THE NATURE AND THEORY OF THE GENERAL CIRCULATION OF THE ATMOSPHERE

BY

EDWARD N. LORENZ

WORLD METEOROLOGICAL ORGANIZATION
1967

NOTE

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the World Meteorological Organization concerning the legal status of any country or territory or of its authorities, or concerning the delimitation of its frontiers.

TABLE OF CONTENTS

FOREWORD

It has become increasingly recognized in recent years that a deeper understanding of the general circulation of the atmosphere and the associated system of climates is a *sine qua non* to further major and much-needed progress in the science of meteorology as a whole and its many practical applications. Indeed, this aim is one of the main features of the research programme of the World Weather Watch, the over-all implementation of which now constitutes the main preoccupation of the World Meteorological Organization (WMO).

The publication of this monograph on *The nature and theory of the general circulation of the atmosphere* is therefore most timely and will doubtless be warmly welcomed by all scientists concerned with the atmosphere. The fact that its author is Dr. Edward N. Lorenz of the Massachusetts Institute of Technology, an outstanding scientist in this field, makes its appearance of particular significance.

The monograph is a unique compilation of existing knowledge in this branