shower in shear zone about 80° from the vertical (X-band radar). (*M.I.T. Weather Radar Research.*)

precipitation echo signal is vertical, the cell is embedded in air moving with constant velocity and direction. Where the echo signal slopes, the precipitation particles are descending into air moving at a different velocity along the direction of the radar beam. This difference may be due to a velocity differential with height, a direction differential, or a combination of both. The slope of the precipitation is a function of the fall velocities of the particles giving the echo and the wind velocities and directions involved. Unless the fall velocity is known, only relative wind velocities and directions may be calculated from observations of this type [21, 41]. In Fig. 8 it will be noted that in the actual storm the trajectory of the particles was almost horizontal in the shear zone; the 10:1 vertical expansion increases the slope on the scope by this factor.

With increasing use of narrow-beam radars for weather observation, hitherto unsuspected phenomena have been observed in conjunction with warm-front precipitation. Precipitation is occasionally observed to form in narrow, closely spaced parallel bands as shown in Fig. 9. The formation is suggestive of billow clouds formed by shear waves, and is in all likelihood partially due to the same action as that which causes clouds of this type. The phenomena are rather transient, usually persisting for only ten or fifteen minutes in any particular area. Observation of this occurrence usually requires narrow beam width radars (less than 2°) and careful adjustment of antenna elevation angle and gain. Also, rather short pulse lengths must be employed for best presentation. It is interesting to note that perfectly

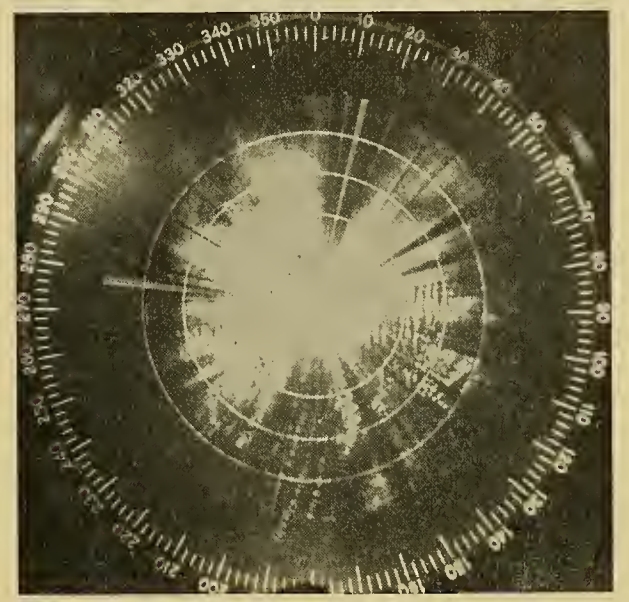


FIG. 9.—Photograph of PPI scope showing striking formation of precipitation echoes formed in parallel bands. Radial spokes are "shadows" cast by buildings and chimneys near the radar. Bands are suggestive of billow clouds, but are actually rain showers oriented north-south. Note the even, light rain echo signal which occupies the northern half of the scope. (1220 EST, 12/7/49, X-band radar, range 120 miles, 10-mile markers.) (*M.I.T. Weather Radar Research.*)

uniform precipitation conditions are seen, though rarely over any extended area. Finer and finer detail becomes evident as the resolving power of radars is increased by employment of narrower beam widths. It may be that reasonably uniform rain can extend only over an area of a few tens or hundreds of square meters.

Another phenomenon observed in conjunction with warm-front precipitation (and also under other conditions), is the descriptively designated "bright band." It receives this name from its appearance on RHI scopes; a typical example is illustrated in Fig. 10. Figure 11 shows its appearance on the R scope of a radar with a vertically directed antenna. It may be seen that the bright band is an approximately horizontal layer of exceptionally strong echo signal. The RHI scope has shown the following conditions to exist during bright-band detection:

1. The band may be the only precipitation return detected by the radar.

2. Precipitation may be seen below the band with no return above (and occasionally vice versa).

3. The bright band may be accompanied by precipitation echo signals both above and beneath it.

While certain detailed processes of bright-band formation are the cause of dispute, all investigators agree that it is in some way connected with the change of state which occurs at the freezing level. The band is invariably observed nearly at or slightly below this

level as determined by radiosonde and aircraft (and sometimes surface) observation [22, 25, 26]. The observed thickness of the band can never be less than the

FIG. 10.—Photograph of RHI scope showing horizontal layer of strong echo signal (bright band) at altitude of about 11,000 ft. Tall vertical showers are indicative of convective activity. (1930 EST, 9/25/47, X-band radar.) (*M.I.T. Weather Radar Research.*)

pulse resolution of the radar used to observe it, and since it has occasionally been determined to be no more than this amount for a given radar system, there is evidence that the thickness of the band may vary from as little as a few tens of meters to several hundred meters in thickness or more. It has been observed in connection with thunderstorms after convective activity has subsided [10, 19]. Two theories have been advanced to explain this phenomenon:

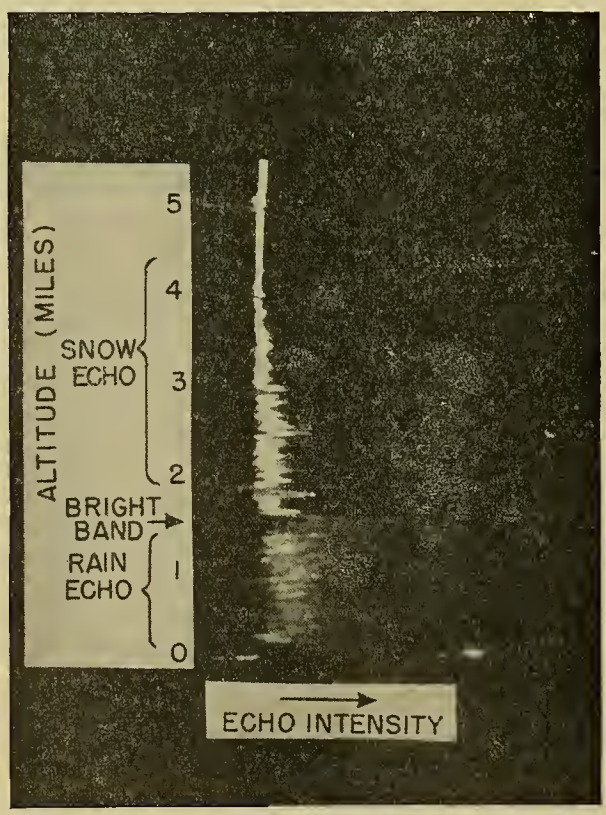


FIG. 11.—Photograph illustrating appearance of R scope when antenna is directed vertically and echoes are received from rain and snow. The weaker echo signal of the snow is due to its greater range and lower dielectric constant. The echo from rain shows much greater intensity variations because of the wider range of fall velocities of the water particles. (X-band radar, range 5 miles, 1-mile markers.) (*M.I.T. Weather Radar Research.*)

1. The region of strong echo is due to drop formation in the colloidally unstable layer of heterogeneous ice-water mixture. Any precipitation detected above that altitude would therefore probably be caused by convective transport [19].

2. Snow particles, too small to give more than a weak echo, fall to the zero degree isotherm, and melt at or just below this level. While melting, they have the low fall velocity of snowflakes, but the high reflectivity of water. Coalescence, often observed near melting temperatures, also serves to increase the reflectivity. After the snowflakes have completely melted, the fall velocity increases and the drops become widely spaced. The result is a region of weaker echo below the level of melting [10, 24, 52].

Most investigators are inclined to favor the latter theory, which the weight of observational evidence seems to support. There is little doubt that, in the absence of other observations, the bright band provides an excellent indication of the approximate height of the freezing level. To a lesser extent, it is also an indicator of the stability of the atmosphere; a thin band is indicative of relative stability, and a thick one (or the absence of a band) suggests convective activity near the freezing level [10]. At the present time, however, this is only a rough qualitative measure.

Under some conditions, the warm front provides the closest approximation to uniform precipitation rates which are so necessary at present for measurements made to study the relationship between rainfall intensity and the power of the echo signal. Efforts are being made to monitor the received power of storm echo signals continuously on two different frequencies [12, 42] (see Fig. 16). The radars are directed and ranged over a recording rain gage to establish relationships between rate of rainfall and echo-signal power. Best correlations between the two have been found during steady rainfall associated with warm-frontal conditions.

3. *Occluded front.* Little has been published concerning radar storm detection during occluded frontal conditions. Because of the complexity of the frontal system few generalizations can be made concerning precipitation types or patterns. Some situations show widespread stratified precipitation, others show just as widely spread convective cells. An excellent place to study fronts of this type would be in Norway or in British Columbia, where radars sited near the coast could be directed westward over the oceans, and observations could be made of fronts unmodified by topography.

B. *Orographic Precipitation.* This type of precipitation, being associated with uneven terrain, is somewhat difficult to observe on PPI scopes if the antenna is directed horizontally. Increase of the antenna elevation angle is necessary to eliminate echo signals from the terrain causing the precipitation. There are no unique

features in this type of precipitation except that, on occasion, it can be fairly uniform. Because of this factor, this type of precipitation might be useful for relating storm echo signals to precipitation intensity.

C. *Hurricanes and Typhoons.* Radar detection of hurricanes and typhoons is even more dramatic than radar detection of cold-front squall lines [66]. The rain-distribution pattern as seen on a PPI scope is illustrated in Fig. 12. It is startling in its similarity to the symbol ϕ

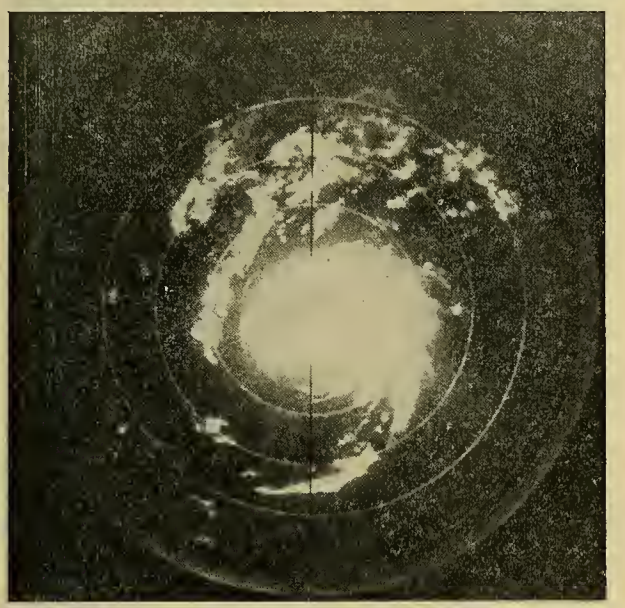


FIG. 12.—Hurricane of 18–21 Sept., 1948. PPI-scope photograph made just prior to passage of the eye over the radar at Key West, Florida. (0900 EST, 9/21/48, range about 125 miles.) (*Official U. S. Navy Photo.*)

which has been chosen to indicate the position of hurricane centers on weather maps.

The spiral rain bands visible in Fig. 12 have been observed in all hurricanes and typhoons detected by radar. These bands move slowly around the center; the cells in the bands move along the bands and into the center in a counterclockwise direction (in the Northern Hemisphere) [37]. As yet, no relationships between the band movement and winds aloft have been established, because of the difficulty of obtaining winds-aloft observations during storms of this type. It is possible that radar will also provide the solution to this problem.

From Fig. 12, it will be noted that a very definite position can be found for the center of rotation as far as the rain is concerned. There is evidence that this center detected by radar does not always coincide with the center of low pressure. Also, a clear "eye" may or may not be present, depending upon the intensity of precipitation at this point.

Radar hurricane detection is probably of greatest value in connection with aircraft hurricane reconnaissance. Early detection of the storm, while it is still far from land and out of range of land-based radars, is considerably facilitated by the use of air-borne radar equipment. It is not necessary for the aircraft to penetrate the storm in order to locate the position of the eye; this often eliminates the necessity of flying under extremely hazardous conditions [57].

By the time a hurricane is within range of land-based radars, its direction and speed are usually well known. However, these radars are useful for precise determination of the storm's position at any instant, and provide very valuable up-to-the-minute data on the storm [58]. It has been found by experience that the best wave length for use in hurricane detection is about 10 cm, at least from the standpoint of rain attenuation. Eventually it may be possible to measure wind speed in different portions of hurricanes by means of radar, and to obtain other clues concerning the structure of these severe storms. The great decrease in loss of life from hurricanes in the southeastern United States is due to highly accurate data concerning movement and velocity of the storms and the early warning of their existence. The importance of the part radar plays in the joint Air Force, Navy, and Weather Bureau hurricane program cannot be minimized.

The importance of radar as a research tool for studying the structure of hurricanes should not be overlooked. The possible use of this new presentation of the storm, in which the rain relative to the wind and pressure distribution may be known, should greatly enhance our understanding of hurricanes. Much work is yet to be done; for example, few, if any, RHI-scope photographs of hurricanes have yet been taken. These would give information concerning convective activity, and whether or not there are shearing forces present as indicated by sloping of the rain cells. It may well be that some of the spiral patterns shown in Fig. 12 are due to the shear and not entirely an indication of the shape of the convective cell.

D. *Instability Showers.* 1. *Air mass.* Air-mass precipitation, being largely convective in nature, generally shows somewhat finer detail in its echo signals on the PPI scope (Fig. 13) than does warm-frontal precipitation. There is an even greater lack of symmetrical arrangement of the convective cells; parallel bands (see Fig. 9) are rarely detected. Shear is rarely observed,

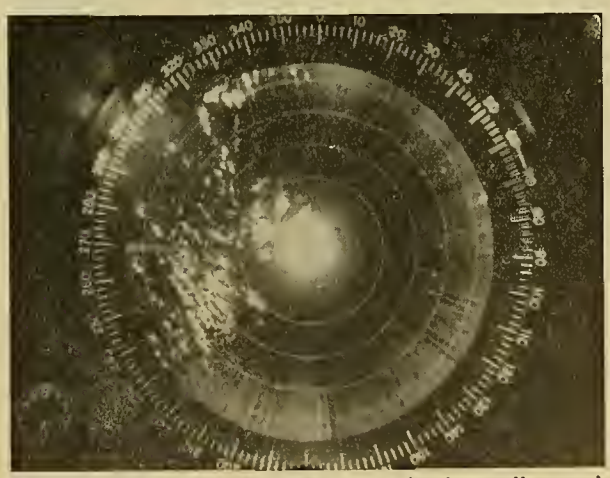


FIG. 13.—Group of air mass precipitation cells moving north-northeast. Note showery appearance in contrast with Fig. 7. Precipitation was in the form of rain showers. Some tendency for band formation is evident. Passage of sharp cold front from west occurred about six hours after this photograph was taken. (1830 EST, 3/8/50, X-band radar, range 60 miles, 10-mile markers.) (*M.I.T. Weather Radar Research.*)

because winds aloft are usually fairly uniform. Determination of the direction and speed of the cells is generally quite simple, especially if the cells are fairly well defined so that their positions may be determined at ten or fifteen minute intervals.

It has been observed in the northeastern United States that warm-sector precipitation, besides being broken into small convective cells, usually consists of rather large, isolated areas. That is, a precipitation area of roughly elliptical shape, perhaps 100 miles in extent, will be broken up into small-size cells of 1–5 miles diameter. The large areas follow each other through the region covered by the radar sweep, with almost no isolated showers occurring between them or on the fringes of the main groups.

2. *Thunderstorms.* The appearance of a thunderstorm echo signal on PPI and RHI scopes is similar to that shown in Fig. 6. These storms, because of their great height, are detectable at greater distances than any other type of precipitation. Well-developed storms have been detected at distances over 300 miles, although for a radar less than 100 ft above the surrounding country, the usual maximum range is nearer 200 miles. Radar played an important role in the extensive research program on thunderstorms in Florida and Ohio [60–62].

Using radar it is possible to determine when a given convective cell reaches the thunderstorm stage. Lightning gives a characteristic echo signal on A and R scopes as shown in Fig. 14 [39]. The echo signal from

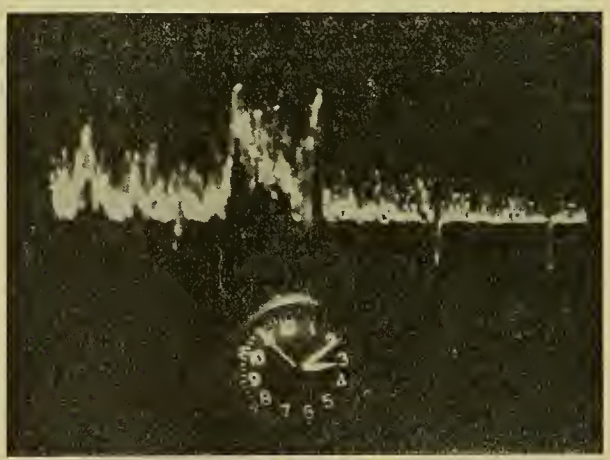


Fig. 14.—Photograph of R scope showing lightning echo signal. Low intensity echo signal from precipitation to right and left of lightning echo signal which persisted for about ¼ sec. (1510 EST, 7/20/49, S-band radar, azimuth 320°, range 50–58 miles, 2-mile markers.) (*M.I.T. Weather Radar Research.*)

lightning is transient, usually lasting less than ½ second, but is easy to detect visually on the scopes. Reflection or scattering evidently takes place from the dielectric gradient established by the stroke, but the mechanism is somewhat uncertain. It should be made clear that it is an actual radar echo signal which is detected, not a radio-static signal emitted by the stroke itself. This is proven by the fact that the lightning indication always occurs on the scope at the range and azimuth of the storm echo signal.

The best presentation of thunderstorm echo signals is found on the RHI scope, as shown in Fig. 15. Again, the reader is cautioned to keep in mind the 10:1 vertical

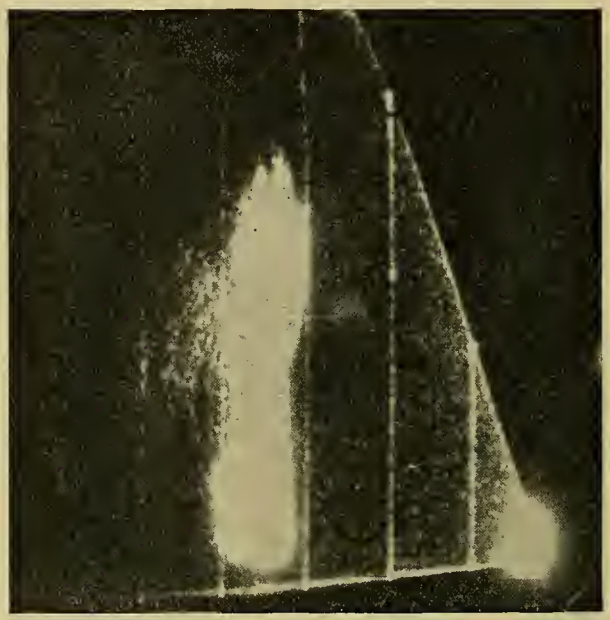


Fig. 15.—Cross section along an azimuth through a thunderstorm. Height of top about 20,000 ft. Note multiple towers at the top, and weak echo signal about 15,000 ft. (1645 EST, 7/31/47, X-band radar, azimuth 350°, range 60 miles, 10-mile markers.) (*M.I.T. Weather Radar Research.*)

expansion of this particular scope. Since this is a fairly representative example of the well-developed mid-latitude thunderstorm, it may be concluded that the precipitation core of such thunderstorms is approximately cylindrical in shape, and is about as wide as it is high. When analyzed by means of accelerated time-scale moving pictures, the RHI-scope echo signals of thunderstorms show discrete volumes of hydrometeors which may move upwards as well as downwards. These probably correspond to the heavy bursts of rain or hail which are occasionally experienced at the surface during intense thunderstorms. Precipitation strong enough to give intense echo signals at appreciable distances from the radar has occasionally been detected at heights exceeding 50,000 ft in mid-latitude thunderstorms. Somewhat rarely the precipitation echo signal will form in the typical anvil shape which the thundercloud itself assumes.

Aircraft have encountered hail in the regions giving intense echo signals, but as yet no reliable means exists for distinguishing the hail from rain or snow by means of radar. A recently developed technique of signal fluctuation analysis, which is discussed in detail under the section on analysis of the storm echo signal, has promise of being a powerful method of detecting the more turbulent regions in thunderstorms.

General Remarks. Because of the comparative recency of applications of radar as a meteorological observation instrument, questions concerning detection of some meteorological phenomena must remain unanswered for the present. Eventually it may be ex-

pected that descriptions of radar detection of such phenomena as tornadoes, waterspouts, dust storms, and other atmospheric anomalies will be available. Forest fires sometimes appear as areas similar to weak rain-echo signals on PPI scopes. The lack of vertical development, coupled with other meteorological information, is usually sufficient to provide correct identification. It is unknown at present whether the temperature gradients or the cinders and ash carried aloft by convective currents cause the echoes.

The small droplet size in fogs and drizzle make these hydrometeors poor targets for radar detection at 3- and 10-cm wave lengths. However, detection by radars operating on wave lengths near 1 cm (K-band) is common. Because of the scarcity of K-band radar, which has not been developed much beyond the laboratory stage, little information is available concerning radar fog and drizzle detection. This equipment may eventually prove valuable for reporting the rate of approach or development of coastal fogs.

The use of radar for determining meteorological conditions in the troposphere has been suggested by Friend [27]. An interesting application of K-band radar to meteorology has recently been developed by the Weather Radar Section of the United States Army Signal Corps Engineering Laboratories [7, 28]. This highly specialized radar system was designed with a vertically directed beam for the purpose of cloud base and top detection. The radar is equipped with an A scope, but in order to provide a permanent record the video signal is fed into a special slow-rate facsimile recorder in such a manner that a vertical time cross section of the clouds is plotted.[3]

Radar provides an ideal means of studying the results produced by seeding of clouds with CO_2, silver iodide, etc. If the seeding is done by aircraft, the position of the plane is known, and if particles of precipitable size form, such vital information as time of formation, height, and position are readily available from the radar scope indications [56].

Another type of echo which has been ascribed to atmospheric conditions [23] has been given the picturesque designation of "angel." "Angels" are most commonly seen with the cloud-detection radar described above and are strong echo signals with no *visible* cause. They occur most frequently in the lower levels of the atmosphere, preferring heights near inversions. They have been ascribed to insects too small to be seen, strong temperature and moisture gradients, and regions of high ionization. Their cause is not yet known.

ANALYSIS OF THE RADAR STORM-ECHO SIGNAL

Besides visual inspection of the scopes, there are other very powerful methods of studying the echo signals resulting from storms. Most of this work is still in the preliminary stages, but enough has been accomplished to warrant description of present techniques.

The Pulse Integrator. Echo-signal intensity may be measured at the radar in at least three different ways: (1) by the intensity of the spot on a PPI or RHI scope; (2) by the amount of vertical deflection on an A or R scope [32]; and (3) by the echo-signal voltage returned to the radar before it is presented on the scopes. The first method must be rejected at present because of fundamental difficulties as well as technical problems. The second method is commonly used, but some doubt exists as to the accuracy of this method and the possibility of observer bias. The third method employs a device designated as the "pulse integrator."

The pulse integrator [70] measures electronically the average strength of the pulsating, fluctuating storm echo signals over a short interval of time. This interval, which determines the number of pulses averaged, may be varied to suit conditions, but from two to four seconds has been found satisfactory. The output of the pulse integrator may be fed into any of several types of visual meters or recorders so that a permanent record is made. The precipitation-echo pulse integrator records are calibrated by feeding signals of known strength into the antenna and comparing the readings. This method of echo-signal intensity measurement is entirely free of observer bias and variability, but results obtained so far have not been satisfactorily correlated with rainfall intensity measurements in an absolute

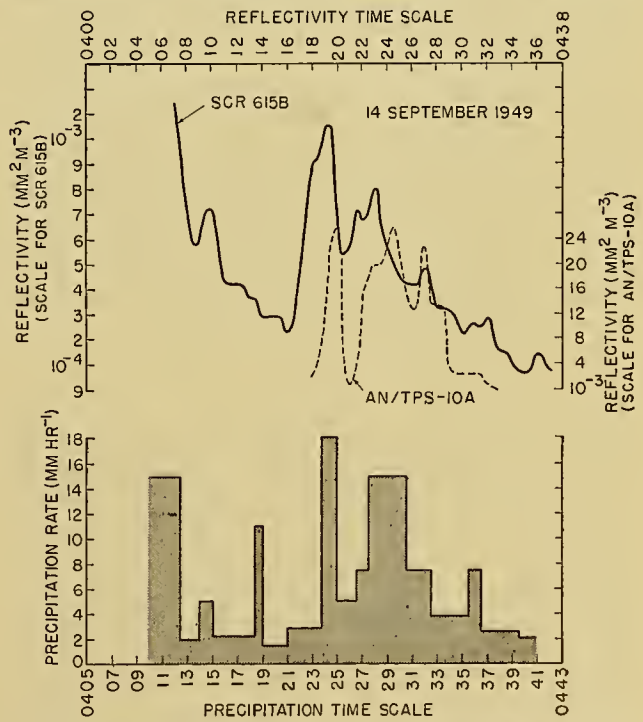


FIG. 16.—Pulse integrator records plotted against precipitation intensity. Echo-signal power records from X-band and S-band radars were converted to units of reflectivity and matched with the precipitation rate as measured with a modified Fergusson rain gage. Note the time-scale shift of about 5 min, used to improve the match between the records. This is explained by the time which elapses before a particularly heavy "burst" of rain, falling from a point well up in the beam, reaches the surface. (*M.I.T. Weather Radar Research.*)

3. For records showing typical results, see Figs. 8 and 9 in the article, "Instruments and Techniques for Meteorological Measurements" by M. Ference, Jr., pp. 1207–1222 in this Compendium.

sense. Fair-to-good *relative* correlations have been found. Figure 16 shows pulse-integrator records of two radars (with operating wave lengths of 3 and 10 cm) compared with a rainfall intensity record obtained from a modified Fergusson rain gage located 12 miles away. The two sets were directed and ranged over the gage. Fluctuations of the signal intensity of both radars show good agreement with variations of the rainfall intensity. The differences between the two radar pulse-integrator curves are due to differences between the radars (frequency, beam width, attenuation, and pulse length), and perhaps to slight errors in orientation and ranging.

Radar Signal Spectrograph. It has been shown both by mathematical analysis and by observation [30; 65, p. 35] that *relative* motions of scattering particles in *range* establish audio and sub-audio frequency fluctuations in the power return of radar storm echo signals. This is caused by the changing phase relationship, or interference patterns, of the return radiation from the separate particles as they move with respect to one another. This relative motion may take place in at least four different ways:

1. Random motion of the scattering particles.
2. Differences in fall velocity of the particles.
3. Small-scale turbulence.
4. Horizontal air streams moving in different directions or with different velocities, or both.

One type of radar signal spectrograph—designated by the abbreviation "rasaph"—is an electronic instrument designed to scan a range of frequencies from 3 to 300 cps and to record on a meter the average power of echo signals at any interval within this range. If the relative motion of the particles is random, the peak signal intensity will lie at zero cycles. As a great deal of random relative motion always exists, the spectrograph trace invariably shows a pronounced peak at the lowest frequencies observable. If large numbers of hydrometeors are moving at some definite velocity with respect to another large group (for example above and below a shear zone), a definite frequency is established in accordance with the following relation:

$$\nu = 2u/\lambda,$$

where ν = fluctuation frequency, u = difference in velocity (along a given azimuth from the radar) between the two groups, and λ = operating wave length of the radar. As is apparent from this relationship, high relative velocities between particles can cause high fluctuation frequencies. However, with pulsed radar there is a limit to the maximum frequency that can be observed; frequencies higher than half the pulse repetition frequency of a particular radar cannot be observed by that radar. (Pulse repetition frequencies usually lie between 200 and 1000 per second.)

Only relative, not absolute, velocities may be determined from the recordings of this instrument. Not enough observations have been made at this writing to enable accurate evaluation of its usefulness, but the following are a few potential uses which it may have:

1. Determination of turbulent conditions in storms.
2. Determination of the distribution of fall velocity for various kinds of hydrometeors. From this, the drop-size distribution may be obtained.
3. Determination of the behavior of freely falling particles as they descend through regions of wind shear.
4. Information concerning the horizontal velocity distribution of hydrometeors may be obtained if wind velocities are known from rawin observations.

Storm Echo Signal Contouring. Early investigators in the field of radar storm detection learned that while no simple relationship between rainfall intensity and echo-signal power existed, regions of more intense precipitation in a given storm could be located with fair accuracy by reduction of receiver "gain" to a point where only the strongest echo signals showed on the PPI scope [36]. By plotting the outline of the storm between successive equal gain reductions, the storm echo signal was reduced to a number of contours; each contour represented a level of equal echo-signal power. In general, it was found that the strongest echo signals, which were loosely associated with more intense rainfall, were near the center of storm-cell echo signals. It was then suggested [5] that the radar be made to do the work of contouring, and to accomplish this in a coarse fashion it was necessary only to block the echo signals of greater than certain power so that they would not appear on the PPI or RHI scopes. Then the area of strongest echo-signal power remained dark in the center of the storm cell presentation. To verify the presence of precipitation in this area, the blocking circuit could be shut off and the scope then regained its normal appearance. Only two levels of signal power could be discriminated by this technique, but this was sufficient for initial tests.

The American Airlines System tested the usefulness of this technique for aircraft storm avoidance, with interesting results [11]. An experimental test aircraft was equipped with an X-band radar modified to produce storm contouring when desired. This plane was deliberately flown into thunderstorms to test the reliability of this equipment for detection of areas of heavy turbulence and precipitation. It was found, at least for the storms penetrated, that areas of heavy turbulence did not necessarily coincide with regions of intense echo signal, as indicated by the contouring. Turbulent regions tended rather to lie where the radar indicated that *steep gradients of rainfall intensity* existed, as shown by narrow, closely packed contours. The contouring proved quite reliable for determining areas of heavy precipitation.

One difficulty with this system is contour distortion by precipitation attenuation, although this is not considered very serious. When X-band (3-cm) radars are used, this is a factor to be considered, because of the heavy rainfall that occurs in most thunderstorms. Additional distortion occurs as the aircraft approaches the storm, because the echo-signal power from the nearest particles increases more rapidly than from those on the far side of the storm, as a result of the range attenuation effect. Studies of these distortions are in progress in

order that the true characteristics of the storm may also be determined from the analysis of the echo signal presentation.

Radarscope Photography. Photography plays an important part in analysis of the storm echo signal because it is the only fully developed, practical method of obtaining accurate and permanent records of the scope displays [40]. Radarscope photography is not difficult; standard films, lenses, apertures, and shutter settings are quite sufficient to produce excellent results. In some respects, photographs of PPI and RHI scopes are more satisfactory for analysis than direct inspection of the scopes, because storm echo signals on all parts of the scope photograph may be studied in relation to each other. On the actual scope only a comparatively small portion is available for study at any instant, unless the antenna is scanning quite rapidly.

Only the sweep trace registers on the negative; the afterglow or persistence of the screen, while visible to the eye for many seconds, does not appreciably affect the film. For a complete PPI photograph the sweep must revolve through 360° during the time the shutter is open in order to enable the entire display to be photographed. With high-speed films (either panchromatic or orthochromatic), aperture openings of the order of f/6.3 are required to obtain photographs with scope-intensity settings comfortable to the unprotected eye.

The most dramatic application of photography to radar storm detection is obtained through cinematographic techniques. Using standard or special motion-picture cameras, one frame of the film is exposed for each successive and complete 360° scan of the antenna. This scanning process may take from 5 to 30 seconds to complete, depending upon the design of the radar. When the film is projected at the normal rate of about 16 frames per second a time-scale contraction of several hundred results. Consequently twenty-four hours of radar storm presentations may be viewed in five or ten minutes.

With this technique the movement and development of precipitation echo signals are beautifully demonstrated. It becomes possible to evaluate accurately such factors as cell life and growth, which may otherwise be studied only by means of laborious and time-consuming plots. This photographic technique has been applied to RHI as well as PPI scopes, with equally interesting results. By this method it has been observed that large numbers of hydrometeors are carried upwards, sometimes with high velocities. This is especially common in the early stages of development of active convective cells.

Another photographic technique which seems to be especially applicable to radar storm observation is photography of PPI and RHI scopes with Polaroid Land Camera film. Using this film, it is possible, without the use of complicated equipment, to have a finished print about one minute after exposure. Tests have shown that these prints are of sufficient contrast to permit accurate measurement of the movement and development of the precipitation-area echo signals. This technique should prove of great assistance to forecasters employing radar for storm observation, since it provides an accurate permanent record of the changes taking place.

Control of Polarization. The energy radiated by the radar is normally linearly polarized either in the horizontal or vertical plane. Only that portion of the received echo energy which is polarized in the same plane is accepted by the receiving antenna. Small raindrops have a unique property which distinguishes them from all other targets, that of perfect rotational symmetry with the line of sight [49, pp. 84–85]. As a result of this symmetry, the intensity and phase of the scattered radiation are not functions of the plane of polarization of the incident radiation.

To make use of this property, the linear polarized energy is altered to circular polarization by reorientation of dielectric slugs inserted in the antenna wave guide, or by "quarter-wave plates" placed in front of the parabolas. The circularly polarized energy back-scattered by spherical raindrops will be reconverted to linear polarization by the plate or slug, but in a direction perpendicular to the original plane of polarization. Therefore it will not be accepted by the radar, and is not presented on the scopes. That portion of the circularly polarized radiation which is scattered by asymmetrical targets will undergo a change of the *sense* of rotation of the vector representing the field of radiation. This back-scattered energy has no special preferred plane of polarization upon reconversion from circular to linear polarization; hence some of the energy will be polarized in the proper plane and will be accepted at the antenna.

Experimental tests have verified the usefulness of the theory outlined above for the detection of ground targets in the presence of rain [1]. As yet, it is not known conclusively whether the technique can be used to differentiate rain from snow particles. It should be pointed out that while the determination of the absolute power of an echo signal is at present quite difficult, measurement of the relative difference in signal strength of two echoes may be accomplished with relative ease and accuracy by means of the pulse integrator. If the back-scattered energy from a region containing both rain and snow is passed through a quarter-wave plate or dielectric slug and the percentage decrease in echo-signal strength of the two regions is unequal, a positive method of identifying the type of hydrometeors by radar is possible.

THE FUTURE OF RADAR STORM DETECTION

A discussion of the future of radar storm detection may logically be divided into three topics: possible future uses for forecasting, possible future equipment development, and possible use for research in physical meteorology. The three are somewhat interdependent, of course.

Use in Forecasting. The value of radar for weather observation can be greatly enhanced by increasing the range through suitable location of the systems. This can be done in two ways: (1) by placing radar sets on

high mountains, and (2) by flying them in aircraft. In either case, the scope presentations would have to be transmitted to weather forecasters; methods for accomplishing this are discussed later. Considering the capabilities of present systems, the entire United States could be covered by about thirty carefully sited ground radars.

With such extended coverage, a weather forecaster would have information concerning the distribution of precipitation in at least the major portions of cyclones. With this information he should be able to observe intensity changes and the speed of the entire system with considerable ease and precision. He should be able to keep a continuous watch on precipitation over a large area and perhaps forecast rainfall totals with accuracy at specified locations. From information gained from previous observations, he should know approximate visibility conditions relative to the precipitation areas. He should be able to determine wind distribution in the lower levels of the troposphere from the movement of the precipitation cells, almost from minute to minute. By means of signal analysis he should be able to differentiate between rain and snow in many regions, and be able to determine zones, kinds, and amounts of icing.

Radar should prove useful as a method of verifying forecasts; it should also be of assistance in the preparation of long-range precipitation forecasts for specific points by providing more information on the effect of topography on precipitation patterns. In addition, from the standpoint of icing and turbulence, it should be of great help in the selection of best flight paths through unavoidable storm areas.

Activities other than aviation should have a more than casual interest in the information which radar can yield. For example, electric power companies, by observing the progress of thunderstorms, should be able to anticipate power-demand peaks in various communities in the path of storms and adjust line loads accordingly. Also, accurate short-range forecasts for various sports and outdoor activities which may be affected by showers are already available in certain areas, and it may be expected that demands for this service will increase in the future.

Because of the ease of radarscope interpretation in location of the position of precipitation areas, television may be an excellent medium by which to convey the image of the PPI scope to the citizen for his own interpretation. Radarscope movies have already been shown several times on television programs in the Boston area.

A possibly valuable application of radar storm detection exists in the field of water conservation, flood control, and irrigation [20, 54, 56]. If rainfall intensity can be even approximately correlated with the intensity of the PPI storm echo image, it should be possible to measure with useful accuracy the total rainfall caused by a storm in a given area within range of the radar. Critical catch basins could be studied by masking all other portions of the PPI scope and integrating the light emitted with a photoelectric cell. If developed, this technique might also be applied to measure the amount of snowfall in mountainous regions, thereby eliminating the need for snow-survey teams in many areas.

Equipment Development. The usefulness of radar for storm detection can be greatly amplified by improving equipment as well as observational techniques. Many instruments and accessories developed for special applications have not yet been tested for their utility in radar storm detection work. It behooves the radar meteorologist to be alert for new developments in such fields as instrumentation, radar, television, communications, and photography, as well as meteorology. Also, he should be ready to suggest improvements or modifications which will increase the utility of radar for meteorological purposes. For example, following are two suggested modifications which would probably facilitate the observation of precipitation echo signals:

Horizon-to-Horizon RHI Scope. This scope, used in conjunction with an antenna which scans from horizon to horizon through the zenith, would greatly improve the cross-section display presently available to the observer. This type of scan, as well as increasing the scanned volume by 100 per cent, would appreciably decrease the time needed for observations in all directions.

Gated Range-Height Indicator. This scope would remove the necessity for vertical expansion of the RHI sweep by limiting the range presented on the scope to 15 or 20 miles. The maximum indicated altitude could be adjustable from 15,000 to 50,000 ft. The scope could be controlled in range to present any 15- to 20-mile range increment from 0 to 100 miles. In this manner the true proportion of range to height would be preserved while the same accuracy would be available for altitude measurements. There is no serious objection to limiting the range in this manner, for experience has shown that only rather limited portions of the RHI scope are of interest at any one instant.

A certain type of radar receiver which was designed for nonmeteorological purposes might be invaluable for radar storm detection [49, pp. 553–554]. This receiver, instead of giving linear amplification of the echo signal, amplifies logarithmically, the echo signals receiving less and less amplification as they increase in power. By use of these "linlog" receivers, or others of similar characteristics, a far greater range of echo-signal power lies within the limits of scope presentation for one particular gain setting. This type of amplification might be useful in the presentation of storm echo signals which may have a vast range of power at any given time. Over a short distance or period of time storm echo signals may cover a range in excess of 60 decibels, from minimum detectable signal to signals which "saturate" the receiver.

At present, one of the main restrictions on the use of radar for storm-detection information by weather forecasters is the lack of suitable means of disseminating the scope information to all persons interested in it. There are several ways in which this could be accomplished [49, pp. 726–735].

1. Verbal description, or code.
2. Television of the PPI scope.

3. Facsimile transmission of the PPI-scope display.

4. Remote scopes, distant from the radar, which get the video information either by wire or radio signals.

The first method is used at present and will continue to be the most practical for a while [59]. The fourth method has been used for remote presentation of scope information for military purposes; but it is expensive and somewhat limited in range [71]. Television is out of the question for the present and the immediate future because of lack of facilities.

Facsimile transmission of PPI photographs presents intriguing possibilities: first, facilities for facsimile transmission are presently available at many military and civilian weather stations; second, the facsimile record presents the recipient with a permanent picture of the scope (which television and remote scopes do not); and third, facilities for facsimile transmission between ground and aircraft have been nearly perfected. The main difficulty at present is getting a PPI picture to transmit, although this problem is practically solved in several ways. One way is by use of the fast development techniques (Polaroid Land Camera). A second way is by use of PPI tubes with very slow decay time phosphors. A third way is through the use of fast special papers such as Alphax.[4] Other more elegant and elaborate methods have also been suggested.

Research in Physical Meteorology. It is anticipated that radar will provide useful information concerning the structure and behavior of that portion of the atmosphere which is not covered by either micro- or synoptic-meteorological studies. We have already observed with radar that precipitation formations which are undoubtedly of significance occur on a scale too gross to be observed from a single station, yet too small to appear even on sectional synoptic charts. Phenomena of this size might well be designated as *mesometeorological.*

In addition to supplying observations in this mesometeorological region, radar is also expected to be of assistance to the physical meteorologist in his studies of the mechanism of precipitation; of the size and number distribution of hydrometeors; and of the behavior of hydrometeors under various conditions of turbulence. It may also provide him with clues concerning the processes of waterdrop and snowflake growth, coalescence, and evaporation.

Sincere and deep-felt appreciation is expressed to members of the M. I. T. Weather Radar Research Project; their helpful comments, corrections, and suggestions were of invaluable assistance in the preparation of this article. For permission to use photographs, the author is indebted to the U. S. Navy Department and to the M. I. T. Weather Radar Research Project operating under Signal Corps Contract W36-039-sc-32038.

4. Manufactured by Alphax Paper and Engineering Co., Brockton, Mass.

REFERENCES

Publications marked with an asterisk are considered to be especially comprehensive, or to contain excellent bibliographies, or both.

1. American Airlines System. Flight Test Rep. No. 64. New York, 1949.
2. Army Air Forces, Headquarters. AAF Manual 101-66-1. Washington, D. C., 1945.
3. —— *Radar Weather Reconnaissance.* AAF Manual 105-101-1. Washington, D. C., May 1945.
4. —— *Radar Storm Detection.* AAF Manual 105-101-2. Washington, D. C., 1945.
5. Atlas, D., *Preliminary Report on New Techniques in Quantitative Radar Analysis of Thunderstorms.* Rep. AWNW-7-4, Pt. I, Dayton, Air Materiel Command, 1947.
6. Austin, P. M., *A Group of Graphs Showing Estimated Radar Return from Precipitation.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 6, 1948.
7. —— *On the Probability of Detecting Bases and Tops of Clouds by Radar at K-Band or at Shorter Wave Lengths.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 2, Oct. 1, 1947.
8. ——*Note on Comparison of Ranges of Radio Set SCR 615B and Radar Set AN/TPS-10A for Storm Detection.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 5, 1947.
9. —— "Measurement of Approximate Rain Drop Size by Microwave Attenuation." *J. Meteor.*, 4: 121–124 (1947).
10. —— and Bemis, A. C., "A Quantitative Study of the 'Bright Band' in Radar Precipitation Echoes." *J. Meteor.*, 7: 145–151 (1950).
11. Ayer, R. W., White, F. C., and Armstrong, L. W., *The Development of an Airborne Radar Method of Avoiding Severe Turbulence and Heavy Precipitation in the Precipitation Areas of Thunderstorms and Squall Lines.* New York, American Airlines System, 1949.
12. Bemis, A. C., "Weather Radar Research at M.I.T." *Bull. Amer. meteor. Soc.*, 28: 115–117 (1947).
13. Bent, A. E., *Echoes from Tropical Rain on X-Band Radar.* Rep. 728, Cambridge, M.I.T. Radiation Lab., 1945.

*14. —— "Radar Detection of Precipitation." *J. Meteor.*, 3: 78–84 (1946).

15. —— *Climate in Relation to Microwave Radar Propagation in Panama.* Rep. 476, Cambridge, M.I.T. Radiation Lab., 1944.
16. —— *Radar Echoes from Precipitation Layers.* Rep. 689, Cambridge, M.I.T. Radiation Lab., 1945.
17. Brooks, H. R., "A Summary of Some Radar Thunderstorm Observations." *Bull. Amer. meteor. Soc.*, 27: 557–563 (1946).
18. Burrows, C. R. (chairman) and Attwood, S. S. (editor), *Radio Wave Propagation.* New York, Academic Press, 1949. (Consolidated Summary Tech. Rep. of Comm. on Propagation of the Nat. Res. Comm.)
19. Byers, H. R., and Coons, R. D., "The Bright Band in Radar Cloud Echoes and Its Probable Explanation." *J. Meteor.*, 4: 75–81 (1947).
20. Byers, H. R., and Collaborators, "The Use of Radar in Determining the Amount of Rain Falling over a Small Area." *Trans. Amer. geophys. Un.*, 29: 187–196 (1948).
21. Byers, H. R., and Battan, L. J., "Some Effects of Vertical Wind Shear on Thunderstorm Structure." *Bull. Amer. meteor. Soc.*, 30: 168–175 (1949).
22. Canadian Army Operational Research Group, *Investigations of the "Bright Band" Discussed in CAORG Report No. 30.* Rep. No. 42, Nov. 16, 1945.
23. Crawford, A. B., "Radar Reflections in the Lower Atmosphere." *Proc. Inst. Radio Engrs.*, 37: 404–405 (1949).

24. Cunningham, R. M., "A Different Explanation of the 'Bright Line.' " *J. Meteor.*, 4: 163 (1947).
25. —— and Miller, R. W., *Five Weather Flights: Measurements and Analysis.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 7, 1948.
26. Eon, L. G., and Tibbles, L. G., *Investigations of the "Bright Band."* Rep. No. 30, Canadian Army Operational Research Group, 1945.
27. Friend, A. W., "Theory and Practice of Tropospheric Sounding by Radar." *Proc. Inst. Radio Engrs., N. Y.*, 37: 116–138 (1949).
28. Gould, W. B., *Cloud Detection by Radar.* Paper presented at Joint Meeting of Amer. Geophys. Un. and Amer. Meteor. Soc., Washington, D. C., April 20, 1949.
29. Hare, F. K., *Movement of Precipitation Areas.* Rep. No. 48, Canadian Army Operational Research Group, 1947.
30. Hilst, G. R., *Analysis of the Audio-Frequency Fluctuations in Radar Storm Echoes: A Key to the Relative Velocities of the Precipitation Particles.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 9A, 1949.
31. —— and MacDowell, G. P., "Radar Measurements of the Initial Growth of Thunderstorm Precipitation Cells." *Bull. Amer. meteor. Soc.*, 31: 95–99 (1950).
32. Hooper, J. E. N., and Kippax, A. A., "Radar Echoes from Meteorological Precipitation." *Proc. Instn. elect. Engrs.*, Pt. 1, 97: 89–95 (1950).
*33. Hooper, J. E. N., and Jones, F. E., *Interim Report: Measurements of Radar Echo Intensities from Rain and Snow.* Rep. No. T 2082, London Telecommunications Research Establishment, 1947.
*34. Jones, F. E., "Radar as an Aid to the Study of the Atmosphere." *J. R. aero. Soc.*, 53: 437–448 (1949).
35. Kirkman, R. A., and LeBedda, J. M., "Meteorological Radio Direction Finding for Measurement of Upper Winds." *J. Meteor.*, 5: 28–37 (1948).
*36. Langille, R. C., Gunn, K. L. S., and Palmer, W. McK., *Quantitative Analysis of Vertical Structure in Precipitation.* Ottawa, Defense Research Board, 1948.
37. Latour, M. H., and Bunting, D. C., "Radar Observation of Florida Hurricane, August 26–27, 1949." Gainesville, Univ. of Florida, College of Engineering, *Bull. Ser.*, Vol. 3, No. 8 (1949).
38. Laws, J. O., and Parsons, D. A., "The Relation of Raindrop-Size to Intensity." *Trans. Amer. geophys. Un.*, 24: 452–459 (1943).
39. Ligda, M. G. H., "Lightning Detection by Radar." *Bull. Amer. meteor. Soc.*, 31: 279–283 (1950).
40. —— *Radar Scope Photography.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 4, 1947.
41. MacDowell, G. P., *A Study of the Slopes of Radar Precipitation Echoes.* Unpublished Thesis, M.I.T. Dept. of Meteor., Cambridge, Mass., 1948.
42. Marshall, J. S., Langille, R. C., and Palmer, W. McK., "Measurement of Rainfall by Radar." *J. Meteor.*, 4: 186–192 (1947).
43. Mather, J. R., "An Investigation of the Dimensions of Precipitation Echoes by Radar." *Bull. Amer. meteor. Soc.*, 30: 271–277 (1949).
44. Maynard, R. H., "Radar and Weather." *J. Meteor.*, 2: 214–226 (1945).
45. Mie, G., "Beiträge zur Optik trüber Medien." *Ann. Physik*, 25: 377–445 (1908).
46. Miller, R. W., "The Use of Airborne Navigational and Bombing Radars for Weather Radar Operations and Verifications." *Bull. Amer. meteor. Soc.*, 28: 19–28 (1947).
47. Perrie, D. W., "The Rain Required for a Radar Echo." *Bull. Amer. meteor. Soc.*, 30: 278–281 (1949).
48. Rayleigh, Lord, "On the Light from the Sky, Its Polarization and Colour." *Phil. Mag.*, 41: 107–120, 274–279 (1871).
*49. Ridenour, L. N., *Radar System Engineering.* New York, McGraw, 1947.
*50. Ryde, J. W., *Echo Intensities and Attenuation Due to Clouds, Rain, Hail, Sand and Dust Storms.* Rep. No. 7831, Wembley (Eng.), General Electric Co., 1941.
51. —— "The Attenuation of Centimetre Radio Waves and Echo Intensities Resulting from Atmospheric Phenomena." *J. Instn. elect. Engrs.*, 93 (3A): 101–103 (1946).
*52. ——"The Attenuation and Radar Echoes Produced at Centimetre Wavelengths by Various Meteorological Phenomena" in *Factors in Microwave Propagation.* Royal Phys. Soc., Report of Conference, 1946. (See pp. 169–188)
53. Siegert, A. J. F., *On the Appearance of the "A-Scope" When the Pulse Travels Through a Homogeneous Distribution of Scatterers.* Rep. 466, Cambridge, M.I.T. Radiation Lab., 1943.
54. Smith, E. D., and Fletcher, R. D., "A Summary of the Uses of Radar in Meteorology." *Trans. Amer. geophys. Un.*, 28: 713–714 (1947).
55. Spilhaus, A. F., "Drop Size, Intensity, and Radar Echo of Rain." *J. Meteor.*, 5: 161–164 (1948).
56. Stout, G. E., and Huff, F. A., *Radar and Rainfall.* Urbana, Illinois State Water Survey Div., 1949.
57. U. S. Navy, *A Note on the Double Eye Phenomenon as Observed During the Hurricane, 11–20 September, 1947.* NAVAER 50-IR-207, Washington, D. C., Aerology, Flight Section, C.N.O., 1947.
58. —— *Report of Hurricane at Key West 18–21 September, 1948.* NAVAER 50-45T-11, Washington, D. C., Aerology, Flight Section, C.N.O., 1948.
59. U. S. Weather Bureau, *Manual of Radar Meteorological Observations.* Washington, D. C., Aug. 1, 1949.
60. —— *A Report on Thunderstorm Conditions Affecting Flight Operations.* Washington, D. C., Thunderstorm Project, Tech. Paper No. 7, 1948.
61. —— *Operation and Activity of the Thunderstorm Project to September 20, 1946.* Chicago, Ill., Thunderstorm Project, Rep. No. 1, Nov. 1946.
62. —— *The Thunderstorm.* Washington, D. C., U. S. Govt. Printing Office, 1949.
63. Van Vleck, J. H., "Absorption of Microwaves by Uncondensed Water Vapor." *Phys. Rev.*, 71: 425–433 (1947).
64. —— "Absorption of Microwaves by Oxygen." *Phys. Rev.*, 71: 413–424 (1947).
*65. Weather Radar Research, *First Technical Report under Signal Corps Contract.* Cambridge, M.I.T. Weather Radar Research, 1946.
66. Wexler, H., "Structure of Hurricanes as Determined by Radar." *Ann. N. Y. Acad. Sci.*, 48: 821–844 (1947).
67. Wexler, R., "Radar Detection of a Frontal Storm, 18 June, 1946." *J. Meteor.*, 4: 38–44 (1947).
68. —— *Optimum Wavelength for Storm Detection Through Rain.* Belmar, N. J., Evans Signal Laboratory, 1946.
69. —— and Swingle, D. M., "Radar Storm Detection." *Bull. Amer. meteor. Soc.*, 28: 159–167 (1947).
70. Williams, E. L., *The Pulse Integrator, Part A: Description of the Instrument and Its Circuitry.* Cambridge, M.I.T. Weather Radar Research Tech. Rep. No. 8A, 1950.
71. Zurcher, L. A., *Storm Detection Radar.* Belmar, N. J., Evans Signal Laboratory, 1949.

THEORY AND OBSERVATION OF RADAR STORM DETECTION

By RAYMOND WEXLER

Imperial College of Science and Technology

Theory of Storm Detection

The basic problem in radar storm detection is the absorption and scattering of a plane electromagnetic wave by a sphere. This problem was first studied by Mie [8] who analyzed the absorption and scattering of light from gold particles suspended in a liquid. Application to the theory of the detection of precipitation was first made by Ryde [11]. The Canadian Army Operational Research Group showed that the theory was valid, at least in respect to order of magnitude, by measurement at the ground of the drop size in rainstorms observed by radar. More exact laboratory verification is presently being undertaken at Harvard University.

The equation for detection of a particle of scattering cross section σ at distance R by electromagnetic energy propagated isotropically from its source is given by

$$\frac{P_r}{A} = \left(\frac{P}{4\pi R^2}\right)\left(\frac{\sigma}{4\pi R^2}\right), \tag{1}$$

where P_r is the power received by the radar and P is the peak power emitted by the radar antenna of effective area A. If the radiation is directional, a gain factor must be included in the numerator of the right-hand side of the equation. This gain is derived from the fact that the radiation is propagated as a beam from the antenna and is G times the radiation from an isotropic source. Because of the fact that the beam is normally completely intercepted by the cloud at ranges under 75 miles, the directional aspect need not be considered.[1] The term $P/4\pi R^2$ will be recognized as the power density of the propagated radiation, and P_r/A the power density of radiation received at the antenna. The radar cross section σ is defined by the equation. It is the area normal to the radiation such that if all the power incident on it were scattered isotropically it would give, at the receiver, the same power as the observed echo.

For complete interception of the beam by the cloud, the equation must summate the total number of scatterers in the volume of air illuminated by the pulse. This volume V may be considered to be a spherical shell at radius R of half a pulse-length thickness:

$$V = 4\pi R^2 h/2, \tag{2}$$

where h is the pulse length of the radar. The reason for the factor 1/2 is that energy is emitted from a time $t = 0$ to $t = h/c$, the pulse duration, where c is the velocity of light. In that time the energy must make a round-trip distance $2R$. The maximum minus the minimum distance illuminated by one pulse is then

$$2(R + \Delta R) - 2R = (t + h/c)c - tc,$$

whence the distance illuminated by the pulse is $\Delta R = h/2$.

If within a unit volume there are n_i scatterers of cross section σ_i all located at random within the beam, then the equation becomes, by combining (1) and (2),

$$P_r = \frac{PAh}{8\pi R^2} \sum_i n_i \sigma_i. \tag{3}$$

If the beam is not completely intercepted by the cloud, the percentage of the beam so intercepted must be inserted on the right-hand side of equation (3).

The equation holds for scatterers randomly located within the beam on the assumption that the distances between drops are large compared to the diameter of the drops and the wave length of radiation. If there were forces causing the drops to fall coherently in pairs, the power received would be twice that indicated by equation (3). However, the assumption of random positions of the drops in a rain cloud appears to be justified.

The equation for storm detection including the effect of attenuation is

$$P_r = \frac{PAh}{8\pi R^2} \sum_i n_i \sigma_i 10^{-0.2KR}, \tag{4}$$

where K is the attenuation in decibels per unit distance. If P_r is the minimum detectable signal of the radar, a figure known for most sets, the distance R in (4) represents the maximum range of detection for the target $\sum_i n_i \sigma_i$.

From Mie's study of the diffraction of electromagnetic waves by a sphere, it is possible to determine the radar cross section in the form of an infinite series of spherical Bessel functions of parameter $\pi m D/\lambda$, where m is the complex index of refraction for the precipitation at a given wave length λ, and D is the diameter of the drop. For values of $D/\lambda \leq 0.05$, the Rayleigh law of scattering is approximately valid:

$$\sigma_i = \pi^5 \left(\frac{\epsilon - 1}{\epsilon + 2}\right)^2 \frac{D_i^6}{\lambda^4}, \tag{5}$$

where ϵ, the dielectric constant, is equal to the square of the index of refraction. At $D/\lambda = 0.08$, the value of σ is about 0.8 that of Rayleigh; it then rises to a value twice that of Rayleigh at $D/\lambda = 0.2$, beyond which it undergoes rapid and large fluctuations about a constant value near $\pi D^2/4$, the radar cross section of a large conducting sphere. This behavior varies somewhat according to the wave length, but qualitatively it char-

1. For a derivation using the directional aspect see, for example, Wexler and Swingle [16].

acterizes the scattering function in the microwave region. Figure 1 shows the ratio of true scattering to Rayleigh scattering for 1.25 cm and 3 cm as computed

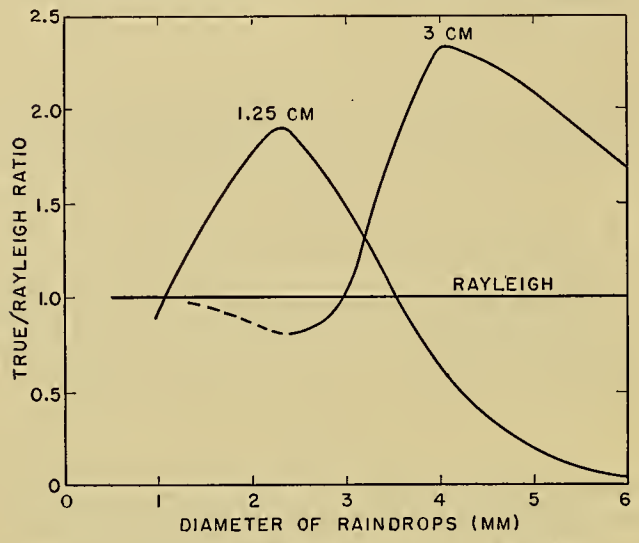


Fig. 1.—Ratio of true scattering to Rayleigh scattering.

by Haddock (unpublished) from calculations made by the National Bureau of Standards. For a wave length of 10 cm, Rayleigh scattering holds almost exactly.

Table I. Radar Cross Section (σ) for Various Rain Intensities
(in 10^{-8} cm^2 per cubic centimeter)

Rainfall intensity (mm hr^{-1})	λ = 1.25 cm		λ = 3 cm	
	Ryde	Haddock	Ryde	Haddock
0.25	0.534	0.570	0.0115	0.0143
1.25	6.58	6.50	0.116	0.156
2.5	18.2	19.5	0.337	0.459
12.5	162	176	4.57	5.59
25	390	415	14.5	16.8
50	901	968	46.2	49.1
100	2000	1920	148	139
150	3130	2890	289	238

Table I gives the radar cross section for different rain intensities based on the mean drop-size distribution of Laws and Parsons [6], as computed separately by Ryde and Haddock. In view of the complexity of the computations, the disagreement is not surprising.

Refraction

In passing through the atmosphere the radio waves are subject to refraction and thereby travel a curved path. The curvature of the ray is stronger the greater the gradient of the refractive index perpendicular to its path (the ray tends to curve towards higher refractive index). In a standard atmosphere of 60 per cent relative humidity the curvature of the ray in the lowest few kilometers of the atmosphere is about one-quarter the curvature of the earth. In tropical regions the curvature is greater; in Florida during summer it averages about 40 per cent of the earth's curvature.

In regions where the vapor pressure decreases rapidly with elevation through a temperature inversion, such as over the trade wind regions of the subtropical high, the curvature of the ray is greater than that of the earth. In such cases the rays, initially nearly horizontal with the earth's surface, may be said to be "trapped" within a narrow layer along the earth and the radar detection of objects near the surface may extend to abnormally long ranges. Generally, however, there is no precipitation on such occasions and, as a result, this "anomalous propagation" is of little importance for storm detection. Occasionally during such conditions, echoes from objects on the surface are detected by the radar and may be mistaken for a rain shower. By careful examination of the A scope,[2] the experienced observer can easily recognize the characteristic fluctuating signal of a rainstorm.

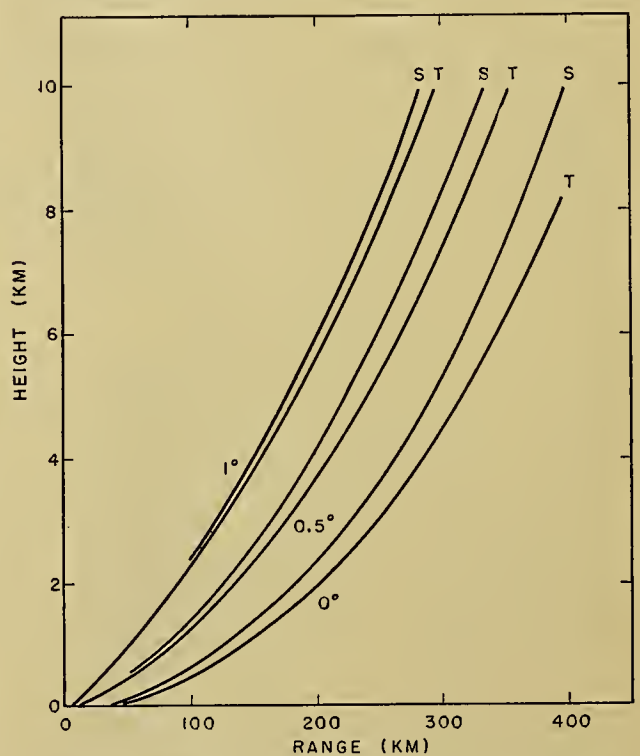


Fig. 2.—Range-height chart for refraction in a standard (S) and tropical (T) atmosphere for elevation angles of 0°, 0.5°, and 1°.

In Fig. 2 a range-height diagram is given for a standard and a tropical atmosphere for rays initially at angles of elevation of 0°, 0.5°, and 1° above the horizontal. As an example of the use of the diagram, it may be seen that a radar with a 1° beam can detect precipitation only between elevations of about 2 to 6 km at a distance of 200 km. If the top of the storm were at 4 km only one-half of the beam would be intercepted at that range.

It is evident that a serious deterrent to the detection of precipitation over long ranges is the fact that beyond a distance of about 200 km (varying from summer to winter) the upper portion of the beam is above most storms, and even at shorter ranges much of the beam

2. On the A scope the abscissa is distance and the ordinate is the relative signal strength at the receiver.

is detecting those portions of the storm which give relatively weak echoes.

Attenuation

Attenuation of microwaves is caused by absorption of oxygen and water vapor in the atmosphere and by absorption and scattering from cloud and precipitation. The water vapor molecule has an electric dipole moment which interacts with the electric field of the radiation and has resonance at 1.33 cm. A stronger absorption band at a wave length of less than 0.2 cm also contributes to the attenuation. The oxygen molecule has a magnetic dipole moment which interacts with the magnetic field of the radiation with resonance at 0.5 cm. Because of the uncertainty of the line breadths of the absorption bands, actual values of the oxygen attenuation are known only for certain regions. Table II gives Van Vleck's [12, 13] values of the attenuation

Table II. Attenuation* Due to Oxygen

Wave length (cm)	Oxygen	Water vapor
1	0.014 -0.036	0.071
1.25	0.010 -0.026	0.21
3	0.0072-0.018	0.0046
10	0.0066-0.014	0.0031

* In decibels per kilometer at 760 mm pressure and 10 g kg^{-1} water vapor, and at a temperature of 293K.

due to oxygen and water vapor. The attenuation due to oxygen is directly proportional to the atmospheric pressure. Van Vleck favors the lower values of oxygen attenuation. At a wave length of 3 to 10 cm the total atmospheric attenuation is therefore 0.01 to 0.02 db km^{-1}.

More serious is the attenuation due to rain. For moderate to heavy rains, computed on the basis of the mean drop-size distribution of Laws and Parsons, Ryde's values for wave lengths of 1.25, 3, and 10 cm are 0.14, 0.023, and 0.003 db km^{-1} mm^{-1} hr^{-1}, respectively. Values of theoretical attenuation given by Haddock in an unpublished paper differ somewhat from Ryde's values.

Considerable error is involved in experimental measurements of rain attenuation since intensity and drop-size distribution vary over the path. In experimental determination of the attenuation at 0.62 cm by Mueller [9] there was excellent agreement with Ryde's theoretical values. Measurements by Robertson and King [10] indicate a scatter of values and give a mean of 0.03 db km^{-1} mm^{-1} hr^{-1} at 3.2 cm. This is much higher than the average theoretical values but within the theoretical limits. For 1.25 cm, Anderson and collaborators [1] found an average attenuation of 0.18 db km^{-1} mm^{-1} hr^{-1} for Hawaiian rainfall. This exceeds Ryde's theoretical maximum values. Reasons for this discrepancy are difficult to see especially since Anderson's experiments appear to be the most careful.

Vertical Structure of Precipitation

Precipitation as observed by radar may be generalized into two categories: showers and continuous precipitation. Continuous precipitation is associated with small vertical velocities of the order of 10 to 30 cm sec^{-1}, while the vertical velocities of showery precipitation may exceed 1 m sec^{-1}. There is, of course, a continuous transition between the two. On an RHI scope,[3] continuous precipitation is characterized by a horizontal layer whereas showers appear as elongated vertical bands. The two types may occur within a few miles of each other and may be found mixed in almost any proportion.

A detailed study of the vertical structure of precipitation has been made by Langille, Gunn, and Palmer [5]. By calibrating the gain of the radar and by assuming that the summation of the sixth powers of the diameters of the drops is proportional to the square of the mass of water illuminated by the beam, they were able to determine lines of equal water content. Some conclusions from these and other observations of the two extremes of precipitation are:

1. Showers

a. Within the space of a mile the liquid water content may change by a factor of more than 100.

b. The liquid water content may be greater aloft than near the ground.

c. Heavier precipitation is generally associated with greater vertical extent of the storm.

d. The motion of the dense portion of a storm (high liquid-water content) is always towards the ground.

2. Continuous Precipitation

a. A bright band, a horizontal layer of high echo intensity 100 m to about 300 m below the freezing level, is an identifying feature of continuous rain. (A separate section below is devoted to the theory and description of this phenomenon in relation to the vertical structure of continuous precipitation.)

b. Radar echoes along the horizontal change less rapidly than echoes from showers.

c. The maximum height of precipitation echoes often remains constant while the rain intensity varies. Such behavior has been observed on both K- and X-band radar with the antenna directed towards the vertical.

The maximum height of the radar echo indicates the level at which particles first approach precipitation size (greater than 0.1 mm) and may vary with the radar. If the radar echo dies off gradually until it is indistinguishable from the noise level, there is no indication as to where the actual top of the cloud may be; but if the echo falls off suddenly (except near the bright band), then it is reasonable to expect that the visual top of the cloud is not far above. The latter type of echo is associated with precipitation from well-developed cumulonimbus clouds, while the former is associated with stratiform clouds and represents a very gradual growth of the precipitation elements.

The maximum height of radar echoes observed in temperate latitudes is found to be 15,000 to 20,000 ft for most widespread winter rains and near 30,000 ft for precipitation from cumulonimbus clouds in summer. The maximum height observed in a thunderstorm at

3. RHI = Range Height Indicator. The abscissa is horizontal range, the ordinate is usually angle of elevation.

Belmar, New Jersey, was 50,000 ft and in Florida it was observed by Byers to be 55,000 to 60,000 ft. Precipitation that reaches the ground as snow generally gives a lower maximum echo height than rain, and snow showers have been observed with maximum heights lower than 5000 ft.

The temperature at the top echo of most mature thunderstorms is generally below −30C. It is at about this temperature, according to W. and E. Findeisen [3], that the number of freezing nuclei become numerous, with large vertical velocities and rapid growth in the high liquid-water content of the thundercloud. However, small thunderstorms have been observed with the temperature at the top echo and the visual top at about −18C [7].

It has been observed that rainfall with maximum echo heights corresponding to temperatures above −9C rarely reaches the ground. The smaller number of ice particles produced at such high temperatures would grow rapidly to precipitation size, drop out of the cloud, and evaporate in the unsaturated air beneath. With greater vertical extent of the cloud above 0C, more numerous ice nuclei are produced which grow to precipitation size over a larger height interval and are numerous enough below the cloud to maintain near-saturated conditions.

Observations of twelve thunderstorms in New Mexico indicate that the first echo from a growing thunderstorm appears at about the −10C isotherm with the top of the cloud at about −30C. The echo then rises rapidly to a height where the average temperature is −30C with the visual top of the cloud at about −32C. Although visual observation of the top is difficult in a degenerating thunderstorm it appears that both the top of the echo and of the cloud descend thereafter.

The Bright Band and Precipitation Growth

A characteristic A-scope presentation of the bright band is shown in Fig. 3. The first intense signal is the transmitted pulse; above that, the weaker echoes from the precipitation are fairly constant in intensity up to the bright band, the intense saturated signal at about 7500 ft. The width of the bright band appears to be less than a thousand feet. Above the bright band the signal falls off gradually and disappears at about 15,000 ft.

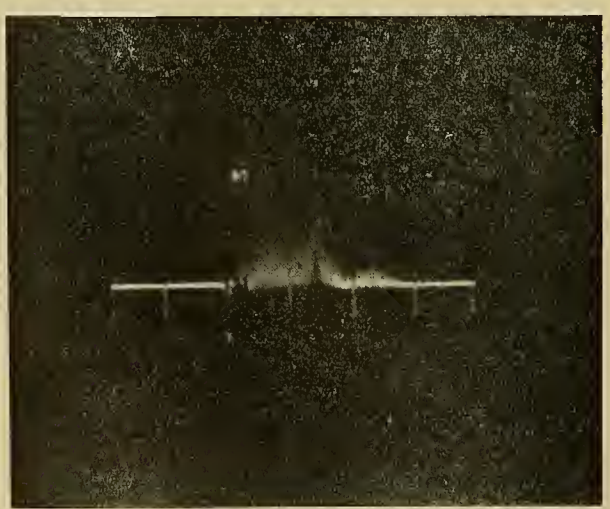


Fig. 3—Bright band observed at Belmar, N.J., January 26, 1949 by Weinstein [14]. The markers are at 1-mile intervals. The ordinate represents signal intensity.

Observation of the bright band was first reported by the Canadian Army Operational Research Group in 1945. From twelve days of observation it was concluded that the echoes from the bright band were from just below the 0C isotherm. It was observed only in stratified clouds although some vestiges of the bright band, considerably weaker in contrast, have been detected in the final stages of a thunderstorm. In some cases the bright band was observed in some sectors but not in others.

Subsequently, aircraft investigations revealed that immediately above the bright band the precipitation was in the form of snow, in the bright band it was melting snow, and below the bright band it was rain. As an explanation it was surmised that "a melting snowflake is covered by a film of water which causes it to act as a raindrop effectively larger than the size of the raindrop resulting from the complete melting of the snowflake."

Ryde [11] rejected this explanation and offered an alternative theory. At temperatures between −4C and 0C a cloud of individual ice crystals aggregate to form snowflakes, thereby giving increased reflectivity. As a result of this aggregation, the relative intensity was estimated to increase from 1 to 40. As the snowflake melts, the reflectivity is further increased owing to the higher dielectric constant of water as compared to ice. Because of this effect, a raindrop resulting from the melting of a snowflake has about five times the reflectivity of the original flake. The decrease in signal strength below the bright band was explained as due to the increased fall velocity of raindrops as compared to snow; hence the number of raindrops per unit volume of air is smaller than that of snowflakes above. Since the fall velocity of raindrops is about five times that of snowflakes, the signal strength would be one-fifth that of the melting snow.

From the theory of storm detection it may be seen that the reflectivity of a melting snowflake is proportional to $K^2/\rho V$ where $K = (\epsilon - 1)/(\epsilon + 2)$; ρ is the density of the mixture, and V is the fall velocity. Ryde assumed that

$$\frac{K}{\rho} = m_i \frac{K_i}{\rho_i} + m_w \frac{K_w}{\rho_w}, \tag{6}$$

where subscripts i and w refer to ice and water, respectively, and m_i and m_w are the respective percentages of the total mass of the snowflake ($m_i + m_w = 1$). This equation is true for a heterogeneous mixture of ice and water but it is not true where the water is on the outside of the snowflake. A particle with an inner core of ice and an outer shell of water has a dielectric constant more nearly that of the water than equation (6) would indicate because of the attenuation of the radio wave towards the interior of the particle (as in the "skin effect"). The increase in signal strength at the top portion of the

bright band is then due to this discontinuous increase of the dielectric constant as a film of water develops on the exterior of the snowflake. The decrease of signal on the bottom portion of the bright band is due to the rapid increase of fall velocity as the snowflake approaches complete melting. It is not necessary, however, to assume (as Ryde did) a nearly constant fall velocity of the snowflake until complete melting, but rather an increase, rapid with respect to one-half the pulse length; this is no serious limitation since the observed width of the bright band is of the same order as one-half the pulse length.

Bowen[4] has recently reported the presence of an upper band observed in Australia on about half the occasions on which the bright band was noted. This upper band, first located at temperatures between −12 and −17C, was observed to fall slowly towards the bright band, merge and intensify, after which the bright band and the precipitation below became more intense. There was a tendency for the process to repeat itself. Bowen ascribed this phenomenon to the freezing of large water droplets of about 400-μ diameter at about −16C and their subsequent rapid growth and fall toward the surface. If this were true the signal strength should first decrease by a factor of 5 and then slowly increase as the ice grew by diffusion.

It is more probable that the upper band is a layer of graupel developed in a region of high liquid-water content in the upper portion of the cloud. It may be shown that there is a critical water content below which crystals maintain their tabular shape and above which crystals convert into graupel of approximately spherical shape. Because of their increased fall velocity and efficiency of catch, graupel, once formed, will grow rapidly, by coagulation with the water droplets. A layer of graupel will deplete the water in its path rather rapidly and a succeeding layer will fall through air of smaller water content and will not acquire as great a mass in the same distance of fall. As an example, a layer of graupel initially of mass 0.001 mg and 150 m in depth (and of reasonable concentration) will grow in a cloud of 0.5 g m^{-3} liquid-water content to a mass of 0.20 mg in a 600-m fall distance. In so doing it will have depleted the water in its path to an average value of about 0.3 g m^{-3}. The next 150-m layer of graupel will grow to a mean mass of about 0.1 mg in the same 600-m fall distance. Since the resolving power of a radar with a 1μ sec pulse is only 150 m, the first layer will be observed to descend as a band of about four times the intensity of the second. Subsequently, the formation of graupel will cease due to depletion of the water below the critical value and will begin again only when the liquid water, replenished by the updraft, rises above that value.

Of meteorological interest is the increase in signal strength from above, as the bright band is approached, as shown in Fig. 3. The ultimate effect of the increase in dielectric constant and fall velocity is an approximate equalization of signal. However, A-scope photographs generally show a rapid increase of signal strength through the bright band, sometimes apparently beginning at the upper edge as in Fig. 3, but often starting from slightly below the 0C level. There is generally a gradual increase from several thousand feet above, where the signal is indistinguishable from the noise level of the radar, to almost the 0C level. Observations made by Austin and Bemis [2] show that the ratios of the signal strength from below the bright band to that above it vary from 1 to 17, with a median near 5. In some cases the signal from above the bright band is vanishingly small. Ryde ascribed the increase in signal strength from above the bright band as due to the aggregation of numerous snow crystals at temperatures of about −4C to 0C, and Austin and Bemis explain the increase in signal strength through the bright band (from above) as due to continued aggregation of snowflakes.

Radar observations of three warm-front rains by Browne in Cambridge, England, give the rate of decrease of signal strength with altitude from the 0C level. In comparing these observations with the rate of growth of ice crystals by diffusion in a supercooled water cloud, calculated according to the methods of Houghton [4], it was found that the calculated growth rates above about the −6C level are not compatible with the observations. The observations were found to be compatible with a growth in an ice cloud such that the crystals consume within a layer the water produced by the updraft and make use of little or no "stored" water from below. This result holds for a rate of mass transfer of air by the updraft that is constant with height. From −6C to −3C the observations are compatible with growth in a water cloud. Between −3C and 0C the observed increase in signal strength is greater than could be obtained by growth of ice crystals at water saturation at those temperatures. This discrepancy is explained by coalescence of some of the ice crystals as they approach the 0C level.

If x crystals of equal mass coalesce, the resultant signal strength is x times the original value. To arrive at an exponential law of drop-size distribution [6], a similar law for coalescence is required. Let the number of coalescences of x crystals of equal mass be given by $n = n_0 e^{-ax}$, then the number of crystals before coalescence is $\sum_x xe^{-ax}$ and the ratio of the radar echo from the coagulated crystals to that from the individual crystals is $\Sigma x^2 e^{-ax}/\Sigma x e^{-ax}$. For any ratio of signal strength from two levels between which coalescence is taking place, the constant a can be computed and the resultant size distribution calculated.

The observations and calculations indicate that coalescence of the particles begins at about the −3C level, possibly because of the cohesive effect of the thin film of water forming on the crystal at these higher temperatures. Below the 0C level, coalescence must continue in order to account for the increased signal below the bright band and the observed drop-size distribution at the ground. These conclusions also agree with snow observations which indicate individual crys-

4. Bowen, E. G., "Radar Observations of Rain and Their Relation to Mechanisms of Rain Formation." *J. atmos. terr. Phys.*, 1:125-140 (1951).

tals arriving at the ground at low temperatures and large snowflakes near the freezing point. The ideas of Ryde, and of Austin and Bemis, on the coalescence of crystals from the −4C level to below the bright band are corroborated; Ryde's ratio of 40 to 1 between −4 and 0C is not indicated by Browne's observations. More observations similar to those by Browne for various cloud depths and precipitation intensities are desirable for a more detailed study of precipitation growth by diffusion and coalescence.

Horizontal Structure of Precipitation

On the PPI scope[5] of the radar, continuous precipitation appears as an extensive brightened area. Generally there are indications of definite edges to the storm, although in winter the precipitation may extend in all directions from the station out to the limit of the range of observation of the radar. An extensive area of warm-front precipitation is often seen to have an abrupt line of cessation.

In showery activity the precipitation may appear as scattered cells, 3–6 miles in diameter, located at random in some sectors of the scope. Occasionally these cells develop to form an area of continuous precipitation and curious patterns often occur. Such a development characterizes the rain areas observed from Belmar, New Jersey, during the afternoon of March 10, 1949 (Fig. 4). A cyclonic storm was located just west of Cape Hatteras at 1230 EST and was moving eastward at about 30 mph. At the 700-mb level a closed cyclonic center was located over southern Ohio, and south-southwest winds of about 20 mph prevailed over the New Jersey area. At 1350 EST showers appeared to be scattered almost at random out to a radius of 125 miles from the station. The remainder of the afternoon saw the virtual decay of the storms in the northern sector and a build-up of the storms to the south. At 1515 EST, what had previously been a loose coalescence of cells was a continuous pattern about 30–50 miles wide but narrowing in its central portion to about 10 miles and trailing off to a thin line at its most distant point, 180 miles distant at 160° azimuth. Farther south a strange pattern in the form of a wavy line of individual cells of precipitation 200 miles or more in length became visible at 1615 EST.

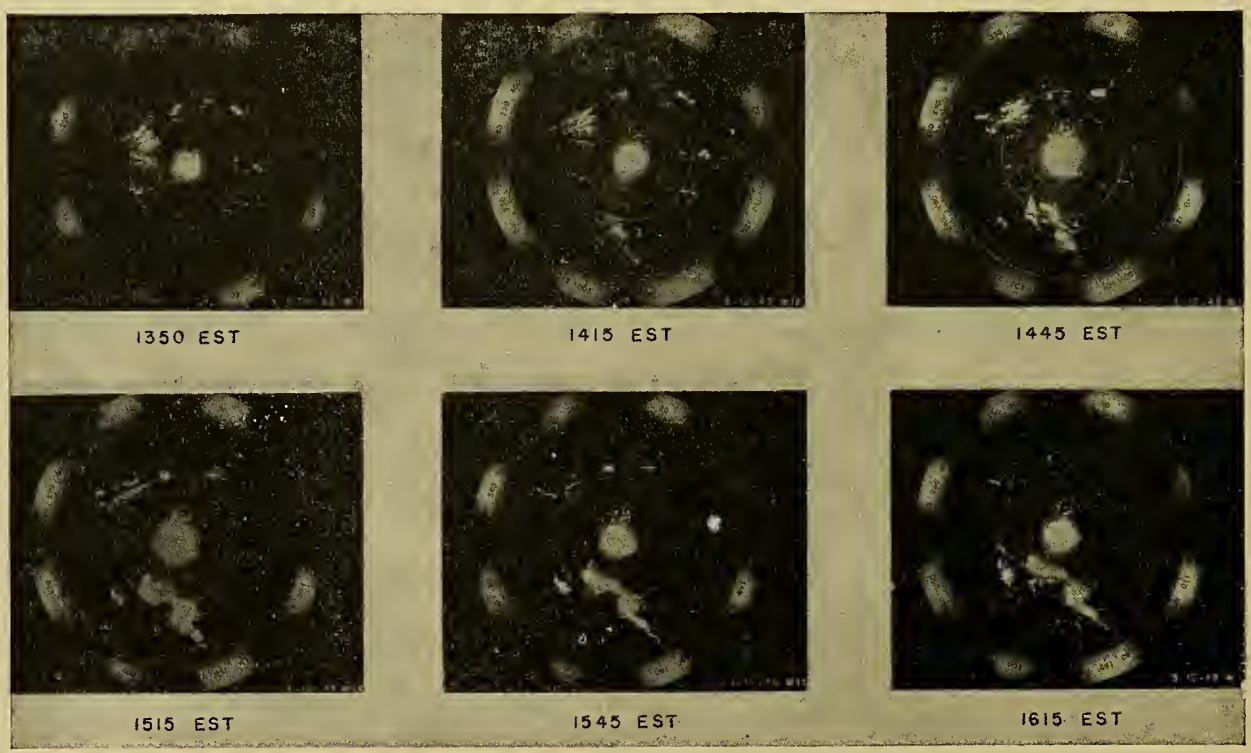


Fig. 4.—Precipitation patterns observed from Belmar, N. J., March 10, 1949. Circular range markers are at 25-mile intervals. (*Courtesy Signal Corps Engineering Laboratories.*)

A line squall that is observed to extend straight across the PPI scope is rare. A more usual situation is a series of line squalls, each perhaps a hundred miles in length, in parallel bands. The edge of the squall in the direction of the high pressure often trails off into a group of cells. The parallel bands of precipitation are generally associated with distinct wind shifts and pressure-gradient discontinuities.

As a front passes over the radar, the range of detection of precipitation along the front diminishes considerably at 3-cm or shorter wave lengths, because of rain attenuation. By comparing the maximum range at a given gain setting before the front has reached the station with the maximum range when the front is directly over the station, it is possible with the aid of equation (4) to

5. PPI = Plan Position Indicator, a polar coordinate plot of azimuth versus range.

determine roughly the average rain intensity over the path [15].

Conclusion

Although the theory of storm detection is fairly straightforward, the complexities of computation from the Mie theory still leave some discrepancies in the amount of reflection to be expected from various precipitation elements and in the values of attenuation. For the purpose of determining quantitatively the target characteristics of a storm, a greater source of error lies in the inaccuracies in measuring the power received at the radar. More exact measurements may open up a new field of estimating the rain intensities of distant storms and of making quantitative studies of the growth processes of precipitation along the vertical.

It would be particularly desirable to measure the rate of decrease with altitude of the signal strength from different clouds of varying depths, precipitation intensities, and atmospheric conditions. Such measurements would give information on the growth of precipitation particles by diffusion and the magnitude of the coalescence effects. Observations of the type, mass, and fall velocity of snow that reaches the ground would give considerable supplementary information for an understanding of the physics of precipitation.

Simultaneous radar and visual (theodolite) observations of cumuliform clouds, similar to those carried out by Workman in New Mexico, should also lead to a better understanding of thunderstorms. Fruitful research problems are (1) the rate of rise of the visual top as related to the echo top in growing thunderstorms, (2) coagulation and chain reaction effects in the heavy rain, again as revealed by the rate of decrease with altitude of signal strength, and (3) the characteristics of precipitation echoes during the formation of the bright band in the degenerate stages of the thunderstorm, as revealed by radarscope photographs taken with reduced gain. Although many complexities in the individual storms present themselves in the study of such problems, the complexities themselves offer fruitful paths of research. The ever-changing panorama of weather is nowhere better revealed than on the radar screen.

REFERENCES

1. Anderson, L. J., and others, "Attenuation of 1.25 cm Radiation through Rain." *Proc. Inst. Radio Engrs., N. Y.*, 35:351–354 (1947).
2. Austin, P. M., and Bemis, A. C., "A Quantitative Study of the 'Bright Band' in Radar Precipitation Echoes." *J. Meteor.*, 7:145–151 (1950).
3. Findeisen, W., und Findeisen, E., "Untersuchungen über die Eissplitterbildung an Reifschichten." *Meteor. Z.*, 60:145–154 (1943).
4. Houghton, H. G., "A Preliminary Quantitative Analysis of Precipitation Mechanisms." *J. Meteor.*, 7:363–369 (1950).
5. Langille, R. C., and Gunn, K. L. S., "Quantitative Analysis of Vertical Structure in Precipitation." *J. Meteor.*, 5:301–304 (1948).
6. Laws, J. O., and Parsons, D. A., "The Relation of Raindrop-Size to Intensity." *Trans. Amer. geophys. Un.*, 24:452–460 (1943).
7. Ludlam, F. H., "Observations of a Thunderstorm from Small Cumulonimbus." *Meteor. Mag.*, 78:357–360 (1949).
8. Mie, G., "Beiträge zur Optik trüber Medien." *Ann. Phys., Lpz.*, (4) 25:377–445 (1908).
9. Mueller, G. E., "Propagation of 6-Millimeter Waves." *Proc. Inst. Radio Engrs., N. Y.*, 34:181–183 (1946).
10. Robertson, S. D., and King, A. P., "The Effect of Rain upon the Propagation of Waves in the 1- and 3-Centimeter Regions." *Proc. Inst. Radio Engrs., N. Y.*, 34:178–180 (1946).
11. Ryde, J. W., "The Attenuation of Centimetre Radio Waves and Echo Intensities Resulting from Atmospheric Phenomena." *J. Instn. elect. Engrs.*, 93(3A):101–103 (1946).
12. Van Vleck, J. H., "Absorption of Microwaves by Oxygen." *Phys. Rev.*, 71:413–424 (1947).
13. —— "Absorption of Microwaves by Uncondensed Water Vapor." *Phys. Rev.*, 71:425–433 (1947).
14. Weinstein, J., "Three-Dimensional 'Bright-Line' Representation." *J. Meteor.*, 6:289 (1949).
15. Wexler, R., "Radar Detection of a Frontal Storm 18 June 1946." *J. Meteor.*, 4:38–44 (1947).
16. —— and Swingle, D. M., "Radar Storm Detection." *Bull. Amer. meteor. Soc.*, 28:159–167 (1947).

METEOROLOGICAL ASPECTS OF PROPAGATION PROBLEMS

By H. G. BOOKER

Cornell University

Introduction

We shall be concerned with the branch of radio-meteorology that deals with refraction of radio waves in the atmosphere. Refraction phenomena depend mainly on the profiles of temperature and humidity near, or comparatively near, the earth's surface. The first section is a description of the phenomenon of extra downward refraction, or *superrefraction*, as experienced at radar stations. The second section explains in broad outline how superrefraction comes about, while the third section focuses attention on the meteorological phenomena mainly responsible for it. In the last section some meteorological investigations that have been made especially in connection with radio propagation are discussed.

Description of Supperrefraction

The expression superrefraction is used because, even under conditions of orthodox propagation, there is some downward bending of radio waves in the atmosphere due to the normal decrease of air density with height. Radio people have a simple way of allowing for this normal downward refraction, so that it is with departures from orthodox atmospheric refraction that we have to deal. The most striking phenomena are associated with increases in downward refraction, and particularly with situations in which the downward refraction of radio rays exceeds the curvature of the earth, although reductions in downward refraction—subrefraction—also occur.

The phenomenon of superrefraction, although known before 1940 [8, 9], was brought vividly to attention during World War II by the development of radar. Under orthodox propagation conditions the region within which a radar may detect aircraft, ships, coast lines, etc., does not normally extend quite as far as the geometrical horizon, and radar vision of targets below the geometrical horizon is usually impossible. When radars came into operation, at first on a wave length of 13 m, the expectation that they would not see beyond the geometrical horizon was broadly fulfilled, although there was some evidence that echoes from objects on the ground were dependent upon atmospheric conditions. It was especially noticed that, during an intense anticyclone in January 1940 (with snow on the ground), echoes were obtained from tall towers some distance beyond the geometrical horizon, and at the same time the normal ability of the station to detect aircraft was somewhat modified. As time went on, radars were developed on progressively shorter wave lengths—7 m, 1.5 m, 50 cm, 10 cm, 3 cm—and it was found that the shorter the wave length in use the more frequently were conditions of unorthodox propagation encountered and the more intense was the degree of superrefraction experienced. Thus, when 10-cm radars were installed along the south coast of England, it was found that they were often able to see the coast of France even where the width of the English Channel was too great for ordinary visual observation of the opposite shore. The phenomenon, when it occurred, was usually most pronounced in the evening, though it sometimes started in the afternoon and extended far into the night. After a while it became clear that these conditions of unusual refraction were associated with fine weather and that they were a good deal more pronounced in summer than in winter. In addition it was realized that radars looking over land instead of over sea also experienced unorthodox propagation conditions in fine weather, but not usually in the afternoon.

Understanding of superrefraction was greatly assisted by the progressive deployment of radars in the various theatres of war. When 7-m and 1.5-m radars were established on coastal sites in the Mediterranean, it soon became clear that superrefraction in this area was much more intense than on the same wave lengths around Britain. The entire south coast of Sicily could frequently be seen with 1.5-m radars on Malta, and from time to time echoes from Greece and Sardinia at ranges of the order of 400 miles were obtained. Ships were sometimes plotted to a range of 200 miles, the range of the geometrical horizon being less than 20 miles. In contrast, ranges on aircraft were occasionally said to be subnormal, though these reports were confused by the possibility that the performance of the equipment may not have been up to standard. It soon became clear that unorthodox propagation was intense, frequent, and widespread over the Mediterranean in summer, though conditions largely returned to normal in winter.

During the course of the war many other areas of the world were found to be subject to superrefraction. These included:

1. Coast of French West Africa in the neighbourhood of the Canary and Cape Verde Islands.
2. Gulf of Aden in summer.
3. Persian Gulf in summer.
4. Northern part of the Arabian Sea during the Indian hot season.
5. West coast of India except during the southwest monsoon.
6. Bay of Bengal.
7. North coast of Australia except during the northwest monsoon.
8. North west coast of Australia all the year round, but particularly in summer.
9. West and south coasts of Australia and coast of

New South Wales in summer (but obliterated by passage of a meridional front).

10. Southeast coast of South Island, New Zealand, during a nor'wester (foehn wind from the Southern Alps).

11. Around North Island, New Zealand, in summer.

Inland it was found almost everywhere that conditions of unorthodox propagation, when they occurred, were largely associated with radiation nights. On such a night moderate superrefraction usually develops fairly quickly after sunset. If there is no tendency for formation of fog, superrefraction usually continues all night, disappearing after sunrise. But if fog forms, superrefraction may disappear during the night and may even be replaced by subrefraction before sunrise, after which conditions begin to return to normal. Fog, even over the sea, is often associated with unorthodox propagation, but this may vary all the way from subrefraction to quite intense superrefraction.

Most of the phenomena of unorthodox propagation outlined above were initially encountered in the course of operational use of ordinary radar equipment. Some examples of the observational material from which it was necessary to work have been given in *Weather* [2]. The qualitative results obtained in this way, however, were followed by quantitative observations made by specially designed equipment using one-way propagation mainly on wave lengths of 9 and 3 cm [10, 12, 16, 17, 20].

It is clear therefore that the simple idea that radars (which operate on metre, decimetre, and centimetre wave lengths) cannot see targets beyond the geometrical horizon may, under suitable conditions of weather and climate, be violated on a stupendous scale.

Explanation of Superrefraction

From the outset there was, of course, no doubt that the phenomena described in the previous section were due to refraction or reflection in the troposphere. In the early days, some confusion was caused by repeated suggestions that reflection from the interfaces constituting ordinary meteorological fronts was the cause. In fact, however, reflections from ordinary frontal surfaces, although sometimes observable at grazing incidence, are not the primary cause of superrefraction. But there is another way in which striking gradients of temperature and humidity can occur in the troposphere. Owing to advection, radiation, subsidence, or some combination of these phenomena, it can easily happen that the surface of the earth has a temperature substantially different from that representative of the superincumbent air mass. An important gradient of temperature then occurs near the surface of the earth. In the same way, the humidity representative of the superincumbent air mass may differ appreciably from that existing close to the earth's surface, and then there is an important gradient of humidity near the bottom of the air mass. Gradients of temperature and humidity that occur in this way are the main cause of unorthodox radio propagation.

Two extremely important parameters in radiometeorology are: (1) the temperature excess of the appropriate air mass in relation to the earth's surface, and (2) the humidity deficit of the appropriate air mass in relation to the earth's surface. The first is the excess of the potential temperature of the air mass over the value corresponding to the earth's surface, and the second is the deficit of the specific humidity of the air mass below that corresponding to the earth's surface (see Fig. 1). Quite a good qualitative idea of the degree of superrefraction can often be derived purely from knowledge of the values of these two parameters, large positive values indicating marked superrefraction. For a more detailed understanding of superrefraction, however, it is necessary to know not merely the temperature excess and humidity deficit but also the associated profiles of potential temperature and specific humidity. From these it is possible, by means of a formula due to Englund, Crawford, and Mumford [8], to deduce the profile of the radio refractive index and hence the propagational properties of the atmosphere.

The formula giving the radio refractive index n as a function of atmospheric pressure p in millibars, tem-

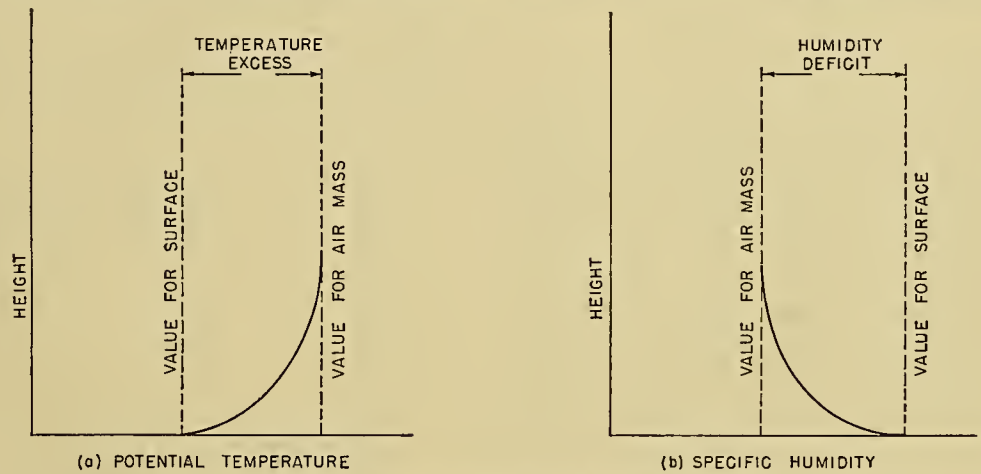


FIG. 1.—Modification of an air mass by surface conditions.

perature T in degrees absolute, and partial pressure e of water vapour in millibars is

$$(n-1)\times 10^6 = \frac{79}{T}\left(p - \frac{e}{7} + \frac{4800e}{T}\right). \qquad (1)$$

We need to know how large the vertical gradients of temperature and humidity must be in order to be of importance in radiometeorology. Consider first a situation in which the potential temperature is uniform with height; how large must the lapse rate of specific humidity be to make the downward curvature of a radio ray equal to the curvature of the earth? Let α be this critical lapse rate of specific humidity. Consider secondly a situation in which the specific humidity is uniform with height; how large must the negative lapse rate of potential temperature be to make the downward curvature of a radio ray equal to the curvature of the earth? Let β be this critical lapse rate of potential temperature, its numerical value being in fact negative. The values of α and β may be derived from (1) and depend somewhat on pressure, temperature, and humidity. They are of the order of

$$\alpha = 0.5 \text{ g kg}^{-1} \text{ per } 100 \text{ ft}, \qquad (2)$$
$$\beta = -5\text{F per } 100 \text{ ft}. \qquad (3)$$

Thus lapse rates of humidity in excess of about ½ g kg^{-1} per 100 ft and inversions of temperature in excess of about 5F per 100 ft are of great importance in radiometeorology as they cause downward bending of radio rays as great as or greater than the curvature of the earth, thereby making possible radio vision round the curved surface of the earth.

On the other hand, even in a well-mixed atmosphere, with specific humidity and potential temperature uniform with height, there is some downward bending of radio rays. This is the normal atmospheric refraction which is involved in orthodox propagation and to which reference has been made earlier in this article. Let κ_0 be the downward curvature of radio rays when there is no lapse rate of specific humidity or potential temperature. It is possible to deduce κ_0 from (1), and, like α and β, it depends somewhat on pressure, temperature, and humidity. It is convenient to compare κ_0 with the curvature κ_e of the earth (5×10^{-6} radians per 100 ft), and expressed in this way the value of κ_0 is of the order of

$$\kappa_0 = \tfrac{1}{6}\,\kappa_e. \qquad (4)$$

Thus, even in a well-mixed atmosphere, there is downward bending of radio rays amounting to about one-sixth of the earth's curvature.

In terms of the parameters κ_0, α, and β we may express the downward curvature κ appropriate to a lapse rate q' of specific humidity and a lapse rate θ' of potential temperature in the form

$$\frac{\kappa - \kappa_0}{\kappa_e - \kappa_0} = \frac{q'}{\alpha} + \frac{\theta'}{\beta}. \qquad (5)$$

This makes

1. $\kappa = \kappa_0$ when $q' = \theta' = 0$,
2. $\kappa = \kappa_e$ when $q' = \alpha$, $\theta' = 0$,
3. $\kappa = \kappa_e$ when $q' = 0$, $\theta' = \beta$.

From Snell's laws of refraction it is easily shown that the downward curvature κ of a ray is also equal, to a high degree of approximation, to the lapse rate of refractive index. Thus (5) also expresses the lapse rate κ of radio refractive index in terms of the lapse rates q' and θ' of specific humidity and potential temperature. In addition to these lapse rates, the actual values of humidity and temperature, as well as the value of the pressure, enter into (5) because κ_0, α, and β depend on these quantities to some extent, as implied by equation (1). But for quite a wide range of meteorological conditions it is often possible to regard κ_0, α, and β as constants, and (5) then expresses κ purely as a function of q' and θ'.

Now under conditions when the temperature and humidity of an air mass have been modified by surface conditions, simple though not unusual profiles of potential temperature and specific humidity are as indicated in Fig. 1. It is here supposed that the air mass is warm and dry in comparison with surface conditions. It is not uncommon for the steepest gradients of temperature and humidity to occur near the surface of the earth, and for them to be such as to make the downward curvature of radio rays near the earth's surface exceed the curvature of the earth. This leads to the state of affairs depicted in Fig. 2. Well up in the air mass the

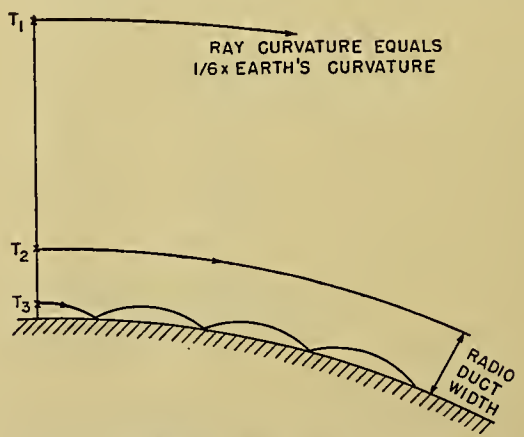


Fig. 2.—Effect of radio duct on radio rays.

lapse rates of potential temperature and specific humidity are zero, and so, in accordance with (5), the downward curvature of radio rays is κ_0, about one-sixth of the earth's curvature. A ray emanating horizontally from a radio transmitter T_1 at such a height would show only a slight tendency to follow the earth's curvature, as indicated in Fig. 2. Now let us bring the transmitter down to a level where modification of the air mass by surface conditions begins to be appreciable. In accordance with (5), the positive lapse rate of specific humidity and the negative lapse rate of potential temperature indicated in Fig. 1 increase the downward curvature of rays above the standard value κ_0. When the transmitter has been brought nearly down to the earth's surface, it reaches a certain level, T_2 in Fig. 2, where the downward curvature of the ray becomes equal to the curvature of the earth. With the transmitter at this

critical level, a ray emanating horizontally remains at the same height above the earth's surface and does not fly off at a tangent. This vital level, where the downward curvature of a ray is equal to the curvature of the earth, forms the top of what is known as the radio duct. As we bring our transmitter down below the top of the duct, a ray emanating horizontally, say from T_3 in Fig. 2, is bent downward to such an extent that it hits the earth and suffers successive reflections from it. The ray, in fact, is trapped within the duct. It is trapping of this sort that causes low-level radars to see low-level targets far beyond the geometrical horizon.

Trapping of rays in a radio duct as illustrated in Fig. 2, although it provides a clear insight into the phenomenon of superrefraction, gives a substantially oversimplified picture [4]. The oversimplification does not reside in the meteorological aspects of the problem, however, and therefore we need not go into the matter in great detail. The principal trouble with Fig. 2 is that it suggests that, under the appropriate atmospheric conditions, superrefraction would be experienced at all radio wave lengths. Now it is well known that, apart from ionospheric refraction, with which we are not concerned, there is no phenomenon of superrefraction at ordinary broadcasting wave lengths. Propagation in the lower atmosphere at ordinary broadcasting wave lengths is almost completely independent of weather. Even under meteorological conditions involving a pronounced radio duct and producing marked superrefraction at radar wave lengths, propagation in the lower atmosphere at broadcasting wave lengths is almost completely orthodox. The reason for this dependence of superrefraction upon wave length is that sufficiently long wave lengths respond only to a crude average of the atmospheric gradient, but sufficiently short ones fully appreciate all the fine structure. Thus, to produce marked superrefraction on a wave length of 1 km would require a radio duct extending from the earth's surface up to a height of the order of 30,000 ft. No approximation to such enormous ducts ever occurs in the atmosphere, and so serious superrefraction at ordinary broadcasting wave lengths is out of the question. On the other hand metre wave lengths can respond quite effectively to unusual gradients of temperature and humidity within the first few thousand feet of the atmosphere, while centimetre wave lengths can be seriously affected by unusual gradients within the first few hundred feet. It thus comes about that superrefraction is a phenomenon that is quite widespread at centimetre wave lengths, fairly widespread at decimetre wave lengths, only moderately widespread at metre wave lengths, quite exceptional at dekametre wave lengths, and virtually unknown at longer wave lengths. It therefore has to be realized that, in addition to meteorological parameters such as temperature excess, humidity deficit, wind speed, which control the distribution of radio refractive index in the lower atmosphere, there are nonmeteorological parameters such as wave length, height of transmitter, height of receiver, which also have a big influence upon radio propagation.

In Fig. 2 we have illustrated a situation in which the steepest gradients of temperature and humidity occur at the surface of the earth. This is by no means always the case however. For example, when superrefraction is produced by a subsidence inversion, the distributions of potential temperature and specific humidity with height are frequently of the type shown in Fig. 3 and

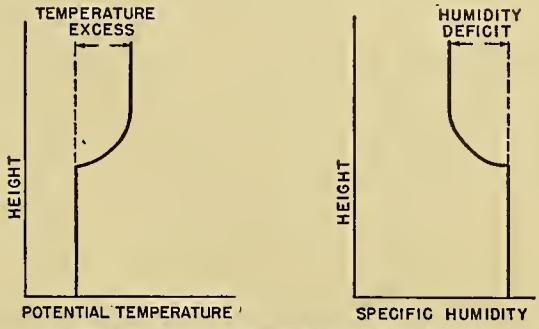


Fig. 3.—Profiles of potential temperature and specific humidity associated with an elevated refracting layer.

this leads to an elevated refracting layer instead of one resting on the surface of the earth. In such a case, the degree of superrefraction experienced in communicating between points on the earth's surface depends on (1) the temperature excess and humidity deficit of the air mass above the layer in comparison with that below the layer, (2) the precise profile of refractive index in the layer, and (3) certain radio parameters (particularly wave length), but it also depends quite vitally upon the height of the layer above the surface of the earth. The lower the height at which a layer exists, the more effective it usually is in causing superrefraction. The reason for this may be illustrated by supposing that the refractive index of the atmosphere suffers a small discontinuous decrease as we ascend through a level S, the refractive index above and below this level being uniform. Such a discontinuous drop in refractive index causes total internal reflection of a ray incident from below at a sufficiently glancing angle. But, on account of the curvature of the earth, even a ray leaving the surface of the earth horizontally attacks the level S at a nonzero angle to the horizontal, and this angle is larger the greater the height of the layer. As a result total internal reflection of a ray leaving the surface of the earth horizontally is impossible if the layer is too high. This picture, although grossly oversimplified, makes it clear why reflecting layers must be quite low in the atmosphere to produce marked superrefraction for communication between points on the earth's surface. Thus a striking subsidence inversion nearly always causes outstanding superrefraction if its base is as low as one thousand feet above the earth's surface, but if its base is at ten thousand feet the inversion is unlikely to be of much practical importance except for communication between airplanes.

To summarize, we may say that superrefraction is usually associated with an inversion of temperature exceeding about 5F per 100 ft and/or a lapse rate of humidity exceeding about $\frac{1}{2}$ gm kg^{-1} per 100 ft maintained over an interval of height of the order of 100 ft

or so at a level in the atmosphere not more than a few thousand feet above the earth's surface and often much less. If the lapse rate of specific humidity becomes sufficiently negative, superrefraction is replaced by subrefraction. The degree of superrefraction is vitally affected by the interval of height over which steep gradients are maintained and by the closeness to the earth's surface at which they occur.

The Meteorological Problem

From what has been said it is clear that we need to be able

1. To recognize those situations which involve a temperature inversion exceeding about 5F per 100 ft within a few thousand feet of the earth's surface.
2. To recognize those situations which involve a lapse rate of humidity exceeding about ½ g kg^{-1} per 100 ft within a few thousand feet of the earth's surface.
3. To state, within say 10 ft, over what interval of height the gradients exceed the values mentioned.
4. To state to an accuracy of say 10 per cent or 10 ft, whichever is the cruder, the height above the earth's surface at which the steep gradients occur.

The above requirements are the least that would make a definite contribution to quantitative radiometeorology, and are insufficient to provide all the information desirable. The complete profiles of temperature and humidity, with an accuracy indicated by the requirements listed above, are highly desirable, together with their statistical variations.

There is, of course, no particular difficulty in measuring the profiles of temperature and humidity as required above at a particular location by means of meteorological instruments attached to towers, balloons, kites, aircraft, etc. Important as such experiments are, however, they do not by themselves constitute an adequate understanding of the problem for radiometeorological purposes. The real problem is to be able to specify the profiles of temperature and humidity with sufficient accuracy purely from the weather data ordinarily available in synoptic meteorology. From this source, it would usually be possible to decide with reasonable accuracy what are the values of parameters such as temperature excess and humidity deficit, illustrated in Figs. 1 and 3. In addition wind speed and direction near the surface and at an appreciable height would normally be available, as well as information about the type of terrain. A good deal of additional information of a general character would also be available, and it is reasonable to suppose that contained in all these data are the values of the parameters which control those features of the profiles of temperature and humidity near the earth's surface that are vital in radiometeorology. The problem is therefore to obtain such a fundamental understanding of how the profiles of temperature and humidity near the base of the troposphere are controlled by such parameters as temperature excess, humidity deficit, wind speed, etc., that knowledge of the values of these parameters is adequate to provide answers to at least the four requirements stated at the beginning of this section.

This can be achieved, if at all, by developing a theory of the profiles of temperature and humidity near the earth's surface capable of standing the test of experimental verification. Such a theory should, ideally, specify the required profiles as a function of certain parameters; if these parameters could be deduced from ordinary synoptic weather data, and if the theory were to give the profiles with sufficient accuracy, the problem would be solved.

There are, of course, certain aspects of the problem which, even at the outset, are not in doubt. The most important physical process directly involved in controlling profiles of temperature and humidity is eddy diffusion. High eddy diffusion means almost uniform potential temperature and specific humidity, and consequently orthodox propagation. Low eddy diffusion permits the existence of substantial atmospheric gradients and so makes superrefraction possible. Eddy diffusion is low in a temperature inversion. Thus a temperature inversion is important in connection with superrefraction for two reasons. First, the gradient of temperature itself causes some downward refraction. But, more important, the associated stability of the atmosphere permits the existence of a steep gradient of humidity, and it is this that is frequently the immediate cause of unorthodox propagation.

The important ways in which marked inversions of temperature and associated hydrolapses can occur near, or relatively near, the surface of the earth arise from the processes of subsidence, advection, and nocturnal radiation. Over land, nocturnal radiation inversions are the principal feature associated with unorthodox propagation. Over sea, particularly in coastal waters, advection of warm dry air from land to sea is of great importance, often complicated by a coastal circulation arising from the land-sea difference of temperature. Subsidence in many cases is a prerequisite for surface inversions for which radiation or advection is the more immediate cause. Subsidence keeps the sky clear and permits the land to heat up to a high temperature during the afternoon and cool rapidly at night. This leads to a marked radiation inversion inland at night and often makes available, during the afternoon, a large excess of land temperature over sea temperature for production of advection inversions over the sea. Moreover subsidence brings warm dry air down relatively close to the earth's surface, frequently forming a subsidence inversion. A well-developed subsidence inversion is itself a potent cause of superrefraction propagation between points on the earth's surface provided it exists at a level less than say 5000 ft above the surface, but many subsidence inversions are too high to be of much direct practical importance in ordinary radio communication.

The problem therefore reduces to this: we need to understand the phenomena of subsidence, advection, and nocturnal radiation sufficiently well to form a workable theory of the associated profiles of temperature and humidity.

Some Special Meteorological Investigations

For the phenomenon of advection of air from land to sea, a special attempt has been made to carry out a program of research such as that outlined above, bearing in mind the particular needs of radio propagation. The most satisfactory results at present available are contained in two papers by Craig [6, 7], although reference may also be made to papers by Sheppard [15] and Booker [3]. The first of Craig's papers describes measurements of temperature and humidity made at various heights and at various distances off the coast of Massachusetts in air that had drifted from the land over the sea, the measurements being carried out in such a way as to reveal the modification in the air mass due to advection over the sea. The second of Craig's papers describes how these measurements were analyzed and fitted to a theory of the modification of the air mass by eddy diffusion. The center of interest in such a theory is the assumption to be made about the distribution of eddy diffusivity with height. Taylor [19] had originally assumed a uniform distribution of eddy diffusivity, but this does not reproduce the logarithmic profiles of temperature, humidity, and wind speed usually observed within the first few feet of the atmosphere [13]. Craig therefore used a model in which eddy diffusivity is proportional to height in the first few feet, but has a uniform value at greater heights. This is probably the most successful representation of the process of eddy diffusion near the earth's surface for stable atmospheric conditions that has so far been achieved.

Other distributions of eddy diffusivity with height have however been employed. Following Sutton [18], Booker [3] has assumed that eddy diffusivity is proportional to a power of height above the earth's surface. He shows that, for advection of a well-mixed air mass from land over a uniform sea, the heights at which the lapse rates of specific humidity and potential temperature have the important values (2) and (3) reach maximum values offshore which are proportional to humidity deficit and temperature excess, respectively, but which are independent of the absolute value of the eddy diffusivity. These maximum heights, which can be used in practice as radio duct-widths (see Fig. 2) over a wide range of distances offshore, do depend on the distribution of eddy diffusivity with height, even though they are independent of the absolute value at any particular level. The distances offshore at which the maximum heights occur, unlike the values of the maximum heights themselves, are dependent on the value of the eddy diffusivity, but this is not too important for some practical applications to radiometeorology. Booker thus concludes that, when a radio duct is produced over the sea by advection of air from the land, the radio duct-width is less dependent on the absolute value of the eddy diffusivity than might appear at first sight, and this result would probably also apply to the profile of eddy diffusivity used by Craig. The radiometeorological effect of changing the profile of eddy diffusivity is, however, quite important.

Conclusion

If all the meteorological phenomena involved in radiometeorology were subjected to as careful a treatment as that applied by Craig [6, 7] to the problem of advection of air from land to sea, it would be possible to feel that the basic meteorological factors underlying radiometeorology were understood as well as could be expected for some time to come.

Reference may be made to the discussions of radiational cooling, fog, subsidence, advection, evaporation, sea breeze, etc., given elsewhere in this Compendium. Additional general references in connection with radiometeorology are listed as [5],]11], and [14] in the references.

REFERENCES

1. Booker, H. G., "Elements of Radio Meteorology: How Weather and Climate Cause Unorthodox Radar Vision Beyond the Geometrical Horizon." *J. Instn. elect. Engrs.*, 93(1): 460–462 (1946).
2. —— "Radio Refraction in the Atmosphere." *Weather*, 3: 42–50 (1948).
3. —— "Some Problems in Radio Meteorology. *Quart. J. R. Meteor. Soc.*, 74:277–315 (1948).
4. —— and Walkinshaw, W., "The Mode Theory of Tropospheric Refraction and Its Relation to Wave-Guides and Diffraction," in *Meteorological Factors in Radio Wave Propagation*. London, The Physical Society, 1946. (See pp. 80–127)
5. Burrows, C. R., (chairman) and Attwood, S. S., (editor), *Radio Wave Propagation*. New York, Academic Press, 1949.
6. Craig, R. A., "Measurements of Temperature and Humidity in the Lowest 1000 Feet of the Atmosphere over Massachusetts Bay." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 10, No. 1, 47 pp. (1946).
7. —— "Vertical Eddy Transfer of Heat and Water Vapor in Stable Air." *J. Meteor.*, 6: 123–133 (1949).
8. Englund, C. R., Crawford, A. B., and Mumford, W. W., "Further Results of a Study of Ultra-Short-Wave Transmission Pheuomena." *Bell Syst. tcch. J.*, 14: 369–387 (1935).
9. —— "Ultra-Short-Wave Radio Transmission through the Non-homogeneous Troposphere." *Bull. Amer. meteor. Soc.*, 19: 356–360 (1938).
10. Katzin, M., Bauchman, R. W., and Binnian, W., "3- and 9-Centimeter Propagation in Low Ocean Ducts." *Proc. Inst. Radio Engrs., N. Y.*, 35: 891–905 (1947).
11. Kerr, D. E., and others, *Propagation of Short Radio Waves. Radiation Laboratory Series No. 13*. New York, McGraw, 1951.
12. Megaw, E. C. S., "Experimental Studies of the Propagation of Very Short Radio Waves." *J. Instn. elect. Engrs.*, 93(1): 462–463 (1946).
13. Montgomery, R. B., "Observations of Vertical Humidity Distribution above the Ocean Surface and Their Relation to Evaporation." *Pap. phys. Ocean. Meteor. Mass. Inst. Tech. Woods Hole ocean. Instn.*, Vol. 7, No. 4 (1940).
14. Norton, K. A., *Advances in Electronics*, Vol. 1. New York, Academic Press, 1948. (See pp. 392–397)
15. Sheppard, P. A., "The Structure and Refractive Index of the Lower Atmosphere," in *Metcorological Factors in Radio Wavc Propagation*. London, The Physical Society, 1946. (See pp. 37–79)

16. Smith-Rose, R. L., and Strickland, A. C., "An Experimental Study of the Effect of Meteorological Conditions upon the Propagation of Centrimetric Radio Waves," in *Meteorological Factors in Radio Wave Propagation*. London, The Physical Society, 1946. (See pp. 18–37)
17. Smyth, J. B., and Trolese, L. G., "Propagation of Radio Waves in the Lower Troposphere." *Proc. Inst. Radio Engrs., N. Y.*, 35: 1198–1202 (1947).
18. Sutton, O. G., *Atmospheric Turbulence*. London, Methuen, 1949.
19. Taylor, G. I., "Report by Mr. G. I. Taylor," in *Ice Observation, Meteorology, and Oceanography in the North Atlantic Ocean. Report on Work Carried Out by the S. S. "Scotia," 1913*. London, Darling and Son, 1914. (See pp. 48–68)
20. Wickizer, G. S., and Braaten, A. M., "Propagation Studies on 45.1, 474, and 2800 Megacycles within and beyond the Horizon." *Proc. Inst. Radio Engrs., N. Y.*, 35: 670–680 (1947).

SFERICS

By R. C. WANTA

U. S. Weather Bureau, Upton, N. Y.

Sferics (less commonly spherics) is a contraction of the word atmospherics meaning natural electrical phenomena detected by radio methods. Sferics have variously been known as clicks, grinders, sizzles, X's, strays, *parasites*, and sturbs, and other names. They comprise natural static which interferes especially with amplitude-modulated radio wave reception. Electromagnetic waves at radio frequencies reaching the earth from outside the atmosphere, of solar or other origin, are sometimes included within the meaning of the term, but will not be discussed here. Sudden electrical discharges resulting in redistribution of charge within and between clouds, between clouds and the air space above or below, and between clouds and earth, give rise to electrostatic, induction, and radiation fields. It is the last which principally form sferics at a distance, and with which we are especially concerned here. The significance of sferics in meteorology is due to their origin in relatively intense convection intimately involving water vapor.

The reception of sferics preceded even the artificial generation of radio waves for transmission over long distances. A. S. Popov, using an elevated wire, noticed the association of lightning flashes with the indications of a coherer at Cronstadt in 1895 [9]. The history of sferics during the remaining years of the last century and the first two and a half decades of this century is sketched in two papers, by Cave and Watson-Watt [9] and by Watson-Watt [40]. Sferics were positively correlated with cyclones, polar fronts, thunderstorms, and cumuliform clouds, and less definitely, with showery precipitation. Rudiments of later methods of sferics study appeared—not only observation of frequency and intensity of sferics, but also coordinated observation of bearings of sources of sferics by means of direction finders. Sferics provided a key which opened the door to an understanding of radio wave propagation. Investigators in more than a dozen countries explored the nature and sources of atmospherics, and official international recognition soon was given to this new field of meteorology. Lugeon has recently written an historical account, which includes a summary of the more important international resolutions concerning sferics [17].

Is the source of sferics necessarily lightning? In Popov's case the answer seemed clear, but a complete answer has not yet been given. Nevertheless, Watson-Watt's dictum that no valid evidence contradicts the hypothesis that all sferics originate in lightning still stands [4, 41], at least as regards sferics received at a distance [24]. Besides the positive correlations already mentioned, supporting evidence was found early in the diurnal and annual variation of sferics; sferics originating over land exhibited afternoon and summer maxima, whereas those originating over water exhibited an early morning maximum. Good correspondence was obtained between prevailing areas of sferics origin and the distribution of thunderstorms given by Brooks [6]. Furthermore, the general features of the wave forms of sferics, as determined, for example, by Appleton and Chapman [2], Lutkin [19], and Schonland and Laby and their collaborators [15, 31, 33], corresponded with the details of the lightning discharge. In this connection, see Hagenguth's article[1] elsewhere in this Compendium. The meteorologist may feel safe in treating all sferics arising more than some one hundred miles from a sferics receiver tuned to very low radio frequencies as evidence of violent convection such as occurs in thunderstorms.

The observation of sferics has taken several forms: determination of direction of arrival, measurement of intensity and rate of occurrence, and display of the wave form of individual sferics. From Brooks' estimate that approximately 40,000 thunderstorms occur over the globe per average day and from the fact that day and night ranges of sferics at radio frequencies of the order of 10 kc sec^{-1} exceed, respectively, 1000 and 3000 miles, some idea is gained of the average frequency of occurrence of sferics.

In English-speaking countries the most commonly used device for determining the direction of arrival is a direction-finding system, justly associated with the name Watson-Watt, which uses two stationary loop antennas at right angles to one another. Each loop characteristically receives the maximum signal in a direction along its own plane, the signal intensity at angles of arrival other than zero varying as the cosine. The input circuits are tuned in the neighborhood of 10 kc sec^{-1}, corresponding to the quasi-frequencies of the lightning stroke. The signal in each loop is separately amplified and then added vectorially on an oscilloscope screen to produce a straight line representing the direction of arrival with 180° ambiguity. More recently, ambiguity of azimuth has been eliminated electronically. The signals are most often observed visually, less often recorded on stationary or continuously moving film. This form of sferics receiver has been used by numerous investigators in many countries [4, 11–13, 17, 32, 41]. (For an early analysis of such observations, see [39].) The Watson-Watt system was employed in networks of stations during World War II in England, the United States, the Caribbean, the western Pacific, China, Japan, and elsewhere. Radio or telephone intercommunication of stations comprising a network permits synchronized observation of the same instantaneous source. One station generally desig-

1. See "The Lightning Discharge" by J. H. Hagenguth, pp. 136–143.

nates the particular signal to be located, choosing those relatively distinct from others in the typically rapid sequence; indication of simultaneity and reliability is obtained in the plotting on gnomonic maps. Observations are generally made at two- or three-hourly intervals for periods of about half an hour. Base lines connecting stations may range in length from a few hundred miles to over a thousand.

A second direction-finding system, developed by Lutkin [20], also employs a pair of loops at right angles to one another, so arranged that a record of the incoming sferics signal is made only when the plane of one of the loops is directed toward the source. The pair of loops rotates at perhaps one revolution per minute, and the system provides a record on a drum turning with the loops. The receptive sector may be reduced to one degree of azimuth. Records obtained in this manner show direction of signal arrival as a function of time; a rough indication of the proximity of the source can be gained from the relative size and density of the marks made by the recording pen. The narrow sector system has been employed by Lugeon [16] and Bureau [8], among others. It provides very simply a record upon which statistics might be based, and rough conclusions may be drawn regarding the location of sources of sferics from observations made at several stations, or from the appearance and density of the record at one station. This system has found use in Switzerland, France, Australia, and other countries.

Other systems measure and record, as functions of time, the frequency of occurrence of sferics having an intensity greater than a present level (Bureau's radiocinemograph and Lugeon's atmoradiograph) [8], and the field strength of sferics (Lugeon and Nobile's radiomaximograph) [18]. Similar systems provide power companies with information on the proximity and course of intense local storms [29]. Rough determination of distance of sferics sources from single stations employing these and the narrow sector systems is made possible by astronomical methods involving the ozonosphere or ionosphere [14, 18]. Observation at a single station of the wave form of sferics indicating multiple reflections from the ionosphere also permits rough determination of distance [15, 23, 33]. Approximations of distance have also been based upon the assumption that the most intense discharges in all but the weakest storms have about the same average power [18].

Accuracy of determination of bearings of sferics varies diurnally because of changing polarization of the arriving radio wave. These errors are largest at night, and especially at sunrise and sunset [22, 23]; nevertheless, useful data may be obtained even then, although the efficiency of observation is reduced. Polarization errors are greatest in an intermediate range between a few tens of miles and some 1500 miles from the receiver, and are smaller at the very low than at the medium and high radio frequencies. Most other sources of error, due to nonhomogeneity of the site with respect to azimuth, and to instrumental and operational errors, can be made negligible. Use of antennas other than loops to decrease polarization errors is being investigated [1]. Regarding propagation, range, and attenuation of low-frequency radio waves, see, for example, [4, 5, 7, 10, 21]. In triangulation procedures with data obtained from direction finders, the errors of closure of polygons formed by laying out the several reported azimuths, together with observers' estimates of confidence in the value and synchronization of each azimuth, have been utilized empirically to assign estimates of accuracy to sferics fixes. Error charts based upon the geometry of a network with respect to any source of sferics have been constructed. These show which pairs of azimuths yield the smallest maximum error, under the assumption that there is a constant error (generally taken as $\pm$ 2°) in azimuth determination at each station. In network operation, the task of designating which sferics shall be observed is often rotated among the stations in the network, in order to avoid favoring a circle of detection about only one station, and to reduce the blanking effects of local storms. Probable errors are assigned for each fix based upon location with respect to the observing network, estimates of instrumental error, size and shape of triangle or polygon of intersection of bearings, and consistency. Significant contributions regarding selection of an optimum point for a fix, given several radio direction-finder bearings, have been made by Ross [28], Stansfield [34], and Barfield [3].

Of what value are sferics to meteorologists? Like radar, sferics instruments provide the equivalent of an extremely dense network of observing stations within the working range; the practical need for such high-density coverage depends upon the significance of the weather element(s) detected, the size of the area covered, and the density of the regular weather observational network over the area. Patterns of activity are more readily perceived than by use of point observations. In the present case, the forecaster insures that his map analysis, especially when based upon widely scattered observations, agrees with the fact of occurrence of vigorous convection at certain places in the pattern. Synoptic maps were revised on this basis as early as 1924 [4]. Petterssen and Berson [26] in a recent study of European and North Atlantic sferics reports and winter synoptic maps found concentrations of sferics sources at the apex of warm sectors of both nascent and occluding cyclones, in the former case even before the appearance of closed isobars at the surface. Maxima were also found on the cold front 400–500 miles from the apex, and secondary maxima 300–500 miles behind the cold front.

Study of the results of observations of a network of direction finders in South Africa led Schonland and his collaborators [32] to conclude that 70 per cent of sferics fixes could be verified by local observations of thunderstorms and precipitation, though perhaps more were real, but that the utility of sferics reports in forecasting tends to be marginal.

The performance of a United States direction-finding network consisting of stations in New Jersey, Florida, Newfoundland, and Bermuda has been studied [38] by comparison of sferics fixes with reports of thunder-

storms, lightning, and cumulonimbus from the relatively dense United States weather-reporting network east of the 95th meridian. For the six-month period, May to October, 1945, 87 per cent of 135 four-station fixes were within 100 miles of a thunderstorm, lightning, or cumulonimbus report within one hour of the sferics observation, and 68 per cent of 19,130 three-station fixes were similarly verified. On the other hand, of 11,006 reports of thunderstorms or lightning, 57 per cent were attended by fixes within 100 miles, and 80 per cent, within 200 miles. The diurnal variation averaged between ten and fifteen per cent on either side of the figures given, with poorest performance at 0600 and 0900 GMT, best at 2100 and 0000 GMT. A survey made in 1945 [38] of 98 U. S. Air Force weather stations located within the area of coverage of this network, selected only for having forecasting problems in the Gulf of Mexico, Caribbean, or western Atlantic, revealed that 42 per cent considered sferics reports over the oceans in middle latitudes to be "extremely useful," and 98 per cent considered them either "extremely useful" or "useful." Almost identical percentages resulted as regards sferics reports over oceans in the tropics. The average length of experience with sferics was nine months. In a series of airplane flights over the ocean bordering on the network, 81 per cent of the sferics reports near the flight path were attended by significant weather within 100 miles.

Interesting remarks on the use of sferics in weather forecasting in Switzerland, including a discussion of the implications of both active and quiet regions, may be found in [16]. The relation of sferics to the Indian monsoon may be found in [35, 36]. For examples of recent North Atlantic weather and corresponding sferics fixes, see [25]. Studies relating to the utility of sferics as regards hurricanes have thus far tended to be inconclusive or negative [30], although further work appears justified in view of reports from the western Pacific during World War II of their use in following the progress of easterly waves. Recently sferics have been studied in tornado investigations, with emphasis on analysis of wave forms.

More studies are needed relating broad meteorological patterns to the distribution of thunderstorms and other intense electrical activity occurring in regions of rapid charge separation. The physical nature of the lightning flash, its visible and photographic forms, the resulting field changes, the propagation of attending electromagnetic waves—these researches have been and are still emphasized. Consequently, sferics research has tended to appear in nonmeteorological publications. Doubtless room for further application of modern direction-finding techniques to sferics exists, if only because present systems are essentially those first employed two or three decades ago. However, if the field of sferics is to find important use in meteorology, studies of regional application of sferics observations to characteristic weather patterns must be undertaken; reference is made to the areal distribution of thunderstorms attending the large-scale flow patterns. Such patterns may include cyclogenesis, especially in regions off the east coasts of the continents. Presumably, departures from the normal in the disposition of sferics may yet be found in the neighborhood of developing tropical storms. As to the utility of sferics over land areas occupied by a dense meteorological observational network, the burden of proof rests with sferics, although the inexpensiveness of simple sferics receivers compared with radar sometimes recommends them for warning of the approach of local storms, especially during daylight hours when the detection of thunder is limited to a distance of the order of ten miles. Because of an essential requirement for a thunderstorm, that is, sufficient available moisture, sferics methods find negligible application in some areas, and only seasonal use in some others. The development of a synoptic climatology of sferics distribution over the oceans would contribute toward maximum utility of sferics methods. Great benefit should also be derived from the analysis of wave forms that indicates the relative frequency of cloud-to-cloud, cloud-to-air, and cloud-to-ground discharges.

Sferics study promises to refine our knowledge of the distribution of thunderstorms over the oceans. Brooks' classical work [6] was necessarily based largely upon ships' reports for the ocean, and it is expected that conclusions based upon this essentially small sample may be improved. A minor disadvantage in the use of sferics for this purpose lies in the loss in accuracy as one proceeds away from a central area or point, but this objection is not final as regards the refinement of our climatological knowledge over the oceans. The development of a climatology of radio static appears desirable [37]. It should become possible to estimate reasonably well for any given place on the globe the monthly, seasonal, and annual noise level, and its diurnal variation. Meteorologists may well assist in the interpretation of noise-level records made over the world. Specific results will depend upon increased acquaintance with the synoptic climatology of thunderstorm occurrence over the oceans.

Present sferics methods are time-consuming when it is desired to locate fixes accurately and to obtain a large sample. Photography has mostly found application for research purposes. Rapid techniques for handling the considerable volume of information obtained even with a simple direction finder will increase the utility of sferics in applications to forecasting. With small sample methods, the reliability of a sferics observation as an indication of electrical activity is greater than that of a negative observation as an indication of quiet conditions. Electronic methods are apparently available for the presentation of the areal distribution of electrical discharges continuously on oscilloscope screens [27]. The electronic determination of frequency distribution of azimuths of discharges above predetermined levels of intensity also seems feasible. It should be possible to present the results of a complete sferics observation made at map time well before the map is plotted and analyzed. Finally, it would be helpful to investigate the proper length of observational sample necessary in order to obtain a reasonably complete picture of the sferics activity occurring at a given time.

The expectation that sferics will become a more familiar tool for the meteorologist appears justified.

REFERENCES

1. Adcock, F., and Clarke, C., "The Location of Thunderstorms by Radio Direction-Finding." *J. Instn. elect. Engrs.*, 94(3):118–125 (1947).
2. Appleton, E. V., and Chapman, F. W., "On the Nature of Atmospherics—IV." *Proc. roy. Soc.*, (A) 158:1–22 (1937).
3. Barfield, R. H., "Statistical Plotting Methods for Radio Direction-Finding." *J. Instn. elect. Engrs.*, 94(3A):673–675 (1947).
4. Boswell, R. W., and Wark, W. J., "Relation between Sources of Atmospherics and Meteorological Conditions in Southern Australia during October and November, 1934." *Quart. J. R. meteor. Soc.*, 62:499–527 (1936).
5. Bracewell, R. N., "The Propagation of Very Long Radio Waves." *J. Instn. elect. Engrs.*, 95(3):326 (1948). (A summary.)
6. Brooks, C. E. P., "The Distribution of Thunderstorms over the Globe." *Geophys. Mem.*, Vol. 3, No. 4 (1925).
7. Budden, K. G., Ratcliffe, J. A., and Wilkes, M. V., "Further Investigations of Very Long Waves Reflected from the Ionosphere." *Proc. roy. Soc.*, (A) 171:188–214 (1939).
8. Bureau, R., "Les foyers d'atmosphériques." *Mém. off. nat. météor., Paris*, No. 25 (1936).
9. Cave, C. J. P., and Watson-Watt, R. A., "The Study of Radiotelegraphic Atmospherics in Relation to Meteorology." *Quart. J. R. meteor. Soc.*, 49:35–39 (1923).
10. Committee on the Relation between Atmospherics and Weather, "The Range of Atmospherics." *Quart. J. R. meteor. Soc.*, 53:327–400 (1927).
11. Harper, A. E., "Some Measurements of the Directional Distribution of Static." *Proc. Inst. Radio Engrs., N. Y.*, 17:1214–1224 (1929).
12. Henderson, J. T., "Direction Finding of Atmospherics." *Canad. J. Res.*, 13(A):34–44 (1935).
13. Kessler, W. J., and Knowles, H. L., "Direction Finder for Locating Storms." *Electronics*, 21:106–110 (1948).
14. Khastir, S. R., Das Gupta, M. K., and Ganguli, D. K., "Location of Thunderstorm Centres from Directional Observations of Atmospherics during Sunrise and Sunset." *Nature*, 159:572–573 (1947).
15. Laby, T. H., and others, "Wave Form, Energy and Reflexion by the Ionosphere, of Atmospherics." *Proc. roy. Soc.*, (A) 174:145–163 (1940).
16. Lugeon, J., "Le radiogoniographe de la Station centrale suisse de météorologie et son utilisation pour la prévision du temps." *Ann. schweiz. meteor. Zent-Anst.* (1939).
17. —— Report for the Toronto Meeting, International Commission for Synoptic Weather Information, Sub-Commission of Sferics, CIR/IMO/T-21, July 18, 1947.
18. —— et Nobile, G., "Le radiomaxigraphe enregistreur d'intensité des parasites atmosphériques de la Station centrale suisse de météorologie." *Ann. schweiz. meteor. Zent-Anst.* (1938).
19. Lutkin, F. E., "The Nature of Atmospherics—VI." *Proc. roy. Soc.*, (A) 171:285–313 (1939).
20. —— "Lightning and Atmospherics." *Quart. J. R. meteor. Soc.*, 67:345–350 (1941).
21. Mimno, H. R., "The Physics of the Ionosphere." *Rev. mod. Phys.*, 9:1–43 (1937).
22. Munro, G. H., and Huxley, L. G. H., *Atmospherics in Australia—I.* Australian Radio Res. Board Rep. No. 5, 1932.
23. Munro, G. H., Webster, H. C., and Higgs, A. J., *Simultaneous Observations of Atmospherics with Cathode-Ray Direction Finders at Toowoomba and Canberra.* Australian Radio Res. Board Rep. No. 8, pp. 9–42, 1935.
24. Norinder, H., "On the Measurement and Nature of Atmospherics Produced by Electric Discharges in Snow Squalls and from Other Sources." *Tellus*, Vol. 1, No. 2, pp. 1–13 (1949).
25. Ockenden, C. V., "Sferics." *Mar. Obs.*, 19:199–206 (1949).
26. Petterssen, S., and Berson, F. A., *Atmospherics in Relation to Fronts and Air Masses.* Reprint, U. S. Office of Chief of Naval Operations, NAVAER 50-1T-38.
27. Pick, L. A., *A New Method for Locating Thunderstorms.* Paper read at Meeting of the Amer. Meteor. Soc., Washington, D. C., April 29, 1947 (unpublished).
28. Ross, W., "The Estimation of the Probable Accuracy of High-Frequency Radio Direction-Finding Bearings." *J. Instn. elect. Engrs.*, 94(3A):722–726 (1947).
29. Ruedy, R., *The Distribution of Thunderstorms and the Frequency of Lightning Flashes*, 2nd ed. Nat. Res. Council of Canada, N.R.C. No. 1282, Ottawa, 1945.
30. Sashoff, S. P., and Roberts, W. K., "Directional Characteristics of Tropical Storm Static." *Proc. Inst. Radio Engrs., N. Y.*, 30:131–133 (1942).
31. Schonland, B. F. J., Hodges, D. B., and Collens, H., "Progressive Lightning. V—A Comparison of Photographic and Electrical Studies of the Discharge Process." *Proc. roy. Soc.*, (A) 166:56–75 (1938).
32. Schonland, B. F. J., and others, "Direction-Finding of Sources of Atmospherics and South African Meteorology." (With a commentary by N. P. Sellick, J. S. Peake, and R. A. Jubb.) *Quart. J. R. meteor. Soc.*, 66:23–43 (1940).
33. —— "The Wave Form of Atmospherics at Night." *Proc. roy. Soc.*, (A) 176:180–202 (1940).
34. Stansfield, R. G., "Statistical Theory of D. F. Fixing." *J. Instn. elect. Engrs.*, 94(3A):762–770 (1947).
35. Subba Rao, N. S., "Atmospherics during the Monsoon Period." *Proc. Ind. Acad. Sci.*, (A) 17:83–105 (1943).
36. —— "A Further Study of Atmospherics during the Monsoon Period." *Proc. Ind. Acad. Sci.*, (A) 18:127–139 (1943).
37. Thomas, H. A., "The World Distribution of Radio Noise." *J. Instn. elect. Engrs.*, 95(3):332–333 (1948). (A summary.)
38. Wanta, R. C., *Forecaster's Introduction to Sferics.* Paper read at Meeting of American Geophysical Union, Washington, D. C., March, 1946 (unpublished).
39. Watson-Watt, R. A., "Directional Observations of Atmospherics, 1916–1920." *Phil. Mag.*, (6) 45:1010–1026 (1923).
40. —— "Abstracts of Papers on the Meteorological Relations of Atmospherics." *Quart. J. R. meteor. Soc.*, 52:199–209 (1926).
41. —— "Weather and Wireless." *Quart. J. R. meteor. Soc.*, 55:273–301 (1929).

MICROSEISMS

Observations and Theory of Microseisms *by B. Gutenberg* 1303

Practical Application of Microseisms to Forecasting *by James B. Macelwane, S. J.* 1312

OBSERVATIONS AND THEORY OF MICROSEISMS

By B. GUTENBERG

California Institute of Technology

Observations and Causes of Microseisms

As soon as fairly sensitive seismographs were available, it was found that the ground is never at rest. Various terms such as *pulsations*, *pulsatory oscillations*, *microseismic movements*, and *microseisms* were used to describe these continuous small movements which are not caused by earthquakes. Hecker, in 1906, divided the microseisms, as they are now usually called, into four groups: (1) those with periods of less than 4 sec, (2) those with periods of about 7 sec, (3) those with periods of about 30 sec, and (4) those with periods of about one minute and more. Later, additional types of microseisms were described. For a bibliography, see [4].

There is little agreement about the causes of microseisms. This is partly due to the fact that at one station certain types of microseisms prevail and are described; at another station, different types. In addition, the seismographs in use have widely different characteristics so that some emphasize waves with periods of a fraction of a second, others waves with periods of 10 sec or even more. (Compare Figs. 1*c* and 1*f*.) The fact that there are several types of microseisms which differ rather little in their appearance has added to the confusion. Many authors have claimed that hypotheses concerning causes of microseisms given by other seismologists are incorrect while, actually, completely different types of microseisms were involved in the discussions. In Table I a summary of some of the more common types of microseisms is presented, and Fig. 1 shows some typical records.

No movements of the ground due to traffic and industry (which are usually also called microseisms) are included here. In addition, "microseisms" of Type 10 (Table I) are frequently caused by the direct effect of air currents in the vault on the instruments, and Type 11 probably contains instances where subfreezing temperatures caused freezing of moisture on the instrument and spurious "microseisms." Movements with periods of more than 1 minute (for example from changing loads at coasts during high seas or high tides) are now usually classified as "tilt." In general, irregular and short-period microseisms are connected with rather local causes. On the other hand, Types 6 and 7 have been recorded thousands of miles from the source, for example in Irkutsk, Central Asia, during storms near the Norwegian coast.

Very little is known about Type 1 (Table I). These microseisms are under investigation by Father Macelwane and his collaborators [25]. Microseisms of Type 2 are observed at stations near coasts and are caused by local surf [15, p. 269].

Types 3 and 4 are probably of local origin. They have been attributed to effects of windstorms and cold fronts near the station, but no detailed theory has been published on the mechanism by which they are produced. They are being investigated at Fordham [24] and at Pasadena. Caloi [5] has found that microseisms with periods of 2–3 sec at Trieste are correlated with seiches in the Gulf of Trieste which in turn are caused by low pressure centers.

Types 5 to 8 usually originate at greater distance

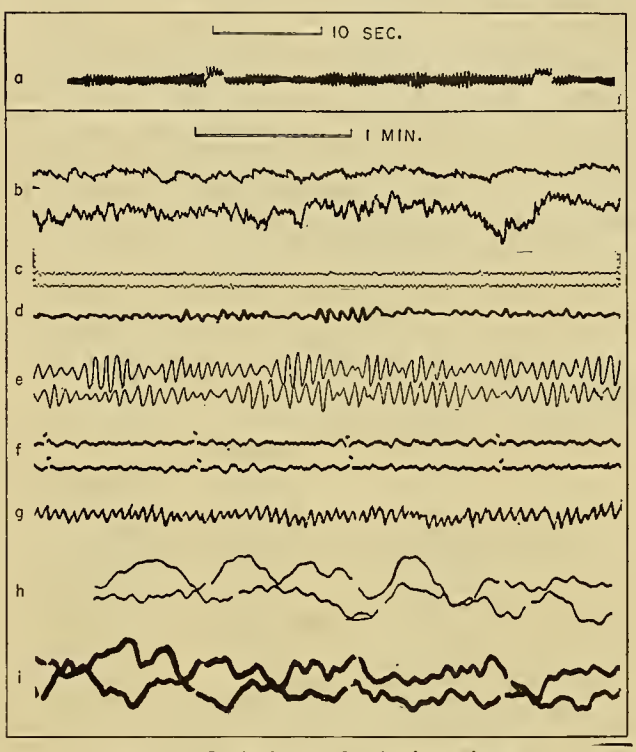


FIG. 1.—Typical records of microseisms.

a. Short-period microseisms (Type 1, Table I), Wood-Anderson torsion seismograph, Florissant, Missouri. (*After Macelwane.*)

b. Microseisms from local surf at Helgoland, January 18, 1918; Wiechert 200-kg horizontal seismograph; successive lines are about 1 hr apart. (*After Gutenberg.*)

c. Microseisms of Type 3 (Table I), recorded at Pasadena on December 26, 1948; short-period Benioff vertical seismograph.

d. Microseisms during windstorm at Göttingen, November 13, 1909; 1200-kg vertical Wiechert seismograph. (*After Gutenberg.*)

e. Microseisms recorded at Zi-ka-wei near Shaughai from typhoon over the ocean, October 3, 1923; Galitzin seismograph; lines are about ½ hr apart. (*After Gherzi.*) Records of Type 6 or 7 are frequently similar to this record.

f. Microseisms of Type 6 or 7 (Table I), recorded simultaneously with Fig. 1*c* at Pasadena by long-period Benioff vertical seismograph.

g. Microseisms during monsoon recorded at Zi-ka-wei on November 22, 1922; Galitzin seismograph. (*After Gherzi.*)

h. Microseisms of Type 10 (Table I). (*After Whipple.*)

i. Microseisms of Type 11 (Table I), recorded at Zi-ka-wei on January 4, 1917; Galitzin seismograph. (*After Gherzi.*)

from the stations. Microseisms of Type 5 have been studied recently in more detail because of their economic importance. They are probably caused by high ocean waves in hurricanes and typhoons. Most of the available theory is applicable to these microseisms.

At most stations more or less regular sinusoidal waves are recorded with periods between 4 and 10 sec, and such microseisms have been described in several hundred papers [4]. Nearly everywhere they have well-expressed maxima during winter and frequently are scarcely noticeable on typical records during summer. At some stations a slight increase during the night has been reported. Except for Type 9, it is generally pointed out that these microseisms are connected with windstorms over the ocean (Fig. 2), but there is no agreement concerning the mechanism by which the energy is transmitted from the storm to the ocean bottom and frequently it is not clear whether Type 6, 7, 8, or still another type (*e.g.*, those attributed to strong winds blowing against coastal mountains) is involved. According to the hypothesis stated in 1903 by E. Wiechert and developed by Gutenberg [15], ocean waves driven by storms toward steep coasts are the source of Type 7. On the other hand, Gherzi [11] believes that fast changes in air pressure (pumping) are a major cause of such microseisms. However, there is some doubt whether Type 8 actually exists, since air-pressure changes amount to only a small fraction of one atmosphere while the pressure changes produced by storm waves in the ocean are of the order of magnitude of one atmosphere.

It is very likely that a mechanism similar to that which produces microseisms during hurricanes also operates in extratropical disturbances. In this case the energy of the storms available in a given area is smaller, but the region affected by it is much more extended. The resulting microseisms are Type 6. On the other hand, Banerji [1] believes that the motion of gravity waves in water is large enough to reach the ocean bottom and cause microseisms of Type 6. Finally, similar microseisms are caused by monsoon winds [1] (Type 9).

In one of the most comprehensive of all studies of microseismic records and their relationship to meteorological phenomena, Båth [2] has shown that the microseisms with periods of 4 to 10 sec recorded at Uppsala, Sweden, belong to at least two types: one due to surf driven against the steep Norwegian coast; the other to cyclones in the North Atlantic which transmit a small fraction of their energy into the water and the ocean bottom. In instances like this, when microseisms from different sources arrive simultaneously at a station, waves with slightly different periods produce interference patterns; as a consequence, wave groups with gradually increasing and decreasing amplitudes are frequently found at certain stations, but rarely at others.

Great progress has been made in the study of Types 5, 6, and 7 by the use of tripartite stations which are described in the article in the Compendium by Father Macelwane.[1] For papers giving summaries of the literature concerning Types 6 to 9 see, for example, [12, 15, 18, 19, 32, 36].

It is possible that microseisms with longer periods include movements produced by strong winds (Type 10). These differ from Type 4 by their much greater periods. Possibly related to them are the compressional movements of more irregular appearance which are recorded by the strain seismographs at Pasadena during windstorms as well as during the presence of convection currents. These movements seem to be the result of compression and dilation of the hill on which the Seismological Laboratory is built.[2] Type 11 includes irregular movements, with periods of the order of one minute, which seem to be connected with freezing of the ground [15, pp. 294–298].

Table I. Types of Microseisms

Type no.	Period (sec)	Kind of movement	Hypothetical causes	Relative distance of causes	Figure
1	0.2–0.5	Regular	?	Local?	1*a*
2	0.2–2	Irregular	Surf	Local	1*b*
3	1–4	Regular	Fronts? Rain? Wind?	Local	1*c*
4	1–4	Irregular	Wind	Local	1*d*
5	2–6	Regular	Ocean waves in hurricanes, typhoons	Distant	1*e*
6	4–10	Regular	Ocean waves in extratropical disturbances	Distant	1*f*?
7	4–10	Regular	Surf driven by wind against steep coasts	Distant	1*f*?
8	4–10	Regular	Air-pressure pulsations?	Medium?	—
9	4–10	Regular	Monsoon and other types of wind	Medium?	1*g*
10	20–100	Irregular	Wind? Air currents in instrument vault?	Local	1*h*
11	40–200	Irregular	Freezing of ground? "Icing" of instruments?	Medium	1*i*

Transmission of Energy from Meteorological Disturbances into the Ground

There is no type of microseism for which a complete theory has been worked out. The few theories which have been formulated mathematically cover only certain phases of the process. These theories may be divided into two groups. The first group deals with the transmission of energy from the meteorological disturbance into the ground, the second with the propa-

1. "Practical Application of Microseisms to Forecasting" by J. B. Macelwane, S.J., pp. 1312-1315.

2. See *Bull. Amer. meteor. Soc.*, 20:424, Fig. 3 (1939).

gation of the microseismic waves in the ground to the point of observation.

Energy Transmitted from Surf. The energy transmitted from surf to the coast [14] depends on the energy of these waves is transferred to the coast; f depends on the steepness of the coast, the loss of energy due to friction in the water approaching the coast, and the wind. The kinetic energy E^* (per second) of ocean

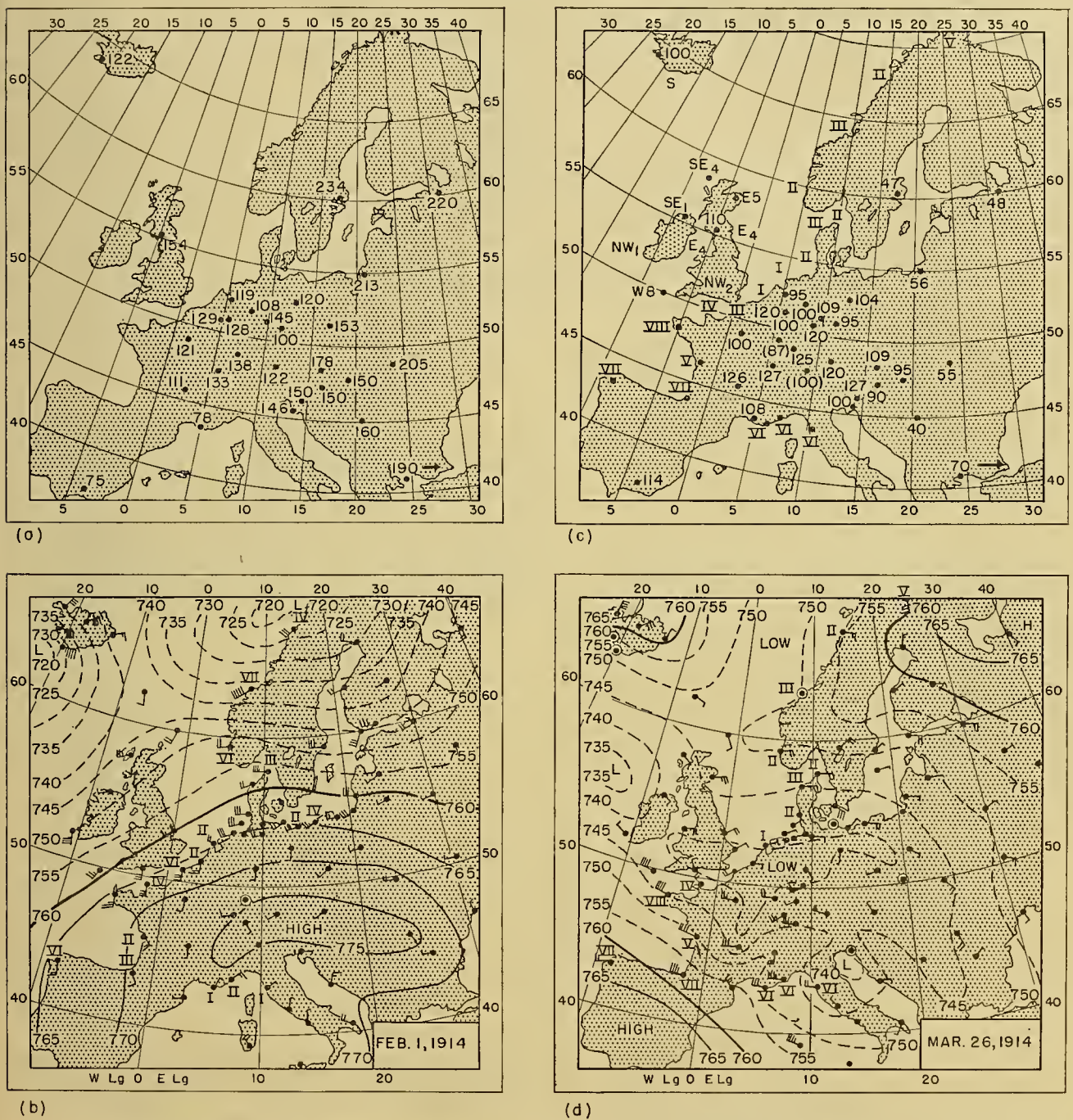


FIG. 2.—Relative amplitudes of microseisms (in per cent of the individual average amplitudes for each station for the same days with winter microseisms) and corresponding weather maps: (*a*) and (*b*) on Feb. 1, 1914; (*c*) and (*d*) on March 26, 1914. State of swell indicated by Roman numerals (on a scale from 0 to IX). (*After Gutenberg.*)

kinetic energy of the waves, the type and slope of the coast, the force of the wind, and its direction relative to the shore. The greatest energy will be transferred to the coast when a strong wind drives high ocean waves against a steep, rocky shore. Only a fraction f of the waves (period T, length L, velocity V, and height H is given approximately by

$$E^* = \frac{gLH^2}{16T} = \frac{gVH^2}{16}. \qquad (1)$$

The energy which is transferred to a coast of length l is given by

$$E = \frac{fgVH^2l}{16}. \quad (2)$$

During storms this may reach the order of $10^{18}f$ ergs sec^{-1}. The amplitude a of the corresponding microseisms with a period T and a wave velocity C at a distance D from the source is then given approximately [14] by

$$a^2 = \frac{TE}{5 \times 10^{11}C^2 \sin D}. \quad (3)$$

If $T = 7$ sec, $C = 3$ km sec^{-1}, $D = 10°$, $E = 10^{18}f$ ergs (as found above), and $a = 1$ micron $= 10^{-3}$ mm, f must be of the order of 10^{-3}, that is, one-tenth of one per cent of the wave energy must be transferred to the coast in order to explain the microseisms by the surf action. This value is not unreasonable. No arguments against this theory have been published and the theory explains the microseisms of Type 7.

Theory of Propagation of Waves from the Surface of the Ocean to the Bottom. In an attempt to calculate the energy transmitted from ocean waves at the surface to the ocean bottom, Scholte [35] supposed that waves of period T exist at the surface of the ocean. He started with the following equation given by Sezawa [37]:

$$\frac{\partial^2 \mathbf{r}}{\partial t^2} = c^2\nabla(\nabla \cdot \mathbf{r}) + g\nabla w, \quad (4)$$

which connects a displacement $\mathbf{r}$ (components u, v, w) in the water with the velocity c of its propagation and includes gravity waves as well as compressional waves. Scholte found under reasonable assumptions that near the surface of the ocean the ratio of the amplitudes of the gravitational and the elastic waves is of the order of 10^4 to 1. If the gravity waves at the surface have a height of 10 m, the elastic waves in the water would have an amplitude of the order of 1 mm. However, the gravity waves decrease exponentially with depth; the elastic waves change relatively little. Consequently, gravity waves should not be noticeable at great depths in the ocean (this result confirmed theoretical conclusions of earlier investigators), but elastic waves should remain perceptible with sensitive instruments down to the bottom of the ocean where they start the usual seismic waves in the surface layers. However, Scholte did not describe this latter process in detail. His paper is the only one giving quantitative results for the elastic microseismic waves at the ocean bottom.

Theory of Propagation of Elastic Waves in a Medium Consisting of an Upper Layer of Water Overlying a Homogeneous Layer of Solid Rock. Press and Ewing [31] have studied the propagation of elastic surface waves originating at the surface of an ocean with a homogeneous bottom layer. They have used an equation developed by Stoneley [38] which may be written in the following form:

$$\tan\left(\frac{2\pi AH}{L}\right) = \frac{\rho A[4(1-B)^{\frac{1}{2}}(1-C)^{\frac{1}{2}} - (2-B)^2]}{B^2(1-C)^{\frac{1}{2}}}, \quad (5)$$

where

$$A = \left[\left(\frac{c}{w}\right)^2 - 1\right]^{\frac{1}{2}}, \quad B = \left(\frac{c}{v}\right)^2, \quad C = \left(\frac{c}{V}\right)^2, \quad (6)$$

H = depth of ocean; L = wave length (measured in a horizontal direction) of waves formed by constructive interference between elementary waves of length l which undergo multiple reflections at the boundary of the liquid layer at an angle of incidence i; c = velocity of microseismic waves; w = velocity of sound in water (density 1.0); V and v are the velocities of longitudinal and transverse body-waves, respectively, in the material below the ocean which has the density ρ and is assumed to extend down to infinite depth. Scholte, too, has found and used this equation [35, equation bottom p. 674] apparently without recognizing the equation as Stoneley's. If the velocity c is assumed, then H/L can be calculated from equation (5). If the depth H of the ocean is very small as compared with the wave length L, equation (5) becomes the well-known equation for Rayleigh waves in the surface of the solid bottom. For a more detailed theory, see Pekeris [30].

Press and Ewing assumed, for example, $w = 1\frac{1}{2}$ km sec^{-1} (velocity of elastic waves in water), $V = 5.3$ km sec^{-1} and $v = 2w = 3.0$ km sec^{-1} in the ocean bottom where the density $\rho = 2.5$, and Poisson's ratio equals $\frac{1}{4}$. In order to get an exponential decrease of amplitudes with depth, that is, surface waves, A in equation (6) must be real ($c \geqq w$). Under their assumptions (which include $l = 1.5T$) the curve giving the wave velocity as a function of H/l consists of two branches (Fig. 3).

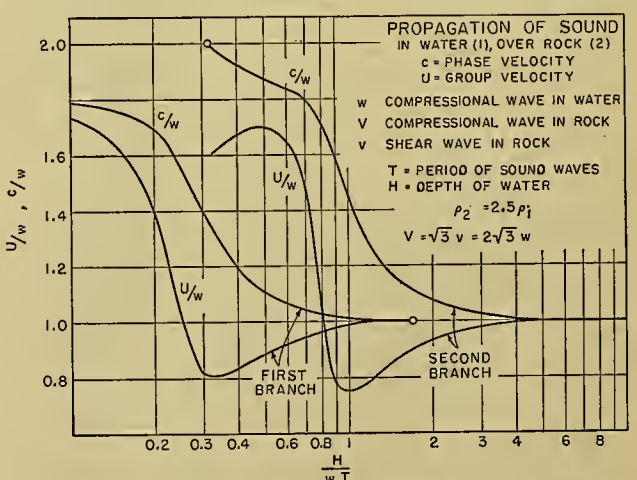


FIG. 3.—Theoretical phase and group velocity in a system consisting of a liquid layer overlying an infinitely thick homogeneous solid (under the specified assumptions). (*After Press and Ewing* [31].)

For the first branch, c' approaches the velocity of Rayleigh waves in the ground if H/l is small. Under the assumptions made in the figure, this requires an ocean depth of not over a few hundred meters for waves with periods of 3 sec. If the ocean depth H increases beyond the wave length ($l = 1.5T$) of the elastic waves in water, the velocity of the microseisms approaches the velocity of sound in water.

The second branch in Fig. 3 begins if the ocean depth exceeds a value of approximately $l/3$ (or about $\frac{1}{2}$ T), which is normally more than 1 km. The corresponding wave velocity of the microseisms is about twice the sound velocity in water or about 3 km sec^{-1}. If the ocean depth is smaller than the critical depth, the value of A in equation (6) becomes imaginary; the waves which correspond to this part of the branch are no longer surface waves. For $H = 0$, this corresponds to the second (physically nonexistent) branch of the solution for Rayleigh waves in a homogeneous layer. If the ocean depth exceeds a value of about four times the wave length, the velocity of the microseisms again becomes nearly equal to the velocity of longitudinal waves in water.

Press and Ewing have pointed out that (as in other seismic surface waves) the group velocity and not the wave velocity should be investigated in a study of the amplitudes of continuous waves. The group velocity U is given by the equation,

$$U = c - L \frac{\partial c}{\partial L}. \tag{7}$$

Under the assumptions of Fig. 3, there is a minimum group velocity in the first branch and a maximum as well as a minimum group velocity in the second branch. All three values can be expected to be associated with an increase in amplitude in the course of time; under the assumptions used in Fig. 3, Press and Ewing find the values given in Table II. All three values for T are

Table II. Periods for Group Velocity Maxima and Minima
(*After Press and Ewing* [31])

Branch	Group velocity U	U/w	H/L	T for $H = 4\frac{1}{2}$ km (sec)
First	Minimum	0.8	0.33	9.1
Second	Maximum	1.7	0.49	6.1
Second	Minimum	0.75	0.98	3.1

within the range which is observed in microseisms. However, although the assumed ocean depth $H = 4\frac{1}{2}$ km is a fair average in many instances, it must be considered that in many oceanic areas the depth is much less and consequently the periods of the microseisms should vary greatly depending on H. Such a great variation is not observed. However, the theory by Press and Ewing is a first approximation only; among other factors, the rather rapid increase in velocity with depth in the ocean bottom must be expected to affect the numerical results considerably. The effect of friction of the water on the sea bottom has no appreciable effect according to Menzel [26].

Scholte has not considered the effects of group velocity. However, since he discusses the vibrations of the ocean bottom relatively close to the source, the effect of the group velocity is probably negligible; it takes many wave lengths before the accumulation of energy becomes appreciable in waves with periods near those of the maximum or minimum group velocity. However, effects of energy accumulation near extreme values of the group velocity have to be considered in the propagation of surface waves in the crustal layers. This has not been discussed by Scholte.

Theoretical Periods of Pressure Waves in the Ocean. Comparison between the observed periods of microseisms and the simultaneous periods of ocean waves by Bernard [3] indicate that the periods of the microseismic waves frequently are about one-half of the periods of the corresponding ocean waves. Darbyshire [6, 7] confirmed the observations of Bernard by using examples where a depression had a distance of the order of 1000 miles from the west coast of the British Isles and comparing the ocean waves which arrived at the coast of Cornwall with the microseisms recorded at Kew. On the other hand, S. M. Mukherjee (unpublished manuscript) found that the periods of sea waves during the southwest monsoon in India are definitely not double the periods of microseisms.

In detailed theoretical investigations Miche [27] has shown that in a stationary wave motion there are second-order pressure variations of twice the wave frequency. These variations are not attenuated to zero with depth. On the other hand Longuet-Higgins [23] (assuming incompressibility) found that beneath two wave trains with the same frequency travelling on the ocean in opposite directions, pressure fluctuations result with a frequency double that of either wave and with amplitudes proportional to the product of the wave amplitudes. Such mechanisms could produce the pressure changes near the surface which are assumed by Scholte in his theory [35].

All the mathematical theories mentioned thus far probably give a rough approximation, each from a different angle, to a part of the causative mechanism which operates in microseisms. The theories of Miche, Longuet-Higgins, and Bernard could be combined with that of Scholte or that of Press and Ewing. In the latter theory, the effect of the maximum and minimum group velocities on the amplitudes would be smaller if the source of the microseisms had strongly prominent periods. Such periods (possibly somewhat modified) should also appear in the microseisms.

Theory of Propagation of Elastic Waves in the Ground

Since microseisms are observed at stations on land, some theoretical results on the propagation of elastic waves in the earth's crustal layers are needed for a discussion of microseisms.

Wave Types Observed in Microseisms. There are two major groups of seismic waves: (1) body waves (travelling through the interior of the material), and (2) surface waves. Group (1) contains longitudinal and transverse waves which have no noticeable dispersion, so that group velocity and wave velocity are practically the same. Group (2) includes Love waves (surface-shear waves) in which the particles move in the surface of the earth perpendicular to the direction of propagation, and Rayleigh waves in which, theoretically at least, the particles describe ellipses with their longer axis in the vertical direction and their shorter axis in the direction of propagation. In seismograms produced by disturb-

ances near the surface (but not from explosions) the surface waves are usually more prominent than the body waves.

The first attempt to find the type of movement which prevails in microseismic waves was made by Geussenhainer [10]. He found that in Göttingen the period of the vertical component of the microseisms normally changed gradually with time while in the horizontal component periods of 6, 7½, and 9 sec occurred more often than periods with intermediate values. If waves with one of these preferred periods existed in the vertical component, the corresponding waves had especially large amplitudes in the horizontal components; if the period of the microseisms in the vertical component was between two fundamental periods of the horizontal, both fundamental periods, sometimes changing repeatedly from one to the other, were recorded in the horizontal components. Geussenhainer believed that the vertical pendulums record mainly free vibrations of the earth's crust. Such free vibrations would require additional theoretical considerations since the resulting movements may be very different from those calculated for forced vibrations, especially if resonance phenomena are involved.

The observed velocities of microseisms are usually between about 2 and 4 km sec^{-1}; that is, in the range of velocities of surface waves with such periods. It is generally believed that the microseisms consist of surface waves. Unfortunately, measurements of the amplitude of microseisms as a function of depth in order to find the expected decrease are inconclusive [15, 19, 20]. The few authors who have studied the question of the wave type in microseisms agree that Rayleigh waves prevail [16, 22, 32]. On the other hand, waves of the Love type are also involved, although their amplitudes are relatively smaller.

Increase in Period with Distance. It has been found that the period of elastic waves usually increases as they are propagated. For example, in Eurasia the periods of microseisms frequently lengthen by several seconds as the waves are propagated from the Atlantic Coast into Asia. This increase is not due necessarily to a stronger absorption of the shorter periods but may be due to an actual increasing of the wave lengths with distance. Among attempts at a theoretical explanation are the wavelet theory by Ricker [33, 34], a theory developed by Munk [28], and a theory by Sezawa [37] (based on the assumption that the increase in period is due to internal friction) which has been extended by Gutenberg [13] and by Gutenberg and Schlechtweg [17]. This theory leads to the following equation for the period T at a distance D from the source:

$$T^2 = T_0^2 + \frac{5\eta D}{\rho V^3}, \tag{8}$$

where η = coefficient of internal friction (about 10^9 poises), ρ = density, V = wave velocity, and T_0 = period near the source. Values of T calculated from equation (8) agree well with the observed increase in period. For microseisms Gutenberg [14] found approximately

$$T^2 = T_0^2 + 0.01D, \tag{9}$$

where D is expressed in kilometers. The increase in period with distance makes it difficult to use microseisms recorded at a distance D from the source to find the period T_0 which prevails near the source.

Effect of Differences in the Structure of the Earth's Crust and of Faults on Microseisms. Microseisms in general have wave lengths of less than about 25 km; since their energy decreases exponentially with depth, they should be propagated mainly within the uppermost 25 km of the earth's crust, and short-period microseisms should be propagated in even a thinner layer. (This is the main reason for the observed dispersion of surface waves.) In geologically disturbed areas the loss of energy due to absorption is greater. This was recognized first for Europe (Fig. 4), where the regions in which the microseisms are relatively large for a given location of the source agree relatively closely with the geological units [15]. For example, microseisms originating in Scandinavia are propagated far into Asia without much loss of energy, but their amplitudes decrease considerably in southerly directions where surface layers of different geological ages and different physical constants are passed by the waves.

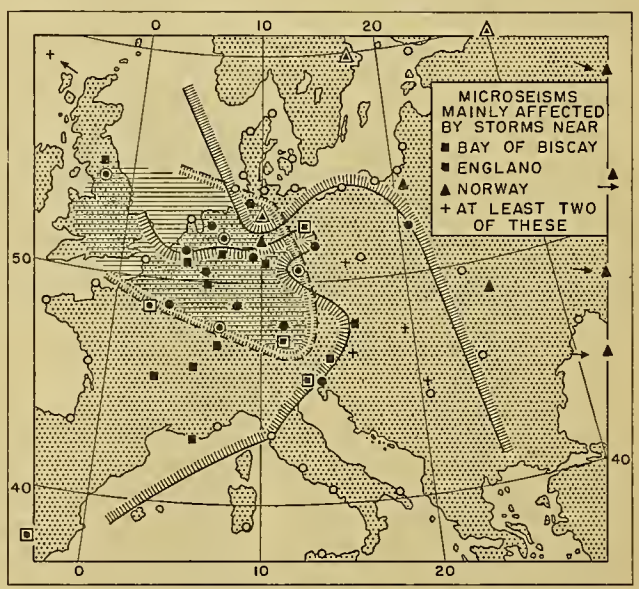


Fig. 4.—Areas with relatively large microseisms in Europe for three different locations of storm centers. (*After Gutenberg.*)

In the Pacific, microseisms with relatively large amplitudes are recorded if the station and the source of the microseisms are on the same side of the andesite line. This line (sometimes called the Marshall line) is the surface trace of a discontinuity (considered to extend to a depth of at least 30 km) which separates the less andesitic material of the upper layers in the bottom of the "Pacific Basin" from the more andesitic material with different physical properties on the continental side. In the western Pacific, the andesite line follows

the ocean deeps off Kamchatka and Japan, then runs to the east of the Marianas and of the Palau Islands and then southeastward off New Guinea and the Solomon Islands. Japanese scientists have pointed out that typhoons to the east of Japan are accompanied by large microseisms at the stations along the east coast of Japan, but only by insignificant microseisms on the west coast of Japan. The station operated at Guam under the project of the United States Navy Department is close to the andesite line. It records microseisms from typhoons which are centered far away in the ocean

Much smaller effects are to be expected from displacements along faults inside the major crustal blocks. Such faults cover greater parts of the earth's surface than is generally believed by nongeologists. Amplitudes of microseisms passing them (*e.g.*, along the Pacific coast of the United States) are not much reduced, especially in instances where the velocities on both sides of the fault zone do not differ much, thus contrasting with the deep structural discontinuities of the type represented by the andesite line surrounding the Pacific Basin. Microseisms are propagated across large parts

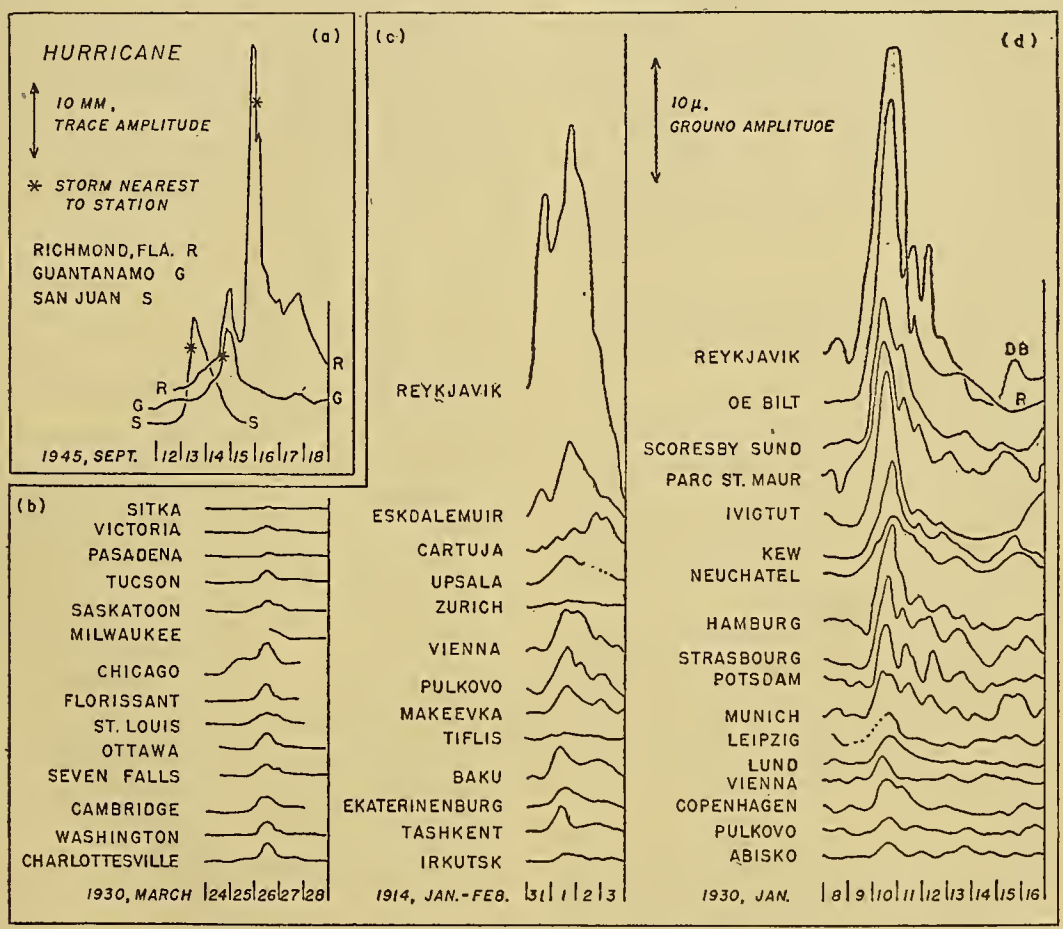


FIG. 5.—Amplitudes of microseisms: (*a*) in the Caribbean during a hurricane (*after Gilmore*), (*b*) in North America during a storm in eastern Canada (*after Gutenberg*), (*c*) in Europe and Asia during a storm approaching northern Norway (*after Gutenberg*), (*d*) in Europe and Greenland during a storm north of Scotland (*after Lee*). (*Taken from Gutenberg* [16].)

southeast of Japan; there is some indication that typhoons moving westward south of Guam produce relatively irregular and small microseisms at Guam until they cross the andesite line in the Caroline Islands.

The Caribbean loop is usually considered to be the boundary of a block with "Pacific" (less andesitic) material in the interior of the Caribbean. This explains why records of microseisms produced by hurricanes in the Caribbean area and recorded there at stations of the United States Navy Department show that the amplitudes of the microseisms decrease rapidly from one island to the other (Fig. 5). For other regions, see [5, 20, 21, 29].

of the United States without much loss of energy beyond regular absorption (Fig. 5b).

Effect of the Ground on Microseisms. Amplitudes of elastic waves at the surface of the earth depend appreciably upon the ground on which the instrument is located [15, p. 272; 20; 21]. For example, ground saturated with water is more likely to vibrate with large amplitudes than is solid rock, thus providing "increased sensitivity." However, the greater complexity of the records from such stations is a disadvantage and frequently more than offsets the advantage of the greater amplitude. For example, the station installed at Corpus Christi, Texas, in 1946 on loose sand recorded such

large microseisms of irregular types that it had to be discontinued. Solid rock is most suitable for the installation of seismographs for practically all purposes.

Desirable Lines for Future Research

The empirical and theoretical results concerning microseisms can be considered to be first approximations at best. While some progress can be expected from various extensive projects now under way, many important studies remain to be undertaken. Among those in which progress would be most helpful are the following, many of which offer no practical difficulties but have not been investigated because of the high cost:

1. Empirical

a. More detailed studies (similar to those described in [2]) on types of microseisms and their causes (leading to improvement of Table I).

b. Study of wave types and characteristics of observed microseismic waves (for all types in Table I), and determination of the "spectrum" of microseisms at a variety of locations.

c. Determination of the direction of approach of microseismic waves of various types at different stations by use of tripartite stations and of different types of instruments at a given station.

d. Additional determinations of velocities of microseismic waves and their correlation with velocities of seismic waves found from earthquakes and from blasts in the same region.

e. More data on the relationship between the periods of the cause of the microseisms in specific instances and the periods of the recorded microseismic waves in various parts of their "spectrum."

f. Study of the relationship between microbarometric waves and microseisms (only a few data exist).

g. Measurements of pressure variations at greater depths in the ocean and their correlation with ocean waves at the surface (amplitudes and periods) as well as with microseisms recorded nearby on land.

2. Theoretical

a. Study of the effect of free vibrations of the ground on microseismic waves, especially resonance effects. (This is a minor problem for earthquake waves where there are rarely "continuous" waves within the range of periods prevailing in microseisms.)

b. Quantitative data on effects of maximum or minimum group velocity on amplitudes (Press and Ewing's theory [31]).

c. Extension of the theory of Press and Ewing to other models, especially under assumption of a gradual or sudden increase of velocity with depth in the materials which form the ocean bottom.

d. Study of the transition of elastic waves from one structure to another, for example, gradually or suddenly from a model of the type considered by Press and Ewing to a typical continental structure.

e. Study of the transition of surface waves from one structure to another with different elastic constants down to a depth of (1) a fraction of the wave length, (2) a multiple of the wave length.

REFERENCES

1. Banerji, S. K., "Theory of Microseisms." *Proc. Ind. Acad. Sci.*, (A) 1:727–753 (1935).
2. Båth, M., "An Investigation of the Uppsala Microseisms." *Medd. meteor. Inst. Uppsala*, No. 14, 168 pp. (1949).
3. Bernard, P., "Étude sur l'agitation microséismique et ses variations." *Inst. Physique du Globe de Paris*, 19:1–77 (1941).
4. *Bibliography on Microseisms.* Seismological Lab., Calif. Inst. of Tech., Div. Geol. Sci., Pasadena, Calif., Contrib. No. 523, mimeogr., 63 pp. (321 entries), 1949.
5. Caloi, P., "Due caratteristici tipi di microsismi." *Ann. Geofis.*, 3:303–314 (1950).
6. Darbyshire, J., *The Correlation between Microseisms and Sea-Waves.* Oceanogr. Sec., British Admiralty Res. Lab. Paper presented at Joint Meeting Assn. Seismol., Meteor., and Phys. Oceanogr., Gen. Assembly, Oslo, August, 1948.
7. —— "Identification of Microseismic Activity with Sea Waves." *Proc. roy. Soc.*, (A) 202:439–448 (1950).
8. Deacon, G. E. R., "Relations between Seawaves and Microseisms." *Nature*, 160:419–421 (1947).
9. —— "Recent Studies of Waves and Swell." *Ann. N. Y. Acad. Sci.*, 51:475–482 (1949).
10. Geussenhainer, O., *Ein Beitrag zum Studium der Bodenunruhe mit Perioden von 4 sec bis 10 sec.* Diss., Göttingen, 1921. (Abstract in *Jb. phil. Fak. Univ. Göttingen*, 2:73–80 (1921).)
11. Gherzi, E., "Microseisms Associated with Storms." *Beitr. Geophys.*, 25:145–147 (1930).
12. Gilmore, M. H., "Microseisms and Ocean Storms." *Bull. seism. Soc. Amer.*, 36:89–119 (1946).
13. Gutenberg, B., "Über Fortpflanzung von elastischen Wellen in viskosen Medien." *Phys. Z.*, 30:230–231 (1929).
14. —— "Microseisms in North America." *Bull. seism. Soc. Amer.*, 21:1–24 (1931).
15. —— "Die seismische Bodenunruhe," *Handbuch der Geophysik*, Bd. 4, SS. 264–298. Berlin, Gebr. Borntträger, 1932.
16. —— "Microseisms and Weather Forecasting." *J. Meteor.*, 4:21–28 (1947).
17. —— und Schlechtweg, H., "Viskosität und innere Reibung fester Körper." *Phys. Z.*, 31:745–752 (1930).
18. Kishinouye, F., "The Unusually Large Microseisms of Oct. 21, 1938, at Hongo, Tokyo." *Bull. Earthq. Res. Inst. Tokyo*, 18:401–418 (1940).
19. Krug, H.-D., "Ausbreitung der natürlichen Bodenunruhe (Mikroseismik) nach Aufzeichnungen mit transportablen Horizontalseismographen." *Z. Geophys.*, 13:328–348 (1937).
20. Lee, A. W., "The Effect of Geological Structure upon Microseismic Disturbance." *Mon. Not. R. astr. Soc.*, Geophys. Supp., 3:83–105 (1932).
21. —— "A World-Wide Survey of Microseismic Disturbances." *Geophys. Mem.*, Vol. 7, No. 62 (1934).
22. —— "On the Direction of Approach of Microseismic Waves." *Proc. roy. Soc.*, (A) 149:183–199 (1935).
23. Longuet-Higgins, M. S., "A Theory of the Origin of Microseisms." *Phil. Trans. roy. Soc. London*, (A) 243:1–35 (1950). (See also "Sea Waves and Microseisms." *Nature*, 162:700 (1948).)
24. Lynch, J. J., "Progress of Investigation of Two-Second Microseisms at the Tripartite Station at Fordham University." *Earthq. Notes*, 20:24 (1949).
25. Macelwane, J. B., and others, *Investigation of the Nature and Origin of Microseisms of Frequency Two to Three Cycles per Second.* Paper presented at the Joint Meeting

Assn. Seismol., Meteor., and Phys. Oceanogr., Gen. Assembly, Oslo, August, 1948.

26. Menzel, H., "Zur Theorie der seismischen Bodenunruhe." *Dtsch. hydrogr. Z.*, 2:169–177 (1949).
27. Miche, M., "Mouvements ondulatoires de la mer en profondeur constante ou décroissante." *Ann. Ponts Chauss.*, 114:25–78, 131–164, 270–292, 369–406 (1944).
28. Munk, W. H., "Note on Period Increase of Waves." *Bull. seism. Soc. Amer.*, 39:41–45 (1949).
29. Murphy, L. M., "Winter Microseisms." *Trans. Amer. geophys. Un.*, 27:19–26 (1946).
30. Pekeris, C. L., "Theory of Propagation of Explosive Sound in Shallow Water." *Geol. Soc. Amer. Memoir* 27, Pt. 2 (1948).
31. Press, F., and Ewing, M., "A Theory of Microseisms with Geological Applications." *Trans. Amer. geophys. Un.*, 29:163–174 (1948).
32. Ramirez, J. E., "An Experimental Investigation of the Nature and Origin of Microseisms at St. Louis, Missouri." *Bull. seism. Soc. Amer.*, 30:35–84, 139–178 (1940).
33. Ricker, N., "The Form and Nature of Seismic Waves and the Structure of Seismograms." *Geophysics*, 5:348–366 (1940).
34. —— "Further Developments in the Wavelet Theory of Seismogram Structure." *Bull. seism. Soc. Amer.*, 33:197–228 (1943).
35. Scholte, J. G., "Over het Verband tussen Zeegolven en Microseismen." *Ned. Akad. Wetensch.*, 52:669–683 (1943).
36. Schulze, G.-A., "Natürliche Bodenunruhe," *Naturforschung und Medizin in Deutschland, 1939–1946. (FIAT Rev.)*. Wiesbaden, Dieterich, 1948. (See Vol. 18, pp. 16–22)
37. Sezawa, K., "On the Transmission of Seismic Waves on the Bottom Surface of the Ocean." *Bull. Earthq. Res. Inst. Tokyo*, 9:115–143 (1931).
38. Stoneley, R., "The Effect of the Ocean on Rayleigh Waves." *Mon. Not. R. astr. Soc.*, Geophys. Supp., 1:349–356 (1926).

PRACTICAL APPLICATION OF MICROSEISMS TO FORECASTING

By JAMES B. MACELWANE, S. J.

Saint Louis University

The term *microseisms* applies to all elastic wave systems which are propagated along the surface of the earth and which, on the one hand, are not caused by earthquakes and, on the other hand, are not purely local man-made disturbances due to traffic or industrial activity. We may also exclude from present consideration the more or less irregular local vibrations produced by natural causes, such as varying pressure of the wind on a particular structure, wind friction on a landscape, freezing and thawing of the ground, jerky movements due to cooling of structures or of hills or other topographic features, although these are often referred to in the literature as a class of microseisms. We are thus restricting our attention to elastic surface waves which are propagated in the earth's crust over appreciable distances. These microseisms seem to be of different types and to appear in discontinuous frequency bands rather than in a continuous spectrum.

The type of microseisms to which most study has been devoted since the time of Bertelli is characterized by wave periods which, in the majority of cases, lie somewhere between four seconds and seven seconds, that is, by frequencies between 140 and 250 milliherz (1 herz equals 1 cycle per second). The reason for this emphasis is not far to seek. These frequencies fall within the optimum response range of most of the earthquake seismographs in common use and hence the microseisms of this type appear as a more or less disturbing background on most of the earthquake records.

Some connection between this type of microseisms and meteorological conditions in general had been noted by many workers beginning with Bertelli. These microseisms are regular in wave form and appear in a succession of groups of a few large waves each with short intervals of slight motion between the groups. They do not appear at all times but in discrete sequences which may last for a period of hours or days, building up to a maximum and dying down again. Such a sequence has come to be known as a *microseismic storm* (Fig. 1).

Four principal theories have arisen concerning the nature and origin of this type of microseisms. The first group of theories related them to the meteorological and geological conditions surrounding the recording station. The rise and fall of amplitude was attributed by some to the simultaneous arrival of vibrations of slightly different frequencies so as to produce an effect of beats. The second theory related the microseisms to steep barometric gradients on land or sea. A third theory, championed principally by Wiechert and other scientists of the Göttingen school, and particularly by Gutenberg through a period of three decades, ascribed

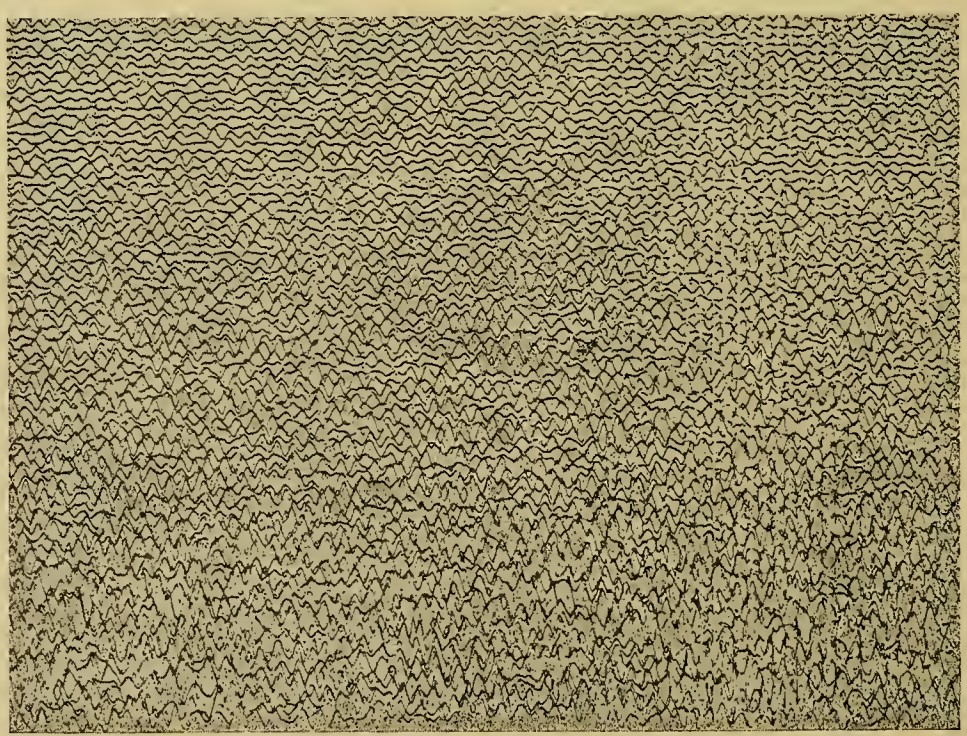


FIG. 1.—Storm microseisms recorded by a Wood-Anderson seismograph. Interval between time marks, 1 min.

this type of microseisms to vibrations caused by the beating of surf on a rocky coast. The fourth theory held that group microseisms are produced in some manner by storm centers at sea and are propagated through the earth's crust to the continents. This theory was espoused by Klotz, Shaw, Lee, Banerji, and especially by Gherzi who suggested as the mechanism a "pumping" or vertical oscillation of the air vortex in a typhoon, hurricane, or other closed circulation over deep water, thus explaining why microseisms cease to be produced when the storm center reaches land.

Many investigators have attempted to decide between these various theories by correlating large amplitudes chronologically with other phenomena, but without any decisive results. Several, for example, Linke, Lee, Archer, Neumann, and Leet, have attempted to determine the direction of arrival of the waves, on the assumption that they are true Rayleigh waves, by determining the relative amplitudes of the three components on the records of a single station. This method also proved inadequate.

A direct attack on the problem of direction of the propagation of the microseisms without any assumption was made independently by Ramirez at Saint Louis and by Trommsdorf at Göttingen in 1938 by means of the observations of the time intervals between the successive arrivals of the same wave at the three corners of a tripartite station, using these time intervals between successive arrivals to calculate both the direction of propagation and the speed of travel. Suitable equations for this purpose were developed by Krug, Gilmore, Macelwane, and Schuyler who also devised a nomogram.

The following equations are valid for any tripartite station. Let A, B, and C (Fig. 2) be the known angles of the triangle named counterclockwise; and let a, b, and c represent the respective opposite sides of known length. Assume that a given microseismic wave front has passed over vertices B and C and is now at A. Drop perpendicular BP from vertex B to the wave front and another perpendicular CQ from vertex C to the wave front. Let X be the angle QAC between the wave front and the side b, and let Y be the angle PAB between the wave front and side c. Let ctn $A = k$ and let the constants (b/c) sec $A = m$ and (c/b) csc $A = n$.

Now let the interval of time (Fig. 3) between the instants at which the chosen wave crest passed over B and A be t_{BA}, and that between the times of its passage over C and A be t_{CA}. Then,

$$\text{ctn } Y = n\frac{t_{CA}}{t_{BA}} - k,$$

and

$$\text{ctn } X = m\frac{t_{BA}}{t_{CA}} - k.$$

Ramirez found that the direction from which the group microseisms arrived at Saint Louis always pointed toward a storm at sea. The U. S. Navy adopted the Ramirez method in 1943. While the method as applied by Ramirez at a single tripartite station gave only the direction, it showed as a matter of fact, in the case of a hurricane off the Atlantic Coast, that this direction continuously pointed toward the center of the storm as it moved up the coast and did not point

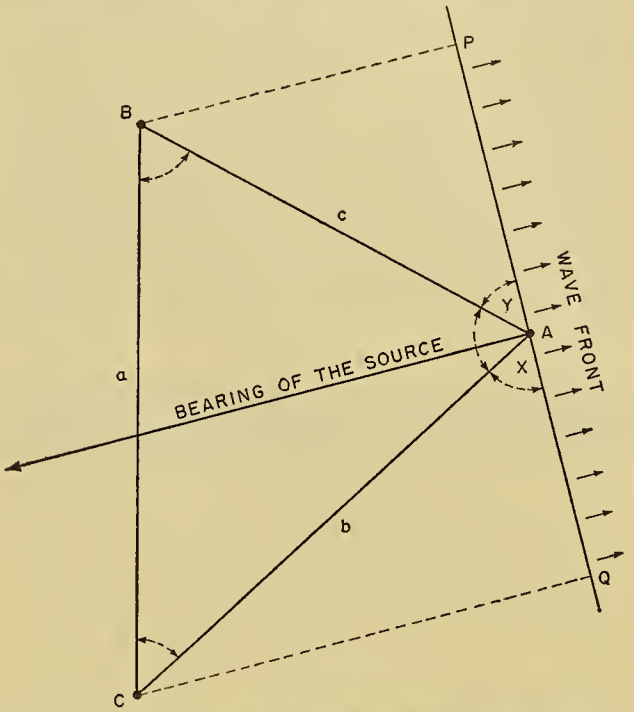


FIG. 2.—A wave front of microseisms passing any tripartite station with a seismograph at A, B, and C, respectively.

toward the area where heavy surf existed. The location of the storm center in latitude and longitude could only be found by cross bearings from more than one tripartite station. By 1945 the U. S. Navy installations in the Caribbean area under Orville and Gilmore had

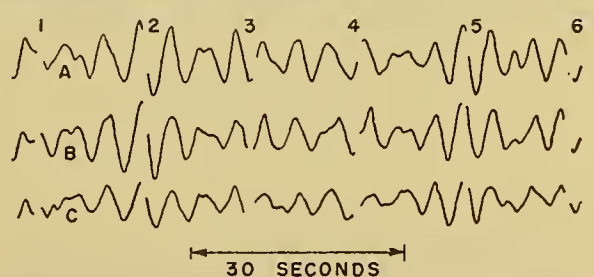


FIG. 3.—Simultaneous time marks showing succession of arrivals at stations A, B, and C. (*Courtesy M. H. Gilmore.*)

obtained satisfactory evidence that these microseisms actually do originate in the storm area at sea. Accordingly the method has been carried to the Pacific Ocean with equal or even greater success. There is no longer any doubt of the origin of this type of microseisms and of its direct applicability to the tracking of hurricanes, typhoons, and other similar storms from the time of their formation until they reach the land.

The seismographs in use by the U. S. Navy are of the horizontal component, electromagnetic, Sprengnether type. Their periods and those of the Leeds and Northrup galvanometers are set at about six seconds

so that their optimum response range will be in resonance with the microseisms. Horizontal component instruments were chosen in preference to vertical component seismographs because they are simpler to operate and are much less sensitive to fluctuations in temperature and hence require much less elaborate vault insulation.

The seismographs at the outlying corners are connected by shielded cable to the respective galvanometers in the main vault of the tripartite station where they record on a triple drum whose speeds of rotation and endwise translation are variable. Moderate speeds are used as long as there is no evidence of a storm. When the amplitudes of the microseisms are seen to increase, the drum rates are greatly accelerated (Fig. 4).

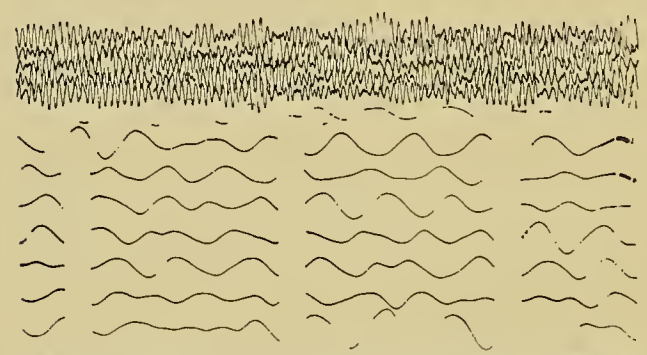


FIG. 4.—Part of a record made at the Miami, Florida, tripartite station of the U. S. Navy, August 23, 1949, showing large microseisms recorded at the stand-by rate (top) and at the hurricane rate (bottom). Time interval between interruptions in both cases, 15 sec.

However, the problem still remains as to the mechanism by which a storm center produces the elastic waves on the ocean bottom and why they are so conspicuously characterized by groups resembling interference patterns (Fig. 5). An argument in favor of swell and decay of activity at the source rather than interference from independent sources within the storm area is the similar appearance of the records in many different directions. A thorough study of the "groups" from this point of view remains to be made.

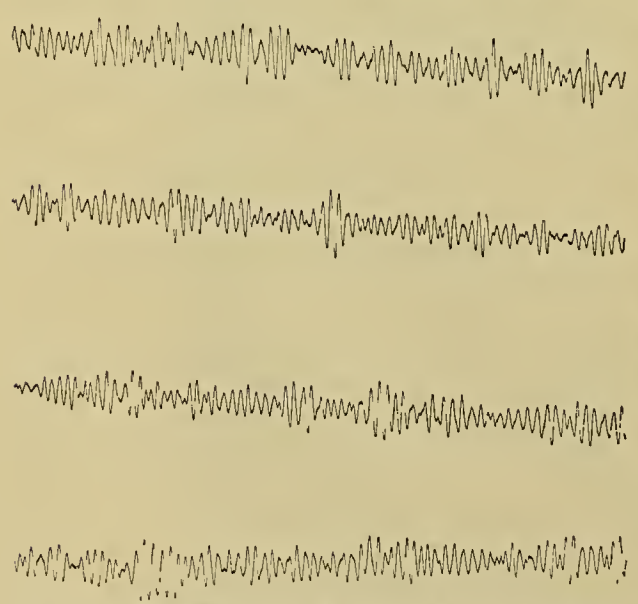


FIG. 5.—Samples from a record made at the Bermuda tripartite station of the U. S. Navy, September 6, 1949, showing the swell and decay "group" effect. Time intervals, 15 sec.

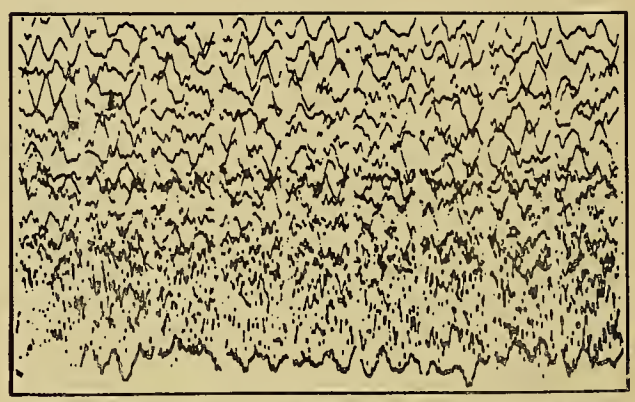


FIG. 6.—Record at Corpus Christi for July 9, 1947, showing increase in amplitude of 1.3-sec microseisms on lower portion of record at the time of a local thunderstorm. Time breaks are at 15-sec intervals.

Suggestions as to the mechanism by which these microseisms are produced have been offered by Banerji, Gherzi, Macelwane, Longuet-Higgins, and others. But the observations necessary to test the validity of any one of them are lacking. In most cases the accumulation of the necessary critical data would be both difficult and costly. For example, observations would be required of the existence or nonexistence of standing ocean waves in a disturbed area of a hurricane or typhoon. The observation of regular "pumping" in the air vortex or of vertical oscillation in a water vortex

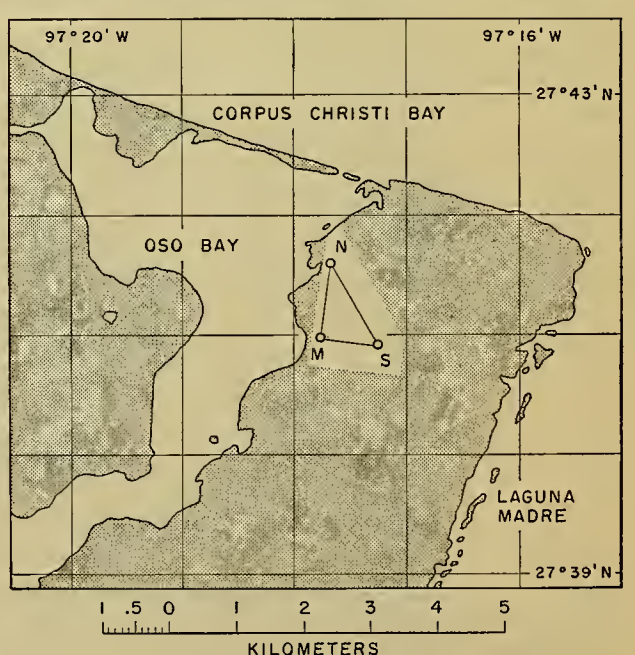


FIG. 7.—Position of the U. S. Navy tripartite station MNS at Corpus Christi, Texas.

underneath the wind system would challenge the resources of any but a governmental organization.

Another type of microseisms is found in a band whose dominant period lies in the neighborhood of two seconds (Fig. 6). Much less is known about these microseisms than about those of longer period. For some time the U. S. Navy maintained a tripartite station near Corpus Christi, Texas (Fig. 7). A preliminary study of the records was made by Jennemann. There were times when this type of microseisms dominated the records sufficiently to permit a determination of the direction of arrival at the station. They appeared to come from west of north. This means that these microseisms were coming not from the Gulf of Mexico

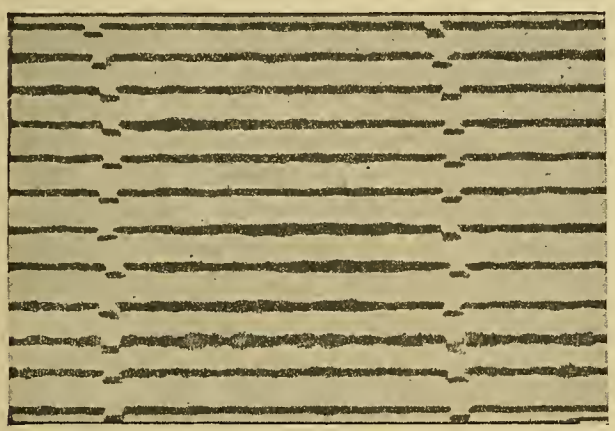


FIG. 8.—Short-period microseisms. Wood-Anderson record, Florissaut station. Time marks every half-minute.

but from the continent. It was possible to correlate these microseisms tentatively with disturbed weather conditions existing to the northward of the station. At Fordham University a tripartite station has been set

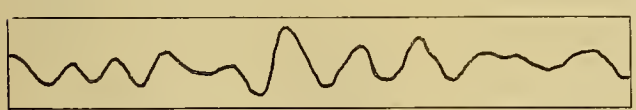


FIG. 9.—Three-tenth-second microseisms as recorded by a Sprengnether-Volk seismograph.

up for the express purpose of studying these two-second microseisms on contract with the Office of Naval Research. Other preliminary studies have been made by Leet and Gutenberg. While it is too early to form a positive judgment, the results so far seem to hold promise for forecasting of conditions related to frontal disturbances.

Preliminary studies are in progress on quite another type of microseisms of frequency of two to three cycles per second on a contract between Saint Louis University and the Office of Naval Research (Fig. 8). They are being studied with a tripartite station equipped with three components at each corner of the triangle and separated by distances considerably less than a wave length with a view to determining how, if at all, they are related to weather conditions and whether they have forecasting value (Fig. 9).

There are other bands of frequencies in the spectrum of microseisms but their possibilities for weather forecasting have not been tested.

Conclusions. There is a limit to what can be done with statistical studies of microseisms. The record of more than three quarters of a century of such studies shows what great care must be taken in their planning and in the interpretation of results. Quite contradictory conclusions have been drawn from parallel correlations. Much the same may be said of vectorial analysis of microseismic records. A very fruitful line of research has been opened up by the use of tripartite stations whose dimensions are less than one wave length. This method is independent of all assumptions except the successive arrival of an identical wave front at the three corners of the tripartite station. It gives a completely independent determination of the direction and speed of wave travel. Cross bearings between two or more tripartite stations locate the source of the microseisms and thus furnish a clue for the investigation of the point of origin within the storm provided there are a sufficient number of tripartite stations available for the observation of one storm. Research on the origin of the microseisms or the mechanism by which they are produced is difficult but very necessary. The possibility of air oscillations acting as a hammer may be investigated by means of microbarographs disposed on small islands in the path of typhoons and hurricanes. A similar system for the investigation of standing waves, or of the oscillations of a water vortex, would be much more difficult and costly. One might think of pressure gauges on a submarine cable laid in a path frequently followed by tropical storms and terminating at a recording station somewhere on land.

REFERENCES

I. The following papers contain extensive surveys of the literature on microseisms and their relation to the weather:

1. GUTENBERG, B., "Untersuchungen über die Bodenunruhe mit Perioden von 4 bis 10 Secunden in Europa." *Veröff. ZentBur. int. seism. Ass.*, 106 SS., Strasbourg (1921).
2. MACELWANE, J. B., "Storms and the Origin of Microseisms." *Ann. Géophys.*, 2:281–289 (1946).
3. RAMIREZ, J. E., "An Experimental Investigation of the Nature and Origin of Microseisms at St. Louis, Missouri." *Bull. seism. Soc. Amer.*, 30:35–84, 139–178 (1940).

II. Important papers on hurricane and typhoon forecasting by means of microseisms are:

4. GILMORE, M. H., "Microseisms and Ocean Storms." *Bull. seism. Soc. Amer.*, 36:89–119 (1946).
5. —— and HUBERT, W. E., "Microseisms and Pacific Typhoons." *Bull. seism. Soc. Amer.*, 38:195–228 (1948).

INDEX

(References to titles of articles appear in bold face.)

A

Absorbers, effect on effective terrestrial radiation, 39
Absorbing layers, upper atmosphere, heights of, 271
Absorption coefficients, 265
  as a function of pressure, 40
  oxygen, 275
  ozone, 275
Absorption cross section, 265
Absorption functions, 36
Absorption laws, 34
  for diffuse radiation, 35
  for parallel radiation, 34
Absorption of solar radiation, 19
  measurement, 50
  spectral evidence, upper atmosphere composition, 270
Absorption spectrum, 264
Actinograph, Robitzsch, 54
Actinometer, Michelson, 53
Actinometric equipment at meteorological stations, 53
  instruments, 52
Actinometric measurements, 50
  accuracy, 50
  atmosphere analysis, 50
  critique of routine, 55
  effect of absorption on solar radiation, 50
  effect of molecular scattering on solar radiation, 50
  effect of varying distance between sun and earth on solar radiation, 50
  heat exchange at earth's surface, 50
  solar constant, 50
  special problems, 55
Actinometric observatory, central, equipment, 54
Actinometric stations, second order, equipment, 54
  third order, equipment, 55
Actinometry, 50
  biological problems and, 52
Actinon, source, 155
Activation energy, 270
Advective-adiabatic extrapolation, prognosis of relative hypsography by, 782
Advective models, barotropic models, 473
  dynamic forecasting and, 471
  Kibel's method, 473
Aerobiology, 1103
  atmosphere as environment for organisms, 1105
  biological factors, 1103
  counting, methods and problems, 1103
  exchange of organisms between earth and atmosphere, 1106
  identifying organisms, methods and problems, 1104
  locating sources, methods and problems, 1104
  meteorological factors, 1106
  origin of molds, 1109
  sampling, methods and problems, 1103
  thermal convection, 1110
  transport of organisms, 1108
Aerology of extratropical disturbances, 599
Aerology of tropical storms, 902
Agriculture, and applied climatology, 987
  and microclimatology, 998
Air, albedo of pure dry, 25
  atmospheric, composition of, 3
Air—*continued*
  earth electric current, 113
  effect on scintillation, 68
  electrical conductivity, 102
  ionization of, 102
  total oxidation, 1130
  tropical, properties, 881
Air current, modifications, forecasting, 779
Air density distribution, 360
Air density, from meteor deceleration, 360
  seasonal variations, 360
Air masses, Antarctic, 923
  Artic, 948
  in equatorial meteorology, 881
  modification, 973
    forecasting, 779
  use for description of climate, 968
Air motion, effect on ozone, 283
Air pressure, effect on downcoming radiation, 40
Air transport, by slope winds, 666
Aircraft, electrification of, flying through precipitation, 133
  meteorological instruments, 1223
Aircraft icing, analysis of formation, 1190
  at low temperatures, 194
  availability of liquid water, 1190
  cloud-drop diameter, measurement, 1198
  cloud-drop size, 1201
  concentration of liquid water in clouds, 1199
  conditions for formation, 1192
  effect on airplane components, 1193
  forecasting, 793
  forecasting icing conditions, 1201
  icing conditions, physical characteristics, 1197
  intensity, 1198
  liquid-water concentration, measurement, 1198
  meteorological aspects, 1197
  meteorological conditions, research needed, 1202
  operational aspects, 1190
  physical aspects, 1190
  physical characteristics of icing conditions, 1197
  research needed, 1195
  special cases, 1193
  water impingement, 1191
Airglow, 269
Airline operations, and instability lines, 647
Airplane, thunderstorms and, 690
Aitken dust counter, 165
Aitken-nuclei, 183; *see also* Condensation nuclei
  destruction processes, 188
  effect of pressure changes on, 186
  form, 184
  formation in smoke, 186
  growth, 184
  horizontal distribution, 184
  mechanical sources, 187
  optical effects, 188
  physical state, 184
  size, relation to electrostatic charge, 187
  vertical distribution, 184
Albedo, 25
  of cloudless atmosphere, 25
  of dust, 26
  of earth's surface, 27
  of planet Earth, 27
Albedo—*continued*
  of pure dry air, 25
  of the Earth, 31
  of various cloud types, 26
  of various surfaces, 27
  of water vapor, 26
  total atmospheric, 26
Alpine tree line, 953
Alpine tundra, 953
Alps, air circulation during daytime, 663
Altimeter, 1226
Altitude(s), effect on electrical conductivity of air, 104
  extreme, pressures at, 308
    temperatures at, 308
Ammonia content of atmosphere, 7
Ammonium ions, effect on condensation nuclei formation, 186
Amorphous frost, 207
Anafront, 771
Angels, 1277
Ångström compensation pyrheliometer, 52
Ångström pyrgeometer, 55
Antarctic, air exchange, 936
  air masses, 923
  anticyclone, 918
  atmosphere, exchange, 936
    oxygen content, 933
  atmospheric circulation, 917
  blizzards, 919
  circulation, 925
    prospects for study, 937
  cyclones, 922
  exploration, 917
  frontal zones, 922
  geography, 917
    glaciation, 934
  ice, recession, 934
    thickness, 933
  interior, summer winds, 921
    weather conditions, 921
  precipitation, 933
  pressure waves, 926
    characteristics, 926
    movement, 926
    theories, 927
  seasons, 920
  snow surface, altitude, 934
  soundings, 930
  storms, 924
  stratospheric warming, 931
  surges, 926
  temperatures, 929
    march of, 932
    summertime, 931
    wintertime, 930
  upper winds, 920
  water temperature, 932
  weather observation stations, 708, 918, 937
  winds, 919
Anticorona, 71
Anticyclogenesis, 770
Anticyclones, 621
  and forecasting, 626
  and summer showers, 626
  Autarctic, 918
  "back-door" cold fronts and, 627
  definition, 621
  degeneration, 778
  high level, formation, 612
  generation, 778
  Indian summer, 626
  motion, 625
  origin, 621

Anticyclones—*continued*
polar, 621
forecasting motion of, 625
transformation to dynamic, 625
role in general circulation, 621
temperature structure, 621
thermal structure, 621, 627
transformation, 625
warm, 622
advective theory, 622
dynamic theory, 622
radiative theory, 622
water-vapor structure, 316
westerlies and, 623
Anticyclonic shear, upper atmosphere, 585
upper limit, 604
Anticyclosis, 780
Antitrades, air motion, 862
Anti-twilight arch, 73
Arago point, 79
distances, effect of ground reflection on, 82
position, 82
Arctic, air, seasonal variations, 947
winter, 948
atmospheric circulation studies, 942
Baffin Bay, 960
central basin, circulation pattern, 943
circulation patterns, 944
concepts, 943
typical, 949
variations, 949
climate and tree growth, 954
climatic control, 953
climatological problems, 952
climatology, ecological, 953
general literature, 952
in Greenland, 961
Eurasian, air, seasonal variations, 948
expedition studies, 944
extent of, 956
literature, climatology, Greenland region, 952
in North America, 952
Northern Russia, 952
over Arctic Ocean, 952
meteorological problems, 950
meteorology, 942
700-mb contour patterns, 946
North America, air, seasonal variations, 947
North Water, 960
observational data, 944, 950
permafrost distribution, 958
reconnaissance flights, 944
sea ice, scope for research, 960
sea-ice distribution, 958
sea-level pressure, 945
sea-level pressure charts, 944
soundings, 930, 947
tree growth, 954
tree line, 955
vegetation, zonal divisions, 955
weather observation stations, 708, 944
Argon, content of, in atmosphere, 5
Aspiration condenser, 147
Aspirator, Gerdien, 147
Weger, 147
Astronomical refraction, 64
Atmosphere, albedo of cloudless, 25
ammonia content, 7
analysis of, 50
argon content of, 5
as environment for organisms, 1105
as hydrodynamic fluid, laboratory investigations, 1235
carbon dioxide content, 3, 4
carbon monoxide content, 7
cloud-free, refraction phenomena in, 64
composition of, 1
variations in, 3
compressible, wave motions in, 414
Atmosphere—*continued*
conductivity, measurement of, 146
constituents, diffusion coefficients, 321
molecular properties, 321
of variable concentrations, 6
contrast, reduction by, 92
cooling, 41
by water-vapor radiation, 43
density, from meteor deceleration, 360
dispersion of light rays by, 67
downcoming radiation from, 38
electric field, in fair weather, 107
measurement, 148
electrical resistance of vertical column, 106
energy transfer within, modes of, 544
exchange of heat with sea surface, 1057
free, electrical measurements of, 150
free oscillations in, 306
future developments, 9
gases of constant percentage, 5
heat gain through convection, 1065
helium content of, 5
hydrogen content, 6
ion concentration, measurement, 146
ion mobility, measurement, 146
ionizers, 146
ions, 120
isotopic composition of gases, 8
Jupiter, 394
kinetic energy balance, 544
krypton content, 5
light behavior in, 91
major planets, circulation, 395
Mars, 392
methane content, 5
middle, origin of high temperature in, 253
mold distribution, 1109
neon content, 5
Neptune, 394
nitrogen content, 5
nitrogen dioxide content, 7
nitrous oxide content, 5
of other planets, 391
ozone content, 6, 275
radio waves in, refraction of, 1290
radioactive substances in, 155
radioactivity, 155
measurement, 155
by emanometry, 155
by induction method, 157
problems, 159
radon, decrease with height, 158
radon balance, 158
radon content, 158
variations, 157
removal of organisms from, 1107
Saturn, 394
scattering of light in, 76
solar radiant energy modification by, 13
sound propagation in, 366
sound wave energy in, 368
stability of permanent, horizontal, isobaric motion, 446
subjective phenomena, 61
sulphur dioxide content, 7
suspensions, 146
temperature change, by water-vapor radiation, 44
upper *see* Upper atmosphere
Uranus, 394
variation of oxygen content, 3, 933
Venus, 391
water vapor density in, 8
wheat stem rust transport in, 1108
xenon content, 5
Atmospheric air, composition of, 3
chemistry *see* Chemistry, atmospheric
circulation *see also* Circulation
general, studies of, 551
density distribution, 360
density, seasonal variations, 360
Atmospheric air—*continued*
disturbances, classifying, 457
large-scale, stability properties of, 454
electricity *see* Electricity, atmospheric
flow, circular vortex, 455
hydrodynamic equations, 454
gases, escape, 246
isotopic composition, 8
motions, experimental analogies, 1235
oscillations *see* Tides, atmospheric
ozone, theory of, 281
perturbation theory, basic assumptions, 402
perturbations, factors involved in, 445
hydrodynamic instability and theory of, 444
pollution *see* Pollution, atmospheric
stability, effect on atmospheric pollution, 1146
testing, 694
tides *see* Tides, atmospheric
turbulence, heat flux due to, 536
Atomic collisions, types, 267
number, 263
particles, 263
combination, 269
dissociation of, 268
ionization of, 268
spectra, 265
two- and three-body collision, 269
processes, reversibility, 267
transitions, spectra associated with, 264
Attachment energy, 263
Audibility zones, abnormal, sound waves through, 371
Aureole, 71
Aurorae, 345
and magnetic storms, 347, 352
appearance, 347
Birkeland's model experiments, 351
charged corpuscle movement and, 352
corpuscular theory of, 351
E-layer of ionosphere during, 353
F_2-layer of ionosphere during, 353
forms, 347
flaming, 347
with ray-structure, 347
without ray-structure, 347
in upper atmosphere, 255
ionosphere and, 352
magnetic disturbances and, 258
position in space, 349
direction of arcs, 349
height statistics, 349
method of height determination, 349
position of radiation point, 349
sunlit aurora, 350
spectrum, 350
Störmer's calculations and, 352
use in studying upper atmosphere, 251
VHF-band reflections from, 354
Aviation medicine, 1121
Avogadro's number, 263

B

Babinet point, 79
position, 82
Bacteria, marine, exchange between earth and atmosphere, 1106
soil, exchange between earth and atmosphere, 1106
Ball lightning, 141
Ballo-electricity, 1134
Balloons, 1214
altitude performance, 1214
ascensional rates, 1215
recording, use in determining ozone distribution, 278
studies, in upper atmosphere, 303
on temperature, 303
temperature effects, 1214
tracking direction finder, 1208

Band spectra, 265
Baroclinic disturbances, 457, 463
  fluids, hydrodynamic equations, 403
  instability, in fluid motion, 468
  waves studies, 459
    by Charney, 459
    by Eady, 459
    by Fjørtoft, 459
Barometric oscillations, use in studying upper atmosphere, 251
Barometric variations, lunar daily, 515
  solar daily, 521
Barometry, tropical cyclones, 888
Barotropic disturbances, 460, 463
Barotropic fluids, hydrodynamic equations, 403
Barotropic models, advective models, 473
  linearized, 474
  nonlinear, 478
Barotropic waves, stability of, 462
Bead lightning, 141
Bellani pyrgeometer, 55
Bénard cells, 701, 1255
  Sutton's theory, 1261
Bifilar electrometer, 144
  Wulf's, 145
Biological problems, actinometry and, 52
Birkeland, model experiments of, 351
Bishop's ring, 71
Bioclimatology, human, aviation medicine, 1121
  breathing, 1116
  clothing, 1115
  cooling, 1115
  effect of aperiodic weather processes, 1120
  effects of extreme cold, 1122
  effects of extreme heat, 1122
  effect of periodic weather processes, 1120
  erythema solare, 1118
  general effects of climate, 1117
  glare, 1119
  heat balance, 1112
  heat loss by convection, 1112
  heat production, 1112
  heat transfer by evaporation, 1113
  heat transfer by radiation, 1113
  meteorotropism, 1120
  ozone and, 287
  physical aspects, 1112
  radiation effects, 1118
  research needed, 1122
  space medicine, 1121
  sultriness and heat, 1114
  sunburn, 1118
  total heat balance, 1114
  vitamin D, 1119
Blocking waves, in planetary jet stream, 432
Böenregen, electrical characteristics, 128
Bolides, 356
Bora, 669
  anticyclonic, 670
  cyclonic, 670
Boreal forest, 953, 957
  southern limit, 958
Boundary layer, in air flow, 492
Bowen ratio, 1058, 1073
Boys cameras, use in determining mechanism of lightning discharge, 142
Branching, 136
Breezes, land and sea, as function of geographical location, 658
  description, 655
  desirable studies, 670
  effect of season on, 658
  effect of time of day on, 658
  explanation, 656
  intensity, 658
  phase shift, 661
  pressure differences during, 657
  pressure distribution during, 657
Breezes, land and sea—*continued*
  range, 658
  single circulation cell, 660
  temperature differences during, 657
  theory of, 659
  tropical, 861
  vertical extent, 658
sea, cold front-like, 659
  Conrad's minor, 658
  direction, 655
  effect of cloudiness on, 658
  of the first kind, 658
  of the second kind, 659
Brewster point, 79
Bright band, and precipitation growth, 1286
  on radarscope photographs, 1273
Bright segment, 73
Brocken-specter, 71
Brückner cycle, 383, 817

C

Cage method, 149
Capillary electrometers, in measurement of atmospheric electricity, 144
Carbon dioxide, atmospheric content, 3
  effect on radiation diagrams, 37
  effect on radiation in stratosphere, 46
  infrared spectra, 296
  variation in atmospheric content over the sea, 4
  variations, effect on climate, 1015
Carbon, isotopic variations in atmosphere, 9
Carbon monoxide content of atmosphere, 7
Cathode-ray oscillograph, in measurement of lightning characteristics, 142
Ceilometers, 1218
  pulsed light, 1218
  radar, 1218
Cellular convection; *see* Convection, cellular
Centrifugal instability, in fluid motion, 467
Characteristics, definition of, 421
  determination of, 421
  envelope of, 423
  method of, 421
    mathematical foundations, 421
    numerical integration of, 422
    quasi-linear differential equations, 421
Charge determinations, lightning discharges, 140
Charney, baroclinic wave studies, 459
Charts, moisture, use in meteorological analysis, 725
  use in meteorological analysis, 721
Chemistry, atmospheric, 262
    absorption-tube methods for detecting gaseous materials, 1126
    absorption-tube method results, 1128
    condensation method, for determining traces which form condensation nuclei, 1127
    iodine, 1128
    methods, 1126
    ozone, of air strata near ground, 1129
    problems of, 1126
    results from condensation method, 1131
    total oxidation of air, 1130
  gas, general principles, 263
  troposphere and, 262
Chinook *see* Foehn
Chlorine, atmospheric, determining, 1126, 1131
  ions, effect on condensation nuclei formation, 186
Christian era, climatic changes during, 1008
Circuit breakers, reclosing, use in lightning protection, 142
Circular vortex, characteristics, 432
Circulation, Antarctic, 917, 925
  cellular character at sea level, 555
  comparison between Northern and Southern Hemispheres, 555
  general, air drift in middle latitudes, 542
    atmospheric, forms of, 822
    considerations in long-range forecasting, 837
    density and meridional motion, 546
    developing cyclone, kinetic energy for, 545
    during strong polar vortex, 547
    effect of mountain ranges on, 547
    energy principles applied to, 568
    energy, total, global balance, 572
    flow types, and sunspot numbers, 840
    horizontal, 543
    horizontal kinetic energy, transport, 545
    index cycle, 560
    internal dynamic consistency, 542
    irregular variations, 559
    kinetic energy, global balance, 568
    kinetic energy balance, 544
    kinetic energy balance fluctuations, 546
    laboratory investigations, 1240
    lag correlations, 565
    low-index conditions, 547
    mean charts, reasons for use of, 552
    meridional circulation strips, 824
    meridional indices, 564
    meridional transfer of internal heat energy, 546
    models, 542
    momentum and energy balance, 548
    nonseasonal variations, 559
    normal state, 551
    northward transport, 543
    patterns, 551
    physical basis, 541
    physical climatology, 556
    qualitative synoptic studies, 559
    quantitative empirical studies, 562
    solenoidal index, 564
    sources of data, 551
    statistical studies, 564
    summer arctic anticyclone of stratosphere, 824
    trough motion, 563
    unordered heat exchange, 824
    wave motion, 563
    wave speed and time, 562
    westward progression, 561
    winter anticyclone in troposphere, 824
    zonal wind speed and wave lengths, 562
  geostrophic zonal wind speed, 555
  global, 541
  in cell in cumulus stage, 697
  indices, 555
  jet stream *see* Jet stream
  local, 653
  maximum index, 555
  mean, relationship of precipitation to, 810
  meridional and zonal, change between, 827
  meridional indices, 555
  normal, of lower stratosphere, 553
    of troposphere, 553
  patterns, Arctic, 944
    concepts, 943
    general, observational studies, 551
    index cycle, 560
    observational accuracy, 564
    qualitative synoptic studies, 559
    quantitative empirical studies, 562
    translation into weather anomalies, 810

Circulation—*continued*
prognosis, physical methods in, 806
relationship between weather and, 809
studies, Arctic, 942
tropical, empirical laws governing, 861
undulations in circumpolar vortex, 555
upper-level, effect on weather, 723
zonal, at sea level, 555
variation in speed, 560
zonal indices, seasonal behavior, 555
zonal wind-speed profile, 555
Circumpolar vortex, jet stream of, 559
undulations in, 555
Cirrus, ice crystals in, 222
ice nuclei in, 196
City planning, and microclimatology, 1001
Clicks *see* Sferics
Climate, air-mass description, 968
difficulties of, 970
as a dynamical problem, 972
as a physical problem, 972
basic requirements for discussion, 974
during geological time, 1004
during ice ages, 1004
early ice ages, 1006
Eocene, 1004
geographic control, 1012
Jurassic age, 1004
mountain building and, 1013
pluvial periods, 1006
synthesis of weather, 967
world, expression of, 967
Climatic change, continental drift, 1012
correlation with glacier changes, 1020
dependent on external influences, theories, 1010
due to terrestrial causes, theories, 1012
during Christian era, 1008
effect of changes in earth's orbit, 1010
effect of solar radiation on, 1010
geological aspects, 1004
historical aspects, 1004
Palaeozoic ice age, 1014
postglacial, 1007
research needed, 1017
role of carbon dioxide variations, 1015
role of ocean currents in, 1014
role of polar ice caps, 1014
solar-topographic theory, 1016
topographic theory of, 1016
tree-ring indices, 1028
volcanic dust and, 1015
Climatic data, analysis of requirements in various activities, 976
Climatic Optimum, 382, 1007, 1021
Climatological analysis, for operational planning, 980
Climatological material, sources of, 977
Climatological problems, Arctic, 952
Climatology, agricultural, 980
applied, 976
and aerial mapping, 986
and agricultural land usage, 987
and aviation, 988
and clothing design, 983
and deterioration of materials, 985
and flood water run-off, 987
and forest fires, 988
and housing design, 982
and marketing of "weather goods," 986
and objective forecasting, 986
and solar radiation, 984
and wind for power production, 984
classes of problems, 979
data for, 977
definition, 976
frost hazard, 984
historical note, 976
list of topics, 990
techniques, 981
weather insurance, 981
Climatology—*continued*
aviation, demands of, 968
Greenland, 961
human comfort and health, 980, 1117
human pathology and, 981
use of synoptic maps in, 970
Clothing design, applied climatology and, 983
Cloud, and weather types, 1175; *see also* Clouds
boundaries, upper, distribution with altitude, 44
centers of electrification, effect on aircraft, 134
classification, need for uniform practice, 1165
according to form, 1165
convection, 694
classes of, 1261
cross sections, 1174
data, synoptic, aid in forecasting, 1173
droplet, accretion, 205
evaporation of, 204
growth by coalescence, 176
growth formula, 169
growth of, 169
ice phase, 173
liquid-water content, 171
measurement, 171
size, 1201
and liquid water, 171
distribution, 170
by cloud types, 173
in sea fog, 173
distribution curves, 172
in free atmosphere, 172
measurement, 171
edges, evaporation from, 200
heat conduction from, 200
elements, condensation on, 204
effect of thermal processes on, 204
electrical characteristics, 130
heat exchange between air and, 199
melting processes, 205
sublimation on, 204
–environment mixing, 695
formation, experimental, 1255
applications to atmosphere, 1260
comparison of phenomena in air and carbon dioxide, 1260
Lord Rayleigh's theories, 1256
motion in unstable layers of air, 1256, 1258
nature of simple convection cell, 1255
research needed, 1261
Sutton's theory of Bénard cells, 1261
forms, classification, 1161
genetical classification, 1162
international classifications, 1161
physical classifications, 1163
radiational influences, 203
indications of new developments, 1175
of passing low, 1175
mass, effect of elevation on, 698
effect of humidity on, 698
effect of temperature on, 698
modification, artificial, relation to production of precipitation, 235
by water, 230
methods, 229
with dry ice, 229
with silver iodide, 230
motions, as wind indicators, 1172
observation, 1167
Physics Project, 236
experiments on stratus clouds, 236
experiments with cumulus clouds, 239
properties, mixing on, 696
radiation, 42
seeding *see* Seeding, cloud
size, aircraft measurement, 1228
Cloud—*continued*
structure, relation to thunderstorm electricity, 687
synoptics, historical summary, 1174
-to-cloud lightning discharge, 139
transformations, interpretation of observed trends in, 1168
types, drop-size variations in various, 173
Cloudiness, effect on ultraviolet radiation, 1119
Clouds, 163, 1159
albedo, 26
altocumulus, and thunderstorm movement, 1173
altocumulus and rain, 1169
altocumulus castellatus, and rain and thunderstorms, 1169
altocumulus lenticularis, and wet weather, 1170
cirrocumulus, weather conditions with, 1170
classification of forms, assessement of adequacies, 1164
assessment of inadequacies, 1164
cirrus, movements, and temperature changes, 1172
motion, use in forecasting cyclone movement, 1171
speed, and storm speed, 1172
velocities, and weather changes, 1172
convective, dynamics of, 201
slice method of analyzing, 201
development of ice phase in, 194
direction of motion, use in forecasting, 1170
double cirrus, weather associated with, 1169
drop-size distribution, 170
effect on atmospheric radiation, 42
effect on shape of sky, 61
entrainment of air, 694
formed outside space they occupy, 1162
high-level, ice phase in, 196
ice *see* Ice clouds
in forecasting, research needed, 1176
in short-term forecasting, 1167
interior, heat conduction in, 200
iridescent, 71
liquid-water concentration, 1200
liquid-water content, effect of elevation, 698
mammatus, weather associated with, 1169
medium-level, ice phase in, 196
microorganisms in, 1109
mixed, physics of, 192
mixing processes and, 200
observed, application of diurnal cycle to, 1167
of advection, 1163
of convection, 1162
of turbulence, 1163
orographic, 1163
physical classification, 1163
physics of, 165
point of convergence, use in forecasting, 1170
proper to space they occupy, 1162
radiation processes and, 43, 200
reflection of solar radiation by, 26
solar radiation absorption by, 21, 31
supercooled, elimination by cloud seeding, 232
factors controlling life cycle of, 227
temperature lapse rate, effect of mixing, 696
thermodynamic analysis, 697
thermodynamics of, 199
tropical cyclones, 890
undulated sheet, weather associated with, 1169
use in forecasting, 1167
wind-shift line marked by, 1173

Coagulation, effect in nuclei destruction, 189
Cold fronts, "back-door," anticyclones and, 627
laboratory investigations, 1238
Collector, electrical field conditions produced by, 149
flame, 148
point, 148
radioactive, 148
water-dropper, 148
measurement of discharge of electrostatic induction by, 148
Collision, de-excitation by, 267
excitation by, 267
-interval, 268
Compatibility, equations of, 422
Compression waves, 425
Condensation, atmospheric, nuclei of, 182
see also Condensation nuclei
below 0C, 174
initial phase, 169
Condensation mechanism, relation of ice clouds to, 197
Condensation nuclei *see also* Aitken-nuclei
Condensation nuclei, 165
and gas reactions, 186
atmospheric, *p*H, determination, 1131
combination of small ions with, 123
condensation effect, 182
destruction processes, 188
determining, 1127
discovery of, 165
distribution of, 168
effect of pressure changes on, 186
electrostatic charges on, 187
formation in smoke, 186
growth, 185
hygroscopic, 166
materials forming, 166
measurement, 148
methods, 182
mechanical sources, 187
nature, 165
nonhygroscopic, 166
number of, 168, 182
optical effects, 188
physical state, 183
sea salts as, 166
size, 167, 182
measurement of, 167
size distribution, 167
size range, 183
sources, 165
Condensation, on cloud elements, 204
Condenser, aspiration, 147
limiting mobility, 147
Conductivity, atmospheric, measurement, 146
electrical, of air, 102
Cone of escape, 247
Conrad's minor sea breeze, 658
Constant-pressure terminology, 773
Continental drift, hypothesis of, 1012
Contour microclimates, 997
Contrast, reduction by atmosphere, 92
Convection, cellular, 203, 1255
effect of shear, 1258
heat gain by atmosphere through, 1065
heat loss from oceans through, 1064
laboratory investigations, 1255
natural, 507
problem of, 506
theory, 1256
free, laboratory experiments on, 1253
Convergence lines, 874
Cooling power, 1115
Cooling temperature, 1115
Coordinate systems, perturbation equations for, 410
Coriolis force, effect on flow under inversion, 432
Corona, 380
and related phenomena, 71
maxima, 72
Corpuscular theory, of aurorae, 351
Cosmical meteorology, 377
Cosmic radiation, and ions, 103
lunar tidal variations of, 521
Craig, calculations of heating and cooling in ozone layer, 299
Crop yields, weather conditions and, 988
Cryology, 1048
Crystalline frost, 207
cup crystal, 207
dendritic crystal, 207
feather-like form, 207
needle form, 207
plate crystal, 207
Cumulonimbus, ice phase development in, 194
liquid-water concentration, 1199
Cumulus
convection, 694
predicting, 700
development, convectively unstable air and, 699
prediction, 699
research needed, 700
surface heating and, 699
effect of cloud seeding on, 239
entrainment, 694
concept, 695
growth, 697
ice phase development in, 194
increased precipitation from, by cloud seeding, 232
liquid-water concentration, 1199
mixing, concept, 695
solid precipitation from, 223
Cup crystal frost, 207
Curie unit, 155
Current, continuous, with lightning discharges, 138
electric, air-earth, 113
lightning discharges, 139
ohmic, 146
peaks, of lightning discharges, duration of, 140
wave shapes of, in lightning discharges, 140
saturation, 147
semi-saturation, 147
vertical, in thunderstorms, 150
Current circuit, for electrometers, 144
Cyclogenesis, 618, 770
in extratropical latitudes, and hydrodynamic instability, 442
Cyclolysis, 780
Cyclones, *see also* Disturbances
and long upper waves, 605
Antarctic, 922
deepening, change of vorticity field, 612
degeneration, 778
developing, kinetic energy for, 545
development, basic equations, 465
of individual, 608
quantitative theory, 464
dynamic instability, 584
extratropical, 577
aerology of, 599
critical eccentricity, 578
dynamics of simplified models, 577
friction layer of closed vortex, 579
frontal wave movement, 582
frontogenesis, 590
inertial motion in isentropic surfaces, 583
instability line in, 648
pressure changes, theory of, 577
profile of cold front, 594
profile of warm front, 594
profile through frontogenesis, 593
Cyclones—*continued*
extratropical—*continued*
slowing down, 580
structure of maturing frontal, 595
surface pressure tendency, 577
synoptic evolution of upper layers, 587
synoptic example, 587
temperature distribution, 580
thermal structure, 577
three-dimensional movement in, 610
upper-air divergence, 597
vorticity analysis, 582
westerlies associated with, 579
westerly wave of upper atmosphere, 583
families, 608
frontogenesis, 618
generation, 778
high-level, 616
formation, 612
laboratory investigations, 1236
mature, pressure distribution, 611
movement, forecasting, cirrus-cloud movement and, 1171
problem, remarks on, 618
systems, laboratory investigations, 1238
tropical, *see also* Tropical cyclones
permanent circular vortex and, 448
upper streamlines, decrease of vorticity along, 611
wave, pressure distribution, 611
with instability line, 648
Cyclonic circulation, during occlusion process, 610
perturbations, and long upper waves, 607
wind shear, and instability line, 649

D

Dalton's law, 320
Dark and bright segments, movement, 74
Dark segment, 73
height at various sun depressions, 75
schematic version, 75
Dart leader, 136
De-excitation, by collision, 267
radiation with, 267
Dendritic crystal frost, 207
Dendrochronology, 1024; *see also* Tree ring indices
changing drought areas, 1028
cycle analysis, 1028
dry and wet years, 1028
errors of interpretation, 1025
history, 1024
long-term climatic fluctuations, 1028
methods, 1025
possible errors, 1025
principles, 1024
relation to neighboring sciences, 1025
research needed, 1029
Deposit gauge, for measuring atmospheric pollution, 1139
Diabatic processes, changes in relative height associated with, 784
Dielectric materials, used in measuring instruments, 144
Diffusion, atmospheric, 320, 492, 1140
eddy, 321, 501
horizontal coefficients of, 330
effect in nuclei destruction, 189
effects of sinks, 325
effects of sources, 325
equation of 322, 503
forced electromagnetic fields and, 323
from continuous point source, 503, 1140
from multiple sources, 503, 1146
in the vertical, steady conditions, 323
unsteady conditions, general equation, 327

Diffusion—*continued*
in upper atmosphere, 320
laboratory investigations, 1252
matter, in nonadiabatic gradients, 507
maximum mixture, 324
maximum separation, 324
molecular, 320
of particulate matter, 326
partial separation, 324
research needed, 332
small-scale, 503
theory of, 320, 503
three-dimensional, 329, 504
three-dimensional, examples of, 330
general causes, 329
time required for establishing equilibrium, 328
two-dimensional, 503
Diffusion coefficients, observed, 321
Diffusion diagram, 322
Diffusion equilibrium, 320
Diffusion processes, effect of representative wind field on, 330
Diffusion theory, Richardson's, 505
Diffusion velocity, in maximum mixture, 324
vector of, 320
Dissociation energy, 263
Dissociation potential, 263, 269
Dissociative recombination, 270
Disturbances, extratropical, aerology of, 599; *see* Cyclones
Divergence, and pressure change, 643
direct measurement, 639
distribution of, theory of, 642
effects of, 643
influence of scale on, 639
large-scale, distribution of, 641
Dobson spectrophotometer, 276
Doldrums, air motion, 862
Droplet size, electric field and, 133
importance in separation of free electrical charges, 132
Dry ice, cloud modification by, 229
Dust, albedo, 26
effect in scattering solar radiation, 23
volcanic, effect on skylight polarization, 79
Dust counter, Aitken, 165
Dust devils, 679
Dust particles, measurement, 148
Dynamic forecasting *see* Forecasting, dynamic
Dynamic instability, *see also* Hydrodynamic instability, 584

E

Eady, baroclinic wave studies, 459
Earth, albedo, 27, 31
meteorological parameters, 392
shadow, 73
surface, albedo of, 27
heat exchange at, 50
Ebert ion counter, 147
Eddy diffusion, 321, 501
horizontal coefficients, 330
in microclimate, 996
Electric charge separation, 116
Electric current, air-earth, 113
density, over thunderheads, 114
Electric field, and droplet size, 133
atmosphere, measurement, 148
effect on electrical conductivity of air, 104
of atmosphere in fair weather, 107
Electrical apparatus, protection against lightning, 142
Electrical charges, by adhesion, 1134
free, importance of droplet size in, 132
on individual droplets, 130
observed on precipitation, 128
separation of, 132
Electrical conductivity, of air, 102
variations, 103
Electrical potential gradient of air, 108
Electrical properties of snow, 227
Electricity, atmospheric, 128
instruments for measuring, 144
dielectric materials used in, 144
measurement, 150
measuring equipment, auxiliary, 144
methods for measurement of, 144
outstanding problems, 118
present-day research, 150
space charge, measurement of, 149
supply current and thunderstorms, 113
universal aspects, 101
vertical current, measurement, 149
precipitation, 128
in cloud elements, 130
in continuous rain, 129
in electrical storm rain, 129
in snow, 129
in squall rain, 129
in thunderstorms, 129
practical problems, 128
thunderstorm, cloud structure and, 687
current, 150
field, 150
investigations of, 150
lightning discharges, 150
precipitation charge, 150
radio interference, 150
Electrification, of aircraft flying through precipitation, 133
Electrode effect, 108
Electrograms, 109
Electromagnetic cgs system, 144
Electromagnetic fields, forced diffusion and, 323
Electrometer, circuits, 144
types, 144
use in measuring atmospheric electricity, 144
Electron affinity, 263
Electron volt, 264
Electronic configuration, 263
Electrostatic cgs system, 144
Electrostatic charges, of condensation nuclei, 187
Electrostatic induction, discharge, measurement, 148
Elsasser's diagram, 37
Emanometers, types, 156
Emanometrical measurements, circuit diagram for, 156
Emanometry, 155
correction of alpha rays, 156
effect of atmospheric pressure and temperature, 156
saturation correction, 156
time correction, 156
Emission-spectral evidence, upper atmosphere composition, 271
Emission spectrum, 264
Emitting layers, upper atmosphere, heights of, 271
Empire State Building, lightning discharges to, 139
Energy, concept of, 483
in latent form of water vapor, exchange rate of, 1061
kinetic, law of, 484
potential, conversion to kinetic energy, 458
total, global balance, 572
basic equations, 572
transmission of, by means of finite disturbances in inversion height, 429
Energy change, 483
Energy equations, 483
applications of, 489
basic, 486
for averaged properties, 488
in barotropic atmosphere, 462
stress values, 490
Energy laws, basic, 485
Energy levels, 263
Energy principles, application to general circulation, 568
Energy transformations, in the atmosphere, 490
over oceans, 1057
Energy units, 264
Entrainment of air by clouds, 202, 695
in thunderstorms, 683
Eocene age, climate, 1004
Equations, energy *see* Energy equations
Eulerian, use in dynamic forecasting, 480
Equatorial front, 865
Equatorial meteorology, 881
air masses, 881
equations of motion, 883
forecasting, 885
frontal phenomena, 882
Escape, cone of, 247
Eulerian equations, 405
use in dynamic forecasting, 480
Eulerian system, perturbation equations, 407
Evaporation at cloud edges, 200
Evaporation, at sea, direct measurements, 1071
history, 1071
detailed study, 505
determining, 1049
effect on supercooled clouds, 228
factor in hydrology, 1049
from oceans, 1071
Bowen ratio, 1058, 1073
computed from energy considerations, 1073
computed from meteorological elements, 1074
theoretical considerations, 1075
determined by extrapolations of values from land, 1071
energy available for, 1073
from rough surfaces, theoretical considerations, 1077
of cloud droplets, 204
Evaporation factor, as function of wind velocity, 1078
from climatic data, 1079
from meteorological observations, 1078
Evaporation rates, at various latitudes, 1072
Eve number, 146
Excitation by collision, 267
Excitation energies, 263
Excitation potentials, 263
Excited states, 263
Exosphere, 247
Expansion wave, and pressure jump, interaction between, 428
storm, 432
Exponential atmosphere, monochromatic absorption, 265
Extratropical cyclones *see* Cyclones, extratropical
Eye, human, light-distinguishing properties, 93
visual range, calculation, 93
hurricane, dynamical aspects, 907
thermal structure, 905
tropical cyclones, 891

F

Faculae, 379
Fallstreifen, appearance of, 192
trails, shape, 196
Fata Morgana, 66
Feather-like frost, 207
"Fido," 178
Filter method, 149
Fire-danger meters, 988
Fire day, 988
Fire weather, 988
Fireballs, 356

Fjørtoft, baroclinic wave studies by, 459
Flame collector, 148
Flocculi, 379
Flood routing, 1051
  storage routing, 1051
Flood water run-off, applied climatology and, 987
Flow, steady-state, under inversion, equations for, 430
  time-dependent, as forecast tool, 430
    under an inversion, 424
  two-dimensional steady, 431
  under an inversion, 424
    effect of Coriolis force on, 432
Flow patterns, midtroposphere, prognosis, 810
Fluid currents, incompressible, perturbations in, 412
Fluid motion, baroclinic instability in, 468
  centrifugal instability in, 467
  generalized vertical-overturning instability, 468
  gravitational-centrifugal instability in, 467
  gravitational instability, 467
  Helmholtz instability and, 468
  instability, general theory of, 466
Foehn, 667
  clouds associated with, 667
  desirable studies, 671
  development, 668
  free, 668
  high, 668
  islands, 668
  north, 668
  nose, 667
  pauses, 669
  south, 667
  thermodynamic explanation, 667
Fog, 1179
  advection-radiation, 1184
    forecasting, 1186
  advective marine, drop size, 170
  air-mass, 1183
  artificial dissipation, 178
  classification, 1183
  clearing, forecasting, 1187
  cold front, 1184
  definition, 1179
  dissipation, artificial, economics of, 178
  effect of atmospheric pollution on, 1148
  forecasting methods, 1185
  formation, 1179
    drop-size distribution, 1180
    liquid-water content, 1180
    over snow cover, 1181
    physical processes, 1181
    research needed, 1182
    role of condensation nuclei, 1179
  frontal, 1183
  land- and sea-breeze, 1183
  liquid-water content, relation between visibility and, 1180
  mixing-radiation, forecasting, 1187
  postfrontal, 1184
  pre-warmfrontal, forecasting, 1187
    restricted heating, 1185
  radiation, 1184
    drop size, 170
    forecasting, 1185
    graph, 1185
  sea, cloud drop-size-distribution in, 173
    forecasting, 1185
  smog, 1179
  smoke, 1179
  steam, 1183
  synoptic aspects, 1183
  upslope, 1183
Fogbow, 70
Forecast districts, 768
Forecast problem, 731
  essential nature, 732
  failure to assess, 731
  scope, 732
Forecast terminology, 842
Forecast verification, 841
  adjustment for difficulty, 846
  contingency-table summaries, 843
  control forecasts, 846
  correlations, 845
  forecast-reversal test, 847
  methods, 843
  percentage correct, 844
  pitfalls, 847
  probability statements, 845
  problem of, 841
  purposes, 841
    administrative, 841
    economic, 841
    scientific, 842
    selection, 843
  research needed, 847
  root-mean-square-error, 844
  scores, arbitrary, for special purposes, 846
  scoring, 843, 844
Forecasters, demand for, 731
  selection, 743
  training, 743
Forecasting *see also* Forecasts
  aids, 733
  air mass modifications, 779
  air current modifications, 779
  aircraft icing, 793
    conditions, 1201
  analogue, 738
  anticyclogenesis, 770
  anticyclones and, 626, 778
  approach, 744
  area forecast, 767
  by statistical methods, 849
  cold air-mass hydrometeors, 789
  continental factors, 739
  cyclones, 778
  cyclogenesis, 770, 1175
  diffusion conditions in atmospheric pollution, 1151
  dynamic, advective models and, 471
    by numerical process, 470
    computational stability, 477
    frictional effects, 482
    geostrophic approximation and, 470
    linearized barotropic model, 474
    nonadiabatic effects, 482
    nonlinear barotropic equation, integration of, 478
    objective analysis, 479
    use of primitive equations, 480
  equatorial, 885
  extended-range, 802, 814
    by weather types, 834
      research needed, 840
    critique of methods of, 828
    definition, 802
    interaction on a hemispheric basis, 838
    relationship between circulation and weather, 809
    research needed, 812
    scientific bases, 814
    techniques, common factors in, 834
  extrapolation, computational techniques, 738
    mathematical techniques, 738
    techniques, 734
  fog, 1185
  front modifications, 779
  frontal occlusion, 778
  frontal weather, 650
  frontogenesis, 770
  hydrometeors, general, 788
    warm air-mass, 789
  improvements needed, forecasting techniques standardization, 741
Forecasting—*continued*
  improvements needed—*continued*
    recommendations for, 740
    technical, 740
    weather observations, 741
  linear extrapolation, 736
  long-range, 831
    extrapolation of periods, 831
    extrapolation techniques, 734
    multiple correlation tables, 831
    regression equations, 831
    symmetry points, 831
    synoptic extrapolation, 737
    synoptic-statistical techniques, 811
  mathematical, 470
  medium-range, circulation, physical methods in, 806
    upper waves, 806
    French methods, 804
    mean-circulation methods, 829
    methods, 804
    method of singularities, 804
    methods, analogue, 806
      weather type, 806
    problems, 802
    sea-level pressure distribution, 828
    statistical methods, 804
    symmetry-point technique, 806
  midtroposphere flow patterns, 810
  movement of sea-level pressure systems, 776
  objective, 796
    and applied climatology, 986
    combined deductive-empirical methods, 799
    investigation goals, 796
    limitations, 797
    methods of deriving aids, 799
    numerical calculation, 798
    problems, 796
    research needed, 850
    statistical methods, 798
    testing procedures, 850
    types, 798
  ocean waves, 1082
  oceanographic factors, 740
  of various weather constituents, 787
  physical factors, 739
  physical techniques, 739
  practical, general principles, 766
  problems, 731
  projecting centers of action, 811
  regional vs. global approach, 744
  river *see also* River forecasting
  river, 1048
  Rossby's vorticity technique, 739
  scientific, problem of, 744
  short-range, 747
    a procedure of, 766
    aids, objective, 762
    analogues, 737, 750
    basic surface chart, 749
    charts and graphs, 748
    cloudiness, 760
    clouds, 753
    constant-pressure charts, 749, 751
    cyclone movement, 752, 1171
    empirical forecast rules tests, 762
    fog, 761
    guides, climatological-statistical, 762
    main centers of action, influence, 759
    700-mb chart, preparation, 753
    700-mb prognosis, operations in making, 754
    observations, 748
    organization of, 762
    physical techniques, 739
    pilot-balloon charts, 748, 753
    precipitation, 760, 790, 1168
    preforecast study, 758
    present state, 747

Forecasting—*continued*
short-range—*continued*
pressure-change and other auxiliary charts, 749
pressure charts, 749
problem, 747
processing observational data, 748
prognostic chart preparation, 753
radiosonde diagrams, 750
research needed, 763
snow-on-the-ground chart, 749
statistical aids, 762
steps in, 758
steps recommended by Petterssen, 759
surface prognostic chart, preparation, 756, 769
temperature, 761, 788
temperature charts, 749
thermodynamic diagrams, 748
type studies, 750
use of surface data, 749
use of upper-air data, 750
usefulness of various charts, 760
usefulness of various techniques, 760
verification, 762, 841
weather prognosis, 760
wind, 761, 791
zonal index graph, 749
short-term, clouds in, 1167
solar control, 740
statistical, research needed, 852
statistical method, application, 849
statistical techniques of extrapolation, 734
storm swell significance in cyclones, 899
storm tide significance in cyclones, 900
synoptic cloud data as aid in, 1173
synoptic cycle, 766
synoptic techniques of extrapolation, 735
techniques, 731, 733
analogues, 737
extrapolation, potentialities, 737
lag-correlation, 734
linear extrapolation, 736
need for standardization, 741
performance, 732
physical, 739
present practice, 732
synoptic extrapolation, 735
thunderstorm, 690, 1170
tools, 731
tornadoes, 678, 792
tropical cyclone movement, 897, 911
unsatisfactory progress, reasons for, 731
upper-air map prognosis, 780
use of clouds in, 1167
use of microseisms in, 1312
use of radar in, 1279
use of sferics in, 1299
use of stationary time series in, 850
use of surface map in, 769
vertical velocities as tool for, 645
wave formation, 771
weather *see also* Forecasting
weather system movement, dynamical approach, 849
statistical approach, 849
weather-type approach, 835
wind, 791
Forecasts, accuracy of, 769
area, 767
chance, 846
climatological, 846
daily, extended, 733
range, 733
extension to longer periods, 811
general, 767
local, 767
long-range, 733
Forecasts—*continued*
medium-range, analogue methods, 828
combination of synoptics and statistics, 828
persistence, 846
precipitation, contingency table for, 844
short-range, 732
time range, 732
types, 732
verification, 841
Forest fires, and applied climatology, 988
Forest tundra, 953
Forest-tundra ecotone, 953, 957
Forestry, and microclimate, 1000
Fork lightning, 141
Formaldehyde, atmospheric, determining, 1131
Fragmentation nuclei, effect on snow formation, 226
Free-electron density, 334
Freeman waves, in tropical meteorology, 870
Freezing nuclei, 174, 192
effect on snow formation, 224
Frictional effects, dynamic forecasting and, 482
Fringe region, 247
Front, cold, radarscope photographs, 1271
equatorial, 865
occluded, radar pictures, 1274
warm, radar pictures, 1272
Frontal occlusion, 778
Frontal phenomena, intertropical, 882
intertropical convergence zone, 882
meridional, 882
tropical, 882
Frontal precipitation, radarscope photographs, 1271
Frontal strip, 770
Frontal theory, modifications, 867
tropical meteorology, 866
Frontogenesis, 432, 590, 770
profile through, 593
Frontolysis, 780
Fronts, modifications, forecasting, 779
Frost, amorphous, 207
crystalline, 207, *see also* Crystalline frost
-point determination, in lower stratosphere, 313
-point hygrometer, design of, 312
protection, applied microclimatology and, 999
window *see* Window frost
Fulchronograph, in measurement of lightning characteristics, 142, 150

G

Gas chemistry, general principles, 263
Gas diffusion in atmosphere *see* Diffusion
Gas particles, kinetic energy of, 268
Gas reactions, condensation nuclei formation from, 186
Gases, atmospheric, escape, 246
General circulation, *see* Circulation, general
Geographic cycles, 1012
Geographical location, effect on land and sea breezes, 658
Geohydrology, 1048
Geological temperature variations, 1017
Geomagnetic lunar tide, 521
Geostrophic approximation, dynamic forecasting and, 470
Geostrophic current, inertial oscillations of, 440
Geostrophic motion, general conditions for stability of, 437
law of, 435
stability of, 434
Geostrophic zonal wind speed, 555
Gerdien aspirator, 147
Gewitter, electrical characteristics, 128
Glacial periods, climate during, 1004
Glacial sequence, Quaternary Ice Age, 1005
Glaciation, effect of solar radiation variation on, 1010
maximum, climatic conditions in, 381
Glacier research, climatic implications, 1019
Glaciers, as climatic indicators, 1019
research needed, 1022
changes, causes of, 1019
correlation with climatic change, 1020
nature of, 1019
since beginning of Christian era, 1020
within Pleistocene epoch, 1021
within recorded time, 1020
Wisconsin maximum, 1020
Glare, effect on vision, 1119
Glory, 71
Gondwanaland, 1014
Gowan, calculations of equilibrium temperature in ozone layer, 298
Gram-atom, 263
Gram-molecule, 263
Granules, 379
Graupel formation, 205
Gravitational instability, in fluid motion, 467
Green flash, 65
Green segment, 65
Greenland, climatological problems, 961
glacial anticyclone, 961
Hobbs' theory, 961
ice cap, alimentation, 962
radial wind system, 962
wind and weather régime, 961
Grinders, *see* Sferics
Grossturbulenz, 502
Grosswetter, 814
consecutive weather anomalies, correlations between, 817
cosmic influences on, 819
effect of ice conditions, 815
effect of long waves on, 817
effect of moon, 819
effect of ocean currents, 815
effect of periodic fluctuations of climate, 817
effect of planets, 819
effect of polar motion, 816
effect of pressure waves, 816
effect of solar radiation, 820
effect of sunspot cycles, 819
effect of volcanic eruptions, 815
existence of, 814
governing complexes, 815
morphology, 822
problem of rhythms, 816
terrestrial influences, 815
Grosswetterlagen, 803
annual variation, 826
European types, 825
Northern Hemisphere types, 826
types of, 825
Ground, radon exhalation from, 158
Ground inversions, effect on effective terrestrial radiation, 39
Ground reflection, effect on Arago point, 82
effect on skylight polarization, 85
Ground state, 263
Ground wires, use in lightning protection, 142
Grounding, of electrostatic instruments, 145

H

Hail, growth, physics of, 178
Hail insurance, 981

Hailstorms, effect of cloud seeding on, 232
Half-life, 267
Halo phenomena, genetic features, 70
schematic view, 69
Halos, before rain, 1168
Harmonic analysis, generalized, use in times-series prediction, 851
Haze, 1179
radiation by, 42
Haze content, effect on downcoming radiation, 40
Heat, radiation to space, 42
Heat conduction, at cloud edges, 200
in interior of cloud, 200
Heat conductivity, of different materials, 1116
Heat exchange, at Earth's surface, 50
Heat flux, due to atmospheric turbulence, 536
equation, 506
Heat gain, of atmosphere, through convection, 1065
Heat lightning, 141
Heat loss, from ocean, through convection, 1064
Heating, radiational, of stratosphere, 47
Heating degree days, 983
Height, absolute, 772
relative, 772
associated with transisobaric motion, 783
local changes associated with diabatic processes, 784
Heiligenschein, 71
Helium, content of atmosphere, 5
isotopic variation in atmosphere, 8
Helium escape, problem of, 247
Helmholtz instability, fluid motion and, 468
Helmholtz waves, 469
Heterostatic circuit, for electrometers, 144
Hoar crystals, 208
Hobbs' theory of glacial anticyclone, 942, 943, 961
Hochtroposhäre, 287
Horizon, looming, 66
sinking, 66
Horse latitudes, air motion, 862
Housing design, and applied climatology, 982
Hudson Bay, freeze-up, 959
Human bioclimatology *see* Bioclimatology, human
Humidity, and temperature patterns, small-scale, in free air, 318
effect on condensation, 169
effect on electrostatic charge of condensation nuclei, 188
effect on growth of condensation nuclei, 184
effect on growth of various nuclei, 185
in thunderstorm, 686
at tropopause, 315
measurement by aircraft, 1225
measurement, future research, 319
Humidity field, turbulent air flow, 499
Humphrey's volcanic dust theory, 1015
Hurricane rings, characteristics, 432
Hurricane tide, 892
Hurricane wave, 892
Hurricanes, formation, 907
see also Tropical cyclones,
convection theory, 907
frontal theory, 908
hypothesis of Riehl, 909
hypothesis of Sawyer, 909
synoptic conditions during, 908
frontal analyses, 866
movement, 910
recurvature, 911
Hurricanes—*continued*
movement—*continued*
steering method, 910
steering principle, 911
radarscope photographs, 1275
Hydration order, importance for mode of charging nuclei, 1133
Hydraulic jumps, in atmosphere, early work on, 421
Hydrodynamic equations, 404
atmospheric flow, 454
boundary conditions, 406
physical, 403
systems of, 403
Hydrodynamic equilibrium, 446
Hydrodynamic instability, 434
and theory of atmospheric perturbations, 444
as function of latitude, 438
cyclogenesis in extratropical latitudes and, 442
in lower latitudes, 440
quadratic form of Kleinschmidt, 437
Hydrogen, isotopic variation, in atmosphere, 8
Hydrogen content of atmosphere, 6
Hydrogen lines, in spectrum of aurorae, 351
Hydrogen sulfide, atmospheric, determining, 1127
Hydrologic cycle, relation to meteorology, 1048
schematic diagram, 1049
Hydrology, design problems and, 1050
evaporation factor, 1049
fields, 1048
precipitation factor, 1048
relation to meteorology, 1048
snow factor, 1050
Hydrometeorology, depth-duration-area data, 1033
empiricism and theory, 1034
future of, 1045
in United States, 1033
moisture, equations of, 1037
physical background, 1034
precipitable water, 1035
rainfall, nonorographic, 1039
orographic, 1038
recurrence interval, 1043
seasonal distribution, 1043
space distribution, 1042
storage equation, 1037
time distribution, 1042
rainfall depth, 1033
rainfall intensity, 1033
rainfall probabilities, 1042
rainfall volume, 1033
research needed, 1046
reservoirs, wind tides and waves, 1045
scope, 1033
snow melt, 1044
special problems, 1042
vertical motion, 1037
water resources, estimating, 1045
Hydrometeors, cold air-mass, forecasting, 789
general, forecasting, 788
warm air-mass, forecasting, 789
Hygrometer, frost-point, design of, 312
Hypsography, relative, 773
geometrical extrapolation, 780
prognosis by advective-adiabatic extrapolation, 782
prognosis by nonadvective and diabatic extrapolation, 784

I

Ice ages, climate during, 1004
early, climate during, 1006
Ice caps, polar, effect in climatic changes, 1014
Ice clouds, appearance of, 192
growth, 197
physics of, 192
relation to condensation mechanisms, 197
Ice conditions, effect on *Grosswetter*, 815
Ice crystals, formation, 207
by sublimation, 207
physics, 177
in cirrus clouds, 222
size, aircraft measurement, 1229
Ice fogs, 193, 1183
Ice nuclei, behavior of, 192
concentration, 224
effect of expansion-chamber experiments on, 193
effect of temperature on, 192
foreign-particle, temperature dependence, 224
in high-level clouds, 196
in medium-level clouds, 196
in upper troposphere, 193
spectrum, 193
Ice particles, charging, 1134
collision, 116
Ice phase, development in clouds, 194
Ice splinters, rate of formation, 195
Icing, aircraft *see* Aircraft icing
-rate meter, 1227
Idiostatic circuit, for electrometers, 144
Impact-chemistry, 262
Index cycle, 560, 561
circulation patterns, 560
primary, 561
variations, 561
winter's primary, 561
Indian summer anticyclones, 626
Industry, and microclimatology, 1001
Infrared absorption, pressure effects on, 298
Infrared light, 265
Infrared spectra, of ozone, 296
Insects, exchange between earth and atmosphere, 1107
Instability, dynamic, 584
centrifugal, 467
energy, 436
gravitational, 467
hydrodynamic, 434
Instability lines, 647; *see also* Squall lines
and airline operations, 647
and cyclonic wind shear, 649
and pressure gradient, 649
and pseudo-cold front, 650
and thunderstorms, 689
and tornadoes, 647, 651
associated physical factors, 648
cyclone with, 648
general features, 647
physical processes, 650
future research needed, 652
recent studies, 651
synoptic aspects, 648
Thunderstorm Project, 652
Instruments *see* Meteorological instruments
Insulation, wood, use in lightning protection, 142
Insulator, open-air, 145
Inversion, compression waves under, 425
expansion waves under, 425
flow, time-dependent under an, 424
flow under, 424
height, transmission of energy by means of finite disturbances in, 429
simple waves under, 425
steady-state flow equations under, 430
Iodine, atmospheric, 1128
effect on condensation nuclei formation, 186
Ion-capture, 116
Ion clouds, 255

Ion counter, Ebert, 147
Israël, 147
Ion production, measurement, 146
Ion spectrum, measurement of, 147
Ionic equilibrium relation, 103
Ionic mobilities, 120
Ionization, air over earth's surface, 122
meteoric, 257
of air, 102
rate, 121
Ionization energy, 263
Ionization potential, 263, 269
Ionizers, of the atmosphere, 146
Ionosphere, 250, 334
absorption of solar radiation by, 19
aurorae and, 352
D-region, 337
E-layer, during auroral displays, 353
normal, 337
sporadic, 337
E_2-region, 337
electrical conductivity of air in, 106
elementary concepts, 335
F_1-region, 337
F_2-layer during auroral displays, 353
F_2-region, 338
G-region, 338
formation mechanisms, 340
ionization, probable distribution of, 336
layers, rise and fall, effect of atmospheric tides on, 520
methods for studying, 334
principal regions, 335
relations between surface meteorology and, 339
research needed, 340
temperatures, 339
typical data, 335
Ionospheric data, need for more, 257
Ions, atmospheric, problems in field of, 126
chemical adsorption, 130
concentration, atmospheric, measurement, 146
in lower atmosphere, 124
positive to negative ratio, 125
variation with height, 125
cosmic radiation and, 103
formation, rates of, measurements, 147
in atmosphere, 120
large, balance, 123
mean life, 125
measurement, 147
mobility, atmospheric, measurement, 146
mobility values, 121
recombination, measurement, 147
slow, nature of, 121
small, balance, 122
combination coefficients, 122
combination with condensation nuclei, 123
recombination of, 123
Irrigation, artificial, applied microclimatology and, 999
Isentropic charts, use in meteorological analysis, 725
Isentropic convergence, 586
Isentropic divergence, 586
Isentropic shear, anticyclonic, 586
Isentropic surfaces, dynamic instability, 584
inertial motion in, 583
Isentropic upgliding, 592
Isohypses, absolute, 772
relative, 770
total relative, 772
Isotopes, 263
Israël ion counter, 147

J

Jet stream, circulation, variations, 560
in upper troposphere, 604
Jet stream—*continued*
normal characteristics, 553
of circumpolar vortex, 560
planetary, blocking waves in, 432
position and strength, 553
variations, 560
Jupiter, atmosphere, 394
research needed, 397
meteorological parameters, 392
Red Spot, 395, 397
spot periods, distribution, 396
Jurassic age, climate, 1004

K

Kaltlufttropfen, formation, 560, 616, 717
Karandikar, calculations of heating of ozone layer, 299
Katafront, 771
Katallobaric system, 635
Kelvin, theory of atmospheric oscillations, 523
Kern counters, 165, 182
Kibel's method, advective models, 473
Kinetic energy, balance fluctuations, 546
change, 467
dissipation, 571
equation of, 484
global balance, 568
a simple system, 569
equations for the atmosphere, 570
general considerations, 568
of gas particles, 268
of horizontal motions, transfer, 571
Kleinschmidt, quadratic form, 437
coordinate system, direction cosines for, 437
Kolmogoroff's theory, turbulent air flow, 497
Kryophilic surfaces, 225
Kryophobic surfaces, 225
Krypton content of atmosphere, 5

L

Laboratory investigations, atmospheric motions, 1235
cold fronts, 1238
cyclone systems, 1238
cyclones, 1236
experimental cloud formation, 1255
general circulation, 1240
model techniques, 1249
research needed, 1242
rotating-pan experiments, 1244
spherical shell experiments, 1241
thermal convection, 1239
tornadoes, 1236
Lagrangian equations, 405
Lagrangian system, perturbation equations, 408
perturbations in incompressible fluid currents, 412
Land breezes *see* Breezes, land and sea
Landregen, electrical characteristics, 128
Laplace, theory of atmospheric oscillations, 523
Large-ion balance, 123
Layers, double, electrical waterfall effect, 131
phenomena, effect of water purity on, 131
electrical double, 130
Leaders, 136
characteristic data, 100
continuous, 136
dart, 136
mechanism, 136
Lichtenberg figure, in measurement of lightning characteristics, 141
Light, absorption, 264
behavior in atmosphere, 91
emission, 264
rays, dispersion by atmosphere, 67
scattering, in the atmosphere, 76
Lightning, characteristics, measurements of, 141
Lightning arrester, for protection of electrical apparatus, 142
Lightning discharge, 136
branching, 136
channel, 138
characteristics of, 140
charge determination, 140
cloud-to-cloud, 139
cloud-to-ground, phenomena on ground end, 139
composite stroke, 140
continuous current with, 138
current peaks, 139
density, 141
forms of, 141
investigations, 150
ionization process, 136
leaders, 136
mechanism, 136
cloud to ground, 137
multiple stroke, 138
polarity, 141
potential involved in, 138
protection against, 142
radarscope photograph, 1276
return stroke, 137
stroke current, 139
stroke mechanism for high buildings, 138
total charge, 140
wave shape of current peaks, 140
Lightning protection, 142
Lightning stroke *see* Lightning discharge
Limnology, 1048
Line spectra, 265
Liquid-water content, measurement by aircraft, 1226
Long-wave radiation *see* Radiation, long-wave
Long waves, upper and cyclones, 605
Looming, 66
Loschmidt's number, 263

M

Mach lines, 422
Macrometeorology, basic problems, 814
Macroviscosity, in air flow, 496
Magnetic link, in measurement of lightning characteristics, 141
Magnetic storms, aurorae and, 347, 352
Magnetic variations, use in studying upper atmosphere, 251
Maloja wind, 664
Map analysis, pressure jump and, 429
Mars, atmosphere, 392
research needed, 397
meteorological parameters, 392
temperature distribution, 393
winds on, 393
Mean-circulation methods of forecasting, 809, 829
Mean free path, 268
Melting processes, 205
Meridional circulation, idealized, 456
Meridional indices, 564
Meridional variations, of total ozone, 331
Meteor(s), acceleration formula, 359
astronomical background, 356
atmospheric research, 357
based on observation of, 357
atmospheric temperatures and, 306
deceleration, air density from, 360
data, 361
radio measurement, 363
heights, 357
ion cap, 363
ion trail, earth's magnetic field and, 363
mass determination, 360
mass loss rate, 359
photographic observation, 358
methods, 358

Meteor(s)—*continued*
radio observations, 362
showers, 356
radiants, radar observation, 362
spectra, 357
sporadic, 356
stream velocities, radar observation, 363
streams, 356
study, continuous-wave method, 363
Doppler method, 363
whistling meteor method, 363
terminology, 356
trails, radio echoes from, 257
trains, winds and, 358
upper atmosphere research and, 356
visual observations, 357
Meteoric phenomena, use in studying upper atmosphere, 251
Meteoric process, theory of, 359
Meteorite, 356
Meteoroid, 356
Meteorological analysis, 715
centers, 716
coordination of sea-level and upper-air, 720
differential technique, 718
frontal technique, 715
isentropic, 725
moisture charts in, 725
prefrontal squall line, 716
radiosondes and, 726
techniques, 715
transmission, 715
upper air, 717
upper-air technique, 715
use of charts in, 721
use of radar, 717
Meteorological balloons *see* Balloons
Meteorological instruments, 1207
aircraft, 1223
cloud and raindrop size, 1228
humidity measurement, 1225
liquid-water content measurement, 1226
pressure measurements, 1225
special research, 1228
snowflake and ice-crystal size measurement, 1229
temperature measurement, 1223
wind velocity measurement, 1226
altitude measurements, 1226
automatic weather stations, 1217
balloons, 1214
ceilometers, 1218
parachute radiosondes, 1215
radar-wind systems, 1213
radarsonde systems, 1213
radiosonde-radiowind systems, 1207
techniques, 1207
wire sondes, 1216
general trends, 1219
Meteorological optics, general, 61
Meteorological organization, international, need for, 745
Meteorological problems, nonlinear, method of characteristics, 421
Meteorological recorder, 1213
Meteorological research, model techniques, 1249
Meteorological stations, actinometric equipment, 53
Meteorological tool, ocean waves as, 1090
Meteorology, and river forecasting, 1053
Arctic, 942
cosmical, 377
dynamic, problem of, 401
equatorial *see* Equatorial meteorology *and* Tropical meteorology
experimental, development, 229
recent advances, 229
relationship of snow to, 221
marine, 1055
perturbation equations in, 401
polar, 915
thermodynamics of open systems, application to, 531
tropical *see* Tropical meteorology
visibility in, 91
Meteorotropism, 1120
Methane content of atmosphere, 5
future developments, 9
Michelson actinometer, 53
Microbarographs, 370
Microclimate, 993
as climate near ground, 994
contour, 997
dependent, 997
eddy diffusion, 996
effect of type of ground, 995
forestry and, 1000
in vegetation layer, 998
independent, 997
moisture conditions, 995
temperature conditions, 994
wind conditions, 996
Microclimatological research, history, 993
origin, 993
Microclimatology, 993
agriculture and, 998
applications, 998
applied, artificial irrigation and, 999
frost protection and, 999
future problems, 998
in city planning, 1001
industry and, 1001
phenology and, 1000
present state of research, 994
research, methodology of, 1001
wind protection and, 999
Microorganisms, counts, 1109
in clouds, 1109
Microseismic storm, 1312
Microseisms, application to forecasting, 1312
causes, 1303, 1314
effect of differences in earth crust on, 1308
effect of distance on period, 1308
effect of earth faults on, 1308
effect of ground on, 1309
energy transmitted from surf, 1305
in tropical cyclones, 900
observations, 1303
periods of pressure waves in the ocean, theoretical, 1307
research needed, 1310
short-period, 1315
storm, 1313
theory of, 1303
theory of elastic wave propagation in the ground, 1307
theory of wave propagation from surface of ocean to bottom, 1306
theory of wave propagation from water to solid rock, 1306
transmission of energy from meteorological disturbances into the ground, 1304
tripartite studies, 1312
types, 1304
typical records, 1303
wave types observed in, 1307
Midtroposphere flow patterns, prognosis, 810
Mirages, 66
Mistral, 670
Mixing-length hypothesis, in turbulent air flow, 494
Mixing processes, in clouds, 200
Moazagotl cloud, 669
Model techniques, in meteorological research, 1249
prototype simulation, 1250
similitude requirements, 1249
typical laboratory investigations, 1251
Mold transport, in atmosphere, 1109
Molds, origin of, 1109
Moist-adiabatic process, prerequisites for ideal, 199
ascending motions, 201
descending motions, 201
Moisture adjustment, cyclonic-adjustment method, 1040
q-adjustment method, 1040
thunderstorm-adjustment method, 1040
Moisture, equations of continuity, 1037
Molecular diffusion, 320
Molecular forces, dipolar, 1132
dispersion, 1133
for condensation, 1132
induction, 1133
polar, 1132
Molecular particles, 263
see also Atomic particles
Molecular processes, reversibility, 267
Molecular transitions, spectra associated with, 264
Molecules, mean free path, 369
Möller's diagram, 37
Momentum, total angular, conservation of, 460
Momentum-transport theory, 496
Monochromatic absorption, in exponential atmosphere, 265
Moon, effect on *Grosswetter*, 819
enlargement, when near horizon, 62
Mountain building, and climate, 1013
theory of, 385

N

Natural m-sec-v-amp system, 144
Needle circuit, for electrometers, 144
Needle frost, 207
Neon content of atmosphere, 5
Nephelometer, polar, 95
Neptune, atmosphere, 394
meteorological parameters, 392
Neutral points, measurement of position, 80
Newton, theory of atmospheric oscillations, 523
Night-sky, aurora, 345
emissions, altitude, 344
further research needed, 345
intensity, 342
spectrum, 341
wave lengths in, 342
theory, 344
variations, 342
annual, 343
diurnal, 343
latitudinal, 343
twilight effects, 343
with zenith angle, 344
light, variation, 258
luminescence, use in studying upper atmosphere, 251
radiations, from upper atmosphere, 341
spectrum, need for more study, 258
zodiacal light, 345
Nitric acid, atmospheric, determining, 1127
Nitrogen, atmospheric, determining, 1131
bands, in spectrum of aurorae, temperature determination from, 351
content of atmosphere, 5
dioxide content of atmosphere, 7
oxides, effect on condensation nuclei formation, 186
Nitrous oxide content of atmosphere, 5
Noctilucent clouds, use in studying upper atmosphere, 251
Nocturnal cooling process, atmospheric radiation and, 40
Nonadiabatic effects, dynamic forecasting and, 482

Nonhomogeneous thermodynamics system, 534
Nonlinear meteorological problems, method of characteristics, 421
Nordenskjöld line, 955
Northers, 670
Northwesterly flow, over Europe, 603
Nucleation, effect on supercooled clouds, 228
32-nuclei, 193
Nuclei, charging, ballo-electricity, 1134
by adhesion, 1134
mode of, order of hydration for, 1133
condensation *see* Condensation nuclei
counters *see* Kern counters
freezing *see* Freezing nuclei
ice *see* Ice nuclei
ice-particle charging, 1134
sublimation *see* Sublimation nuclei
Numerical integration, by method of characteristics, 422
Numerical process, dynamic forecasting by, 470

O

Observation, weather *see* Weather observation
Observations, data, processing, 748
Occlusion process, cyclonic circulation during, 610
Ocean currents, effect in climatic changes, 1014
effect on *Grosswetter*, 815
Ocean waves, as a meteorological tool, 1090
decay distance, 1082
fetch, 1082
forecasting, 1082
state of sea, 1082
Oceans, energy transformation over, 1057
evaporation from, 1071
heat loss through convection, 1064
Open systems, thermodynamics, application to meteorology, 531
Optics, general meteorological, 61
Orbit, earth, effect on climatic change, 1011
Oscillations, free, in the atmosphere, 306
Oxygen, absorption, 294
absorption coefficients, 275
atomic, ozone and, 272
content, of atmosphere, future developments, 9
variations in atmosphere, 3, 933
isotopic composition, in atmosphere, 8
Ozone, absorption, 293, 294
absorption coefficients, 275
air motion effect on, 283
altitude of mass center, 281
amount, determination of, 276
fluctuations, 278
in upper atmosphere, 278
atmospheric, determining, 1127
theory of, 281
atomic oxygen and, 272
bioclimatology and, 287
conditions in troposphere, 286
constitution of upper atmosphere and, 287
cooling effects, Penndorf's calculations, 299
distribution, effect on heating in ozone layer, 295
in ozone layer, 292
theoretical, 282
effect on condensation nuclei formation, 186
effect on radiation in stratosphere, 46
heating effects, Penndorf's calculations of, 299
height distribution, 262
in stratosphere, 284
Ozone—*continued*
in the atmosphere, 275
in undisturbed photochemical equilibrium, 281
infrared spectra, 296
methods of observation, 275
of air strata near ground, 1129
photochemistry of, 293
radiation flux and, 286
total, annual variations, 331
meridional variations, 331
under conditions of disequilibrium, 283
variations with synoptic situation, 284
vertical distribution, 280
determination, 276
in the troposphere, 7
warm layer and, 287
weather and, 284
Ozone content, of atmosphere, 6
during depressions, 6
during thunderstorms, 7
future developments, 9
methods of determination, 6
variations, annual, 6
diurnal, 6
geographic, 6
Ozone heating, in upper atmosphere, 305
Ozone layer, 292
calculation of heating of, 296
cooling, 296
calculating, 296
data for calculating, 298
equilibrium temperature, Gowan's calculations of, 298
heating, 294
and cooling, Craig's calculations of, 299
Karandikar's calculation of heating of, 299
ozone distribution in, 292
radiative temperature changes in, 292
screening effect, 287
solar effect on, 293
temperature changes, further research, needed, 300
temperature distribution in, 292
Ozonosphere, absorption of solar radiation by, 19
origin of high temperature in middle atmosphere, 253

P

Palaeoclimatic sequence, research needed, 1017
Palaeozoic ice age, climatic changes, 1014
Parasites *see* Sferics
Parcel method, for testing atmospheric stability, 694
Particulate matter, diffusion of, 326
Peat bogs, formation, 1008
Penndorf, calculations of heating and cooling by ozone, 299
Permafrost distribution, 958
Perpendicular velocities, method of, 147
Perturbation equations, 407
Eulerian system, 407
for special coordinate systems, 409
in meteorology, 401
Lagrangian system, 408
long waves on rotating globe, 416
long waves on rotating plane, 415
quasi-geostrophic approximations, 417
quasi-static approximations, 417
trough formula, 415
Perturbation method, use of, 411
Perturbations in incompressible fluid currents, 412
Perturbed particle motion, 435
Petterssen, steps in forecasting, 759
Phenology, microclimatology and, 1000
Photochemical processes, in upper atmosphere, 262
Photochemistry, of ozone, 293
Photographic surge-current recorder, in measurement of lightning characteristics, 142
Photography, of meteors, 358
Photolysis, 268
Photometry, photographic, 80
Photopolarimeter, 79, 80
Photosphere, 379
Physical meteorology, 165
Piezotropy, equation of, 404
Planets, effect on *Grosswetter*, 819
major, atmospheres, 391
atmospheric circulation, 395
Plate crystal frost, 207
Pleistocene epoch, glacier changes in, 1021
Pluvial periods, 1006
Point collector, 148
Points, neutral, distances of in different colors, 86
Polar air, source region, 600
southern limit, 600
subsiding, in troposphere, 315
Polar trough, in tropical meteorology, 869
Polariscope, Savart, 81
Voss, 81
Polarity, of lightning discharges, 141
Polarization, anomalies during twilight, 87
atmospheric, dispersion of, 86
effects of scattering, 79
skylight, 79
at point of maximum elevation, 81
at zenith for different sun's elevations, 81
degree of, 84
deviations, 79
distribution, 81
effect of ground reflection on, 85
effect of volcanic dust, 79
fluctuations of, measurement, 80
magnitude, 81
maximum, correlation between turbidity coefficient and, 84
measurement, 80
research needed, 88
theory of, 83
unsolved problems, 88
visual measurement, 80
Pollen, exchange between earth and atmosphere, 1107
Pollution, atmospheric, 1139
annual variations, 1147
basic equations, 1140
by domestic heating plants, 1151
by multiple sources, 1146
by single sources, 1140
climatological planning for industry, 1152
concentrations distant from the source, 1143
concentrations near the source, 1143
control, 1149
daily variations, 1147
Donora, 1147
effect of atmospheric stability, 1146
effect of rain on, 1147
effect of stack height, 1149
effect of stack temperature, 1149
effect of topography, 1147
effect of wind, 1146
effect on electrical conductivity of air, 109
effect on fog, 1148
effects on meteorological elements, 1148
effect on rainfall, 1148
forecasting diffusion conditions, 1151
from elevated sources, 1142
future requirements, 1152
gases and vapors, 1139

Pollution, atmospheric—*continued*
influence of topography, 1144
limitation, 1149
measurement, 1139
meteorological control, 1149
nuclear effluents, 1150
particulate matter, 1139
coagulation, 1145
deposition, 1145
present position, 1152
research needed, 1154
smog control, Los Angeles, 1150
types of contaminants, 1139
weekly variations, 1147
wind-tunnel investigations, 1145
Postglacial climatic changes, 1006
Postglacial rainfall, 1008
Postglacial succession in Europe, 1007
Potamology, 1048
Potential, involved in stroke formation, 138
Power-law profiles, in turbulence, 496
Precipitation, *see also* Rain, Rainfall
aircraft flying through, electrification of, 133
air-mass, radarscope photographs, 1275
amount, forecasting, 760, 790
Antarctic, 933
continuous, radar detection, 1285
drop coalescence theory, 176
during a pressure jump, 427
electrical, association processes, 131
environmental processes, 131
electrification, basic processes, 130
factor in hydrology, 1048
free electrical charge on, 128
horizontal structure, radar detection, 1288
ice-crystal theory, 175
maximum possible, 1033
orographic, radar pictures, 1274
physics, 165
production of, as related to artificial cloud modification, 235
solid, from cumulus clouds, 223
in free atmosphere, 222
in stratus clouds, 223
types, 221
tropical cyclones, 890
Precipitation charge, in thunderstorms, 150
Precipitation electricity, *see* Electricity, Precipitation
Precipitation processes, 175
Precipitation static, 133
Pre-dissociation, 264
Prefrontal squall line, 427
in meteorological analysis, 716
Pre-ionization, 264
Pressure, at extreme altitudes, 308
atmospheric, measurement by rocket, 306
differences, during land and sea breezes, 657
distribution, during land and sea breezes, 657
during pressure jump, 427
effect on electrical conductivity of air, 104
effect on infrared absorption, 298
function of, absorption coefficient as, 40
in thunderstorms, 685
in upper atmosphere, 303
measurement by aircraft, 1225
sea-level, and 700-mb, 554
Pressure changes, dynamic theory, 631
effect on condensation nuclei, 186
empirical evidence, 636
high-level, 636
high-level processes, 636
historical notes, 630
horizontal advection, 635
mechanism, 630
Pressure changes—*continued*
nonadiabatic temperature changes, 633
theories, 630
thermal theory, 631, 633
vorticity studies, 633
Pressure gradient, and instability line, 649
Pressure jumps, 426
and expansion wave, interaction between, 428
and tornadoes, 431
lines, interaction between, 431
map analysis and, 429
weather associated with, 427
Pressure systems, mechanics of, 575
sea level *see* Sea-level pressure systems
Pressure tendency equations, 632
utility, 632
Pressure waves, effect on *Grosswetter*, 816
use in studying upper atmosphere, 251
Pseudo-cold front, and instability line, 650
Pulse integrator, radar storm-echo signal, 1277
Pulse-packet technique, of radio observation of meteors, 362
Purple light, 73
areal extent, 76
associated with volcanic eruptions, 73
schematic diagram, 76
Pyranometers, 55
Pyrgeometer, Ångström, 55
Bellani, 55
Pyrheliometer, Ångström compensation, 52
silver-disk, 52

Q

Quadrant circuit, for electrometers, 144
Quadrant electrometer, schematic diagram, 145
use in measuring atmospheric electricity, 144
Quantum relation, 264
Quasi-geostrophic approximations, 417
Quasi-geostrophic equation, simplification, 471
Quasi-linear differential equations, method of characteristics, 421
Quasi-static approximations, 417
Quaternary ice age, glacial sequence, 1005

R

Radar, A scope, 1269
antenna gain effect, 1266
aperture of the parabola, effect, 1266
appearance during storms, 1271
attenuation factor, effect, 1268
bands, tropical cyclones, 891, 1275
beam width, effect, 1267
equipment development need, 1280
fraction of beam filled by storm, effect, 1268
measuring meteor shower radiants by, 362
measuring meteor stream velocities by, 363
plan position indicator scope, 1270
pulse length, effect, 1267
R scope, 1270
range-height indicator scope, 1270
received power, effect, 1266
reflectivity per unit volume of storm region, effect, 1268
-scope photographs, air-mass precipitation, 1275
bright band, 1273, 1286
front, cold, 1271
warm, 1272
frontal precipitation, 1271
hurricane, 1275
Radar—*continued*
instability showers, 1275
lightning, 1276
precipitation, patterns, 1274
thunderstorms, 1276
scopes, types, 1269
weather phenomena observation on, 1269
storm detection, background, 1265
bright band, 1286
continuous precipitation, 1285
factors, 1266
future, 1279
horizontal structure of precipitation, 1288
microwave attenuation, 1285
observation, 1283
precipitation growth, bright band and, 1286
refraction, 1284
showers, 1285
theory, 1266, 1283
observational verification, 1269
use in forecasting, 1279
storm-echo signal, analysis of, 1277
contouring, 1278
control of polarization, 1279
photography, 1279
pulse integrator, 1277
spectrograph, 1278
storm observation, 1265
thunderstorm study, 688
transmitted power, effect, 1266
use in Cloud Physics Project, 236
use in meteorological analysis, 717
wave length, effect, 1267
Radarscope photography, 1279
Radarsonde systems, 1213
Radar-wind systems, 1213
Radiation, 11
atmospheric, effect of clouds on, 42
outgoing, 41
relative, scattering of, 38
by haze, 42
calculations, analytical, 45
numerical, 45
cloud, 42
de-excitation with, 267
diagrams, 36
according to Elsasser, 37
according to Möller, 37
diffuse, absorption laws, 35
downcoming, atmospheric, 38
absorption coefficient for, 40
effects, on human bioclimatology, 1118
flux, ozone and, 286
from spectral line, absorption laws, 35
in stratosphere, 46
long-wave, 34
absorption laws, 34
research needed, 47
monochromatic, law, 40
night-sky, from upper atmosphere, 341
of free atmosphere, and synoptic situations, 43
parallel, absorption laws, 34
processes, clouds and, 200
solar, effect of absorption, measurement, 50
effect of molecular scattering on, measurement of, 50
effect of varying distance between sun and earth on, measurement, 50
terrestrial, effective, effect of absorbers on, 39
effect of ground inversions on, 39
effect of temperature on, 39
Radiational influences, on cloud forms, 203
Radiative processes, heat transportation by, 465
Radio duct, wave trapping in, 1293

Radio exploration, of upper atmosphere, 254
Radio fade-out, 248
Radio interference, by thunderstorms, 150
Radio observation of meteors, 362
  pulse-packet technique, of meteors, 362
Radio propagation problems, meteorological aspects, 1290
Radio refractive index, formula for, 1291
Radio sounding, of ionosphere, 336
Radio wave propagation, meteorological problem, 1294
Radio waves, in the atmosphere, refraction, 1290
  reflections from aurorae, 354
  superrefraction *see* Superrefraction
  use in studying upper atmosphere, 251
Radioactive collector, 148
Radioactive substances, in atmosphere, 155
Radioactivity, atmospheric, 155
    measurement, 155
      by emanometry, 155
      by induction method, 157
Radiosonde, flight equipment, 1207
    humidity measurement, 1209
    pressure capsule, 1208
    temperature measurement, 1209
    transmitter, 1210
  ground, wind measurement, 1211
  ground equipment, 1211
  measurement, of water vapor, 313
  method of upper atmosphere exploration, 250
  parachute, 1215
  -radiowind systems, 1207
  shortcomings in cloud observations, 199
Radon, atmospheric, decrease with height, 158
  balance, of atmosphere, 158
  content, atmospheric, 158
    atmospheric, variations, 157
  exhalation, of ground, 158
  source, 155
Rain *see also* Precipitation
  cloud halos before, 1168
  effect on atmospheric pollution, 1147
  electrical storm, electrical characteristics, 129
  insurance, 982
  quiet, electrical characteristics, 129
  shower or squall, electrical characteristics, 129
  torrential, cloud seeding and, 232
Rainbow phenomena, 70
Raindrop size, aircraft measurement, 1228
  measurement, 177
Rainfall, *see also* Rain, Precipitation
  atmosphere clearance of organisms, 1107
  depth, 1033
    maximum, in United States, 1034
  effect of atmospheric pollution on, 1148
  in thunderstorms, 684
  intensity, 1033
  isohyetal maps, 1034
  mass curves, 1034
  nonorographic, 1039
    barrier height adjustments, 1041
    isobar-isotherm relationships, 1041
    latitude adjustments, 1041
    moisture adjustment, 1040
    storm transposition, 1039
  orographic, 1038
    empirical approaches, 1038
    spillover, 1039
    theoretical barrier methods, 1038
  postglacial, 1008
  probabilities, 1042
  recurrence interval, 1043
Rainfall—*continued*
  runoff, forecasting, 1052
    distribution, 1052
  seasonal distribution, 1043
  storage equation, 1037
  tree histories, 1026
  tree-ring indices, 1024
  volume, 1033
Rayleigh scattering, 25
Recombination spectrum, 271
Redox value, determining, 1127
Refraction, astronomical, 64
  phenomena due to atmospheric suspensions, 69
    due to average density gradient, 64
    due to special density gradients, 66
    in cloud-free atmosphere, 64
  radio waves, 1284, 1290
  terrestrial, 65
Replica techniques, for studying snow, 221
Reservoirs, wind tides and waves, 1045
Resonance, 270
Resonance theory, atmospheric oscillations, 524
Return stroke, 136
  characteristics of, 137
Reynolds stresses, 493
Ribbon lightning, 141
Richardson's criterion, 506
Richardson's diffusion theory, 505
Ringelmann chart, for measuring atmospheric pollution, 1140
Rip currents, 1083
River flow, tree histories, 1026
  tree-ring indices, 1024
River forecasting, *see also* Hydrology
River forecasting, 1048, 1051
  flood routing, 1051
  rainfall, runoff distribution, 1052
  role of meteorology in, 1053
  runoff from rainfall, 1052
  service, organization, 1051
  snow-melt, 1053
  water supply, 1053
Road mirages, 995
Robitzsch actinograph, 54
Rocket measurements, in upper atmosphere, 306

S

Saturn, atmosphere, 394
  meteorological parameters, 392
Savart polariscope, 81
Scale domain, of nonturbulent variations, 487
Scattering in the atmosphere, 22, 76, 79
  dust, 23
  molecular, 22
  multiple, 25
  water vapor, 23
Schlieren, 67
Scintillation, 67
  astronomical, 67
  chromatic 67
  mechanism of intensity fluctuations, 67
  terrestrial, 68
  theories of, 68
Sea, breezes *see* Breezes, land and sea
  forecasting, 1083
  ice, distribution, 958
    Hudson Bay, 959
    scope for research, 960
  salts, as condensation nuclei, 166
  surface, exchange of heat with atmosphere, 1057
Sea-level pressure distribution, forecasting, 828
Sedimentation, effect in nuclei destruction, 189
Seeding, cloud, effective, 230
    effects, 230
Seeding, cloud—*continued*
    effects on cumulus clouds, 239
    effects on hailstorms, 232
    effects on stratus clouds, 236
    effects on thunderstorms, 232
    effects on torrential rains, 232
    elimination of supercooled clouds by, 232
    increased precipitation from cumulus clouds with, 232
    limitations, 231
    overseeding, 230
    possibilities, 231
    relation to production of precipitation, 235
    silver iodide, 231
    use in studying atmosphere, 233
    windstorms and, 232
Seismographs, 1313
Sferics, 1297
  direction of arrival, determining, 1297
  frequency of occurrence, 1298
  reception, 1297
  relation to thunderstorms, 1298
  research needed, 1299
  source, 1297
  use in forecasting, 1299
  use in meteorology, 1298
Shadow bands, 67
Sheet frost, 208
Sheet lightning, 141
Shimmer, 68
Ships, weather, 707
  weather observations from, 706
Showers, convective, data on, 699
  instability, radarscope photographs, 1275
  radar detection, 1285
  summer, anticyclones and, 626
Silver iodide, cloud modification by, 230
  use in cloud seeding, 231
Simpson's Ice Age theory, 386
Sinking, 66
Sizzles, *see* Sferics
Sky, apparent shape, 61
    theories concerning, 63
  conditions, effect on apparent shape of sky, 61
  shape, effect of cloud amount on, 61, 64
    phenomena related to, 62
    under various sky conditions, 61
  synoptic types of, 1173
Sky light polarization, 79
  see also Polarization, skylight
Slice method, for testing atmospheric stability, 694
Small-ion balance, 122
Smog, 1179
Smoke, 1179
  condensation nuclei formation in, 186
  filter, for measuring atmospheric pollution, 1139
  shell, use in studying upper atmosphere, 251
Snow, aircraft electrification from, 134
  crystals, 209, 222, 223
    artificial production, 213
      in laboratory, 213
      laboratory apparatus, 214
    capped column, 212
    classification, 209, 221
    columnar crystals with extended side planes, 212
    dendritic, with small plates attached, natural and artificial, 219
    fern-like hexagonal, natural and artificial, 217
    form, relation to external conditions, 214
      upper-air conditions and, 217
    formation, 177, 209, 223
    growth, 215

Snow—*continued*
crystals—*continued*
irregular assemblage of columns and plates, 212
irregular snow particles, 213
malformed, 212
method for taking photomicrographs, 209
natural and artificial, comparison between, 217
needle crystals, 212
practical classification, 213
principles of classification, 209
rimed crystal, 213
replica techniques for studying, 221
scientific classification, 212
spatial assemblage of radiating type, 212
natural and artificial, 219
spatial hexagonal type, 212
stellar crystal with plates at end of branches, natural and artificial, 217
studies, application to meteorology, 217
with irregular number of branches, 212
with twelve branches, 212
electrical characteristics, 129
electrical properties, 227
factor in hydrology, 1050
formation, effect of fragmentation nuclei on, 226
effect of freezing nuclei on, 224
effect of spontaneous nuclei on, 225
effect of sublimation nuclei on, 224
factors controlling, 223
physics, 177
melt, estimate, 1044
forecasting, 1053
relationship to experimental meteorology, 221
Snowflake size, aircraft measurement, 1229
Sodium, atmospheric, 272
chloride condensation nuclei, growth curves, 168
Soil bacteria, exchange between earth and atmosphere, 1106
Solar activity, variable, nature of, 379
periodic character, 379
Solar constant, 13, 17, 380
determination, 50, 52
methods of measurement, 17
numerical value, 18
value, 380
variation, 18, 380
Solar control, of upper atmosphere, 247
Solar effect, on ozone layer, 293
Solar energy, at earth's surface, 28
average conditions, 29
through cloudless sky, 28
through clouds, 28
extraterrestrial, on horizontal surface, 19
in ultraviolet, 295
variations, research needed, 388
weather changes and, 379
Solar prominences, 379
Solar radiant energy, modification by earth, 13
modification by earth's atmosphere, 13
Solar radiation, *see also* Sun, radiation
absorption, spectral by water vapor, 20
total water-vapor, 21
absorption by clouds, 21, 31
absorption by earth's atmosphere, 19
absorption by earth's surface, 22
absorption by ionosphere, 19
absorption by ozonosphere, 19
at ground, 30
effect on glaciation, 1010
effect on *Grosswetter*, 820
Solar radiation—*continued*
factor in climatic change, 1010
outside the earth's atmosphere, 13
scattering, 22
by dust, 23
by liquid water droplets, 24
multiple, 25
spectral, by water vapor, 23
spectral molecular, 22
total molecular, 22
total water vapor, 23
terrestrial effects on, 19
upper-atmospheric heating, 31
Solar rays, illumination of upper atmosphere by, 248
Solar-topographic theory of climatic changes, 1016
Solar ultraviolet radiation, effect on upper atmospheric gases, 248
Solenoidal index, 564
Sound pressure, waves, 374
Sound propagation, in the atmosphere, 366
research needed, 375
through upper atmosphere, 304
Sound rays, in moving air, 367
in quiet air, 366
Sound velocity, 366
and temperature, 374
Sound wave energy, absorption, 368
effect of distance between rays, 368
in atmosphere, 368
Sound waves, abnormal audibility zones, 371
in gases, theory, 366
in moving air, 367
in quiet air, 366
natural, observation, 371
observation, interpretation, 371
recording, instruments for, 370
temperatures derived from, 305
through stratosphere, 371
through troposphere, 371
use in studying upper atmosphere, 251
Sounding balloon, use in studying upper atmosphere, 251
Space charge, measurement, 149
Space medicine, 1121
Spectra, atomic, different ranges of wave length, 265
atomic particles, 265
infrared, of carbon dioxide, 296
of water vapor, 296
molecular particles, 265
Spectral line, radiation from, absorption law, 35
Spectrograph, radar signal, 1278
Spectrophotometer, Dobson, 276
Spectrum, aurorae, 350
hydrogen lines, 351
intensity effects, 350
nitrogen bands, temperature determination from, 351
sodium lines, 351
wave lengths, 350
night-sky, 341
Spicules, 379
Spillover, rainfall, orographic, 1039
Spontaneous nuclei, effect on snow formation, 225
Spores, exchange between earth and atmosphere, 1107
Squall lines, 689; *see also* Instability lines
prefrontal, 716
and pressure jump lines, 427
Stability properties of large-scale atmospheric d sturbances, 454
Standard operative temperature, 1115
Static, precipitation, 134
Steady-state flow, under inversion, equations for, 430
Stooping, 66
Storms, Antarctic, 924
effect on electrical conductivity of air, 110
electrical state in, 129
observation, radar, 1265
radar appearance during, 1271
surge method, of predicting swells, 1098
tropical, frontal analyses, 866
see Tropical cyclones
Stratocumulus, liquid-water concentration, 1199
Stratosphere, composition of air in, 8
lower, frost point determination in, 313
normal circulation, 553
ozone in, 284
radiation in, 46
radiational heating of, 47
sound waves through, 371
Stratus, effect of cloud seeding on, 236
solid precipitation in, 223
forecasting time of dissipation, 1187
liquid-water concentration, 1199
Strays *see* Sferics
Streak lightning, 141
Stroke, return, 136
Sturbs *see* Sferics
Sublimation, formation of ice crystals by, 207
nuclei, 174, 192
effect on snow formation, 224
on cloud elements, 204
Sulfur dioxide, effect on condensation nuclei formation, 186
Sulfur dioxide apparatus, for measuring atmospheric pollution, 1140
Sulfur dioxide content of atmosphere, 7
Sulfuric acid, atmospheric, determining, 1126
Sun, chromosphere, 13
corona, 13
enlargement when near horizon, 62
ionized layers, heights and pressures, 15
photosphere, 13
radiation, *see also* Solar radiation
infrared, 14
fluctuation in far ultraviolet, 15
fluctuations in infrared, 17
fluctuations in near ultraviolet, 16
fluctuations in radio waves, 17
fluctuations in visible, 17
fluctuations of emitted, 15
near infrared, 13
short-wave radio, 14
ultraviolet, 14
unmeasured, 15
visible, 13
Sun, reversing layer, 13
ultraviolet intensity, 1119
Sunlight, average spectral distribution, 13
Sunrise, in upper atmosphere, 248
Sunset, upper atmosphere, 248
Suushine recorders, 55
Sunspot cycles, effect on *Grosswetter*, 819
Sunspot number, relative, 380
and meridional and zonal flow types, 840
Sunspots, 379
follower, 380
leader, 380
Superrefraction, areas subject to, 1290
description of, 1290
explanation of, 1291
propagation, 1294
radio duct, 1292
Surf, forecasting, 1083
Surface friction, heat transportation, 465
Surface map, geometrical extrapolation, 769
physical extrapolation, 770
prognosis, 769

Swell, correlation between wind strength and, 1097
forecasting, 1083
forerunners of wave frequency-spectrum method applied to, 1091
visible, wave height-period applied to, 1090
Synoptic cycle, in forecasting, 766
Synoptic representation, models, 711
techniques, 711
Synoptic upper-air model, 784

T

Telephotometers, 95
Temperature, and humidity patterns, in free air, 318
Antarctic, 929
at extreme altitudes, 308
atmospheric, free oscillations and, 306
measurement by rocket, 306
meteors and, 306
balloon studies on, 303
based on ultraviolet absorption, 305
changes, radiative, in ozone layer, 292
cloud lapse rates, 696
conditions, microclimate, 994
derived from sound wave velocities, 305
differences, during land and sea breezes, 657
distribution, in extratropical cyclones, 580
in ozone layer, 292
in west-wind zone, 599
during pressure jump, 427
effect on effective terrestrial radiation, 39
effect on electrical conductivity of air, 104
effect on sound propagation through upper atmosphere, 304
field, turbulent air flow, 498
forecasting, 761, 788
in anticyclone, 627
in geological time, 1017
in ionosphere, 339
in thunderstorms, 685
in upper atmosphere, 303
measurement by aircraft, 1223
sound velocity and, 374
tree-ring indices, 1024
in Arctic, 1027
tropical cyclones, 889
vertical advection, 644
Témperature résultant, 1115
Theodolite, electronic, 1211
Thermal convection, laboratory investigations, 1239
Thermal low, 781
Thermal processes, effect on cloud elements, 204
Thermal slope wind *see* Wind, thermal slope
Thermodynamic analysis, cloud, 697
Thermodynamics of clouds, 199
Thermodynamics of open systems, 531
application to general circulation, 536
homogeneous systems, 531
differences between open and closed systems, 532
first law, 531
polythermic systems, 533
pseudoadiabatic transformations, 533
second law, 532
temperatures, 533
wet-bulb, 533
equivalent, 533
nonhomogeneous systems, 534
first law, 535
heat flux due to atmospheric turbulence, 535
second law, 535
Thermometer, recording black-bulb, 55
vortex, 1224
Thoron, source, 155
Thunderheads, electric current density, 114
Thunderstorms, 681
and the airplane, 690
cell, life cycle, 682
circulation, 681
cumulus stage, 682
development of new cells, 686
development, preferred areas, 688
dissipating stage, 683
downdraft, thermodynamics, 683
effect of cloud seeding on, 232
electrical characteristics, 129
electrical cycle, 115
electric structure, 115
electricity *see* Electricity, thunderstorm
entraining, thermodynamics of, 683
forecasting, 690
hearth, 689
mature stage, 682
movement, altocumulus and, 1173
pressure, 685
Project, 687
instability line, 652
radar pictures, 1276
rainfall, 684
relation of sferics to, 1298
squall lines and, 689
structure, 681
study, by radar, 688
supply current and, 113
temperature, 685
tropical cyclones, 892
unsolved problems, 691
weather, near surface, 684
wind field, 684
effect of environmental, 688
Tidal variations, in geologic time, 1011
Tides, atmospheric, 255, 257, 306, 510
adiabatic nature, 519
equilibrium solar, 510
geomagnetic lunar, 521
gravitational, 510
Kelvin's theory of, 523
Laplace's theory, 523
lunar, annual mean, 515, 516
annual variation, 515
at Paris, 511
computation methods, 514
daily variations, 510
distribution over North America, 517
geographical distribution, 515, 516
harmonic analysis, 513
harmonic dial, 513
nontropical, 512
probable error, 514
rise and fall of ionospheric layers, 520
search for, 511
temperature effect on, 519
tropical, 512
lunar tidal variations of cosmic rays, 521
lunar tidal wind currents, 519
Newton's theories, 523
research needed, 527
resonance theory, 524
solar daily barometric variations, 521
solar day variations, 510
solar semidiurnal, harmonic analysis, 513
harmonic dial, 513
oscillation, annual variation, 523
oscillations, 521, 522
theory, 523
Time series, non-stationary, 851
schematic, 850
stationary, 850
Topographic theory of climatic change, 1016
Tornadoes, 673
air preceding, 674
and pressure jump lines, 431
appearance, 675
audibility, 676
axis, 676
building safety, 678
definition, 673
flow pattern, 675
forecasting, 678, 792
frequency, 673
generation, theories on, 676
injuries from, 678
instability lines and, 647, 651
insurance, 678
laboratory investigations, 1236
loss of life from, 678
low pressure in, cause of, 677
maintenance, theories on, 676
motion, horizontal, as energy source, 677
vertical, energy for, 676
paths, 676
preliminary signs, 673
pressure, central, 674
pressure drops and, 674
properties, 674
property damage from, 678
protection of life, 678
protection of property, 678
public safety measures, 678
rain and, 674
related phenomena and, 673
scent, 676
source of energy, 651
source of rotation, 651
synoptic weather situations, 673
thunderstorms and, 674
tracking, 678
tropical cyclones, 892
visibility, 675
weather conditions associated with, 673
wind speed, 674
Towering, 66
Trades, boundaries, 864
Trade-wind inversion, 864
Transference, 270
Transisobaric displacement of air, 783
Transmission lines, protection against lightning, 142
Transmittance meters, 94
Tree histories, rainfall in North America, 1026
river flow in North America, 1026
Tree line, Arctic, 955
Tree-ring indices, *see also* Dendrochronology
Arctic temperatures, 1027
of rainfall, 1024
of river flow, 1024
of temperature, 1024
physiological drought, 1028
Tripartite stations, use in study of microseisms, 1313
Tropical air, source region, 600
Tropical air parcel, movement, 610
Tropical cyclones, 887
aerology of, 902
air-mass arrangement around, 888
barometry, 888
classification, according to development stage, 887
according to intensity, 887
clouds, 890
detection, 896
eye, 891, 905, 907
forecasting movement, 897, 910
frequency, 895
height, 893
information sources, 887
inundations, 892
methods of analysis, 903
microseisms, 900
normal storm track, variation from, 897

Tropical cyclones—*continued*
observations, 902
origin, 894
precipitation, 890
pumping, 888
radar bands, 891, 1275
rain area, dynamical aspects, 905
heat transfer, 906
microstructure, 905
thermal structure, 904
upper outflow, 907
wind distribution near surface, 906
region of origin, 895
research needed, 900
schematic development of swells, 899
season of origin, 895
steering currents, 898
storm swell significance, 899
storm tide significance, 900
surface characteristics, 887
surface indications, 896
temperature, 889
theories of formation, 907
thunderstorms, 892
tornadoes, 892
tropopause heights, 893
upper-air indications, 897
vertical structure, theory, 892, 904
wind circulation, 889
Tropical meteorology, 859
see also Equatorial meteorology
air-mass method, 859, 863
achievements, 873
analysis methods, 859
climatological analysis method, 859
climatological method, 859, 860
achievements, 872
tropical storms and, 862
convergence lines, 874
cumulus base height, 861
direct circulation cell, 871
easterly wave, 868
equatorial front, 865
Freeman waves, 870
frontal surface, 865
frontal theory, 866
modifications, 867
perturbation method, 859, 868
achievements, 875
polar trough, 869
surface air temperature, 860
temperature discontinuities, 868
wind persistence, 860
Tropics, weather observation, 709
Tropopause, humidity in, 315
Troposphere, anticyclone in, water vapor structure, 316
chemistry and, 262
intrusion of moist air tongues into dry air, 317
normal circulation, 553
ozone conditions, 286
sound waves through, 371
subsiding polar air, 315
upper, jet stream, 604
vertical distribution of ozone, 7
water-vapor distribution, 315
Trough, intensifying, flow pattern in, 458
temperature pattern in, 458
Trough formula, Rossby's, 415
Trough motion, atmospheric, 563
Tundra, 953
maritime, 957
Turbidity coefficient, correlation between maximum polarization and, 84
Turbulence, and diffusion, 492
aerodynamical background, 492
boundary layer, 492
eddy diffusion, 501
flow near a boundary, 493
heat transfer, 499, 502
influence of density gradient on nature of flow, 506
Turbulence—*continued*
in the general circulation, 502
Kolmogoroff's theory, 497
laminar flow, 492
large-scale processes, 500
macroviscosity, 496
mathematical treatment, 493
microscale of turbulence, 497
mixing-length hypothesis, 494
momentum transfer, 499
momentum-transport theory, 496
nature of, 492
physical features, 498
humidity field, 499
temperature field, 498
velocity field, 498
power-law profile, 496
Reynolds stresses, 493
scale, 497
small-scale diffusion processes, 503
statistical theories of, 496
velocity profile, near a boundary, 495
near rough surface, 495
near smooth surface, 495
vorticity-transport hypothesis, 496
water-vapor transfer, 499
Twilight, polarization anomalies during, 87
Twilight arch, 73
Twilight phenomena, 72
description, 73
problems, 74
results of observations, 73
schematic view, 73
theories, 74
Typhoon *see* Tropical cyclones

U

Ultraviolet, solar energy in, 295
Ultraviolet absorption, temperatures based on, 305
Ultraviolet daylight integrator, for measuring atmospheric pollution, 1140
Ultraviolet intensity, sun and sky, 1119
Ultraviolet radiation, effect of cloudiness on, 1119
solar, effect on upper atmospheric gases, 248
Ultraviolet spectrum, use in studying upper atmosphere, 251
Umkehr curves, 277
Umkehr effect, 276
Unifilar electrometer, 144
Wulf's, 145
Upper atmosphere, 245
absorbing constituents, amounts, 272
absorbing layers, heights, 271
anticyclonic shear, 585
aurorae, 255
balloon studies, 303
composition, 7, 251
absorption-spectral evidence regarding, 270
emission-spectral evidence regarding, 271
constituents, 270
diffusion in, 320
diffusive separation, 256
dissociation of atomic particles, 268
emitting constituents, amounts of, 272
emitting layers, heights of, 271
extreme ultraviolet radiation, 256
fast-charged particles in, 255
illumination by solar rays, 248
ionization, 254
meteoric, 257
atomic particles, 268
luminescence, 254
meteor use in research in, 356
molecular escape, 247
night-sky radiations from, 341
origin of structure, 245
ozone, 275
amount, 278
Upper atmosphere—*continued*
ozone—*continued*
and constitution of, 287
vertical distribution, 280
ozone heating, 305
phenomena, 245
photochemical processes, 262
physical features, 249
physics, general aspects of, 245
pressures, 303
radio exploration, 254
reactions, 270
rocket measurements, 306
solar control, 247
sound propagation through, 304
study of meteors, 253
studying, by V-2 rocket, 250
direct methods, 250
indirect methods, 251
sunrise in, 248
temperature distribution, 252
temperatures, 303
unsolved problems, 256
water vapor in, 311
weather and, 256
winds, 255, 358
Upper-air analysis, 717
Upper-air maps, prognosis of, 780
Uranus, atmosphere, 394
meteorological parameters, 392

V

V-2 rockets, use in determining ozone distribution, 278
use in studying upper atmosphere, 250
Vacuum tube electrometers, use in measuring atmospheric electricity, 144
Vanishing constant, 148
Venus, atmosphere, 391
research needed, 397
meteorological parameters, 392
Vertical current, atmospheric, measurement, 149
recorder, 149
Vertical-overturning instability, generalized, 468
Vertical velocity, large-scale, and divergence, 639
adiabatic method of computing, 640
as a forecast tool, 645
direct measurement of, 640
distribution of, 641
theoretical explanation, 642
effects of, 643
advection of temperature, 644
advection of velocity, 643
changes in cloudiness, 644
pressure changes, 644
influence of scale on, 639
kinematic method of determining, 640
Virga, appearance of, 192
Visibility, in meteorology, 91
research needed, 95
Visual range, calculation, 93
in practice, 95
instruments for measuring, 94
Volcanic dust, effect on skylight polarization, 79
role in climatic changes, 1015
Volcanic eruptions, effect on *Grosswetter*, 815
Vortex, circular, atmospheric flow and, 455
permanent circular, application to tropical cyclones, 448
thermometer, 1224
Vorticity, absolute, for mean zonal flow, 461
isolines for vertical component of, 460
analysis of extratropical cyclones, 582
conservation, 461
equation, 458

Vorticity—*continued*
transport hypothesis, in turbulent air flow, 496
Voss polariscope, 81

W

Warm layer, ozone and, 287
Water, cloud modification by, 230
droplets, liquid, scattering of solar radiation by, 24
-dropper collector, 148
exchange of energy in a latent form of, 1061
purity, effect on double-layer phenomena, 131
resources, estimating, 1045
still, freezing of, 207
supercooled, in clouds, 173
supply forecasting, 1053
turbulent, freezing, 207
vapor, absorption spectral, of solar radiation by, 20
albedo, 26
distribution, in troposphere, 315
in upper atmosphere, 314
effect in scattering solar radiation, 22
effect on radiation diagrams, 37
effect on radiation in stratosphere, 46
in upper atmosphere, 311
infrared spectra, 296
radiation, cooling of atmosphere by, 43
structure, of an anticyclone, 316
upper atmosphere, determination by aircraft ascents, 313
measurement, 311
variations, in density, 8
Waterfall effect, 131
Waterspouts, 679
Waves, barotropic, stability, 462
blocking, in planetary jet stream, 432
complex, under inversion, 425
compression, 423, 425
easterly, structure, 869
expansion, 423
and pressure jump, interaction between, 428
gravity, atmospheric perturbation and, 445
inertial, atmospheric perturbation and, 445
long, on a rotating plane, 415
long upper, cyclonic perturbations and, 607
horizontal dimensions, 606
prognosis based on, 786
motions, in compressible atmosphere, 414
number, 264
ocean, as a meteorological tool, 1090
characteristics, predicted and observed, 1086
decay relationships, 1085
direction measurement, 1093
forecasting, 1082
problems, 1087
frequency-spectrum method applied to forerunners of swell, 1091
future research needed, 1099
generation relationships, 1084
height, significant, 1086
height-period, applied to visible swell, 1090
instrumentation, 1091
interpretation of records, 1093
measurement, 1091
origin, 1092
period, comparison of storm path with, 1096
significant, 1086
propagation diagrams, 1094
instrumentation, 1098
methods, 1098
range, 1099
storm intensity, 1099
storm location by, 1098
Waves—*continued*
ocean—*continued*
recording mechnnism, 1093
significant, forecasting, computations, 1083
spectrograms of, 1094
spectrum, 1086
steepness, relation to wave, age, 1086
storm surge method of forecasting swells, 1098
surface weather map and relation to, 1094
pressure, Antarctic, 926
shape of current peaks, in lightning discharges, 140
shear, atmospheric perturbation and, 445
simple, 423
under inversion, 425
sound, atmospheric perturbation and, 445
Weather, control of, 229
frontal, forecasting, 650
glacial-interglacial sequences, 381
ozone and, 284
relationship of circulation to, 809
Weather anomalies, consecutive, correlation between, 817
persistence tendencies, 817
repetition tendencies, 817
successive in distant regions, correlations between, 818
Weather changes, anomalous, solar energy variations and, 379
Weather conditions, crop yields and, 988
varying, effect on cloud radiation, 43
prediction of various, 787
Weather fluctuations, abnormal, 383
climatic, 382
daily, 383
geographical pattern, 381
geological, 381
periodic character, 381
secular, 383
Brückner cycle, 383
climatic, solar hypothesis, 386
theories, 385
factors controlling, 384
geological, solar hypothesis, 385
theories, 384
intra-century, solar hypothesis, 386
monthly, solar hypothesis, 387
postglacial, solar hypothesis, 386
seasonal, solar hypothesis, 387
secular, solar hypothesis, 386
solar hypothesis, 385
sunspot cycle, 383
weekly, solar hypothesis, 387
week-to-week, theories, 385
Weather forecasting *see* Forecasting
Weather insurance, 981
Weather key-days, 826
and weather development, 826
Weather map analysis, *see* Meteorological analysis
Weather observation, aircraft, 707
Antarctic stations, 708
Arctic stations, 708
ships' reports, 706
tropics, 709
world, gaps in, 705
world network, 705
Weather spectra, long-term predictable, 852
Weather stations, automatic, 1217
Weather types, and clouds, 1175
catalogue, 839
use in analogue selection, 839
extended-range forecasting by, 834
meridional flow, 835
schematic diagrams, 836
zonal flow, 835
Weger aspirator, 147
West wind, anticyclonic shear, upper limit, 604
West wind—*continued*
zone, temperature, 599
wind distribution, 599
Westerlies, and anticyclones, 623
baroclinic, middle latitude, 578
disturbances, formation of, 605
eddy motion, 838
geostrophic, average distribution, 601
mid-troposphere, 561
orographic disturbances, 605
thermodynamic disturbances, 605
waves associated with cyclones, 579
zonal current disturbances, 605
Wet-bulb thermometer temperature and equivalent temperature, 533
Wheat stem rust, transport in atmosphere, 1108
Whirlwinds, 678
Wilson test-plate, 149
Wind *see also* Breeze
bora, 669
circulation, tropical cyclones, 889
conditions, in microclimate, 996
distribution, in west-wind zone, 599
during a pressure jump, 427
effect on atmospheric pollution, 1146
field, environmental, effect on thunderstorms, 688
in thunderstorms, 684
representative, effect on diffusion processes, 330
foehn, 667
see also Foehn
for power production, 984
forecasting, 761, 791
indicators, cloud motions as, 1172
jet-effect, 670
local, 655
definition, 655
desirable studies, 670
Maloja, 664
meridional circulation, 599
mistral, 670
mountain, 662, 663
circulation, components of, 662
desirable studies, 670
theory of, 664
night, vector frequency of, 358
northers, 670
orographic influences, 667
protection, applied microclimatology and, 999
-shift line, marked by clouds, 1173
slope, air transport by, 666
speed, effect on ultraviolet daylight losses, 1146
strength, correlation with ocean swells, 1097
thermal slope, 662
-tunnel investigations, of atmospheric pollution, 1145
upper, Little America, 920
upper atmosphere, 255
meteor trains and, 358
valley, 662, 663
desirable studies, 670
theory of, 664
velocity measurement by aircraft, 1226
Window frost, 207
Windstorms, cloud seeding and, 232
insurance, 982
Wire sondes, 1216
Wulf's bifilar electrometer, 145
Wulf's unifilar electrometer, 145

X

X's *see* Sferics
Xenon content of atmosphere, 5

Z

Zodiacal light, 345
mystery of, 258
Zonda *see* Foehn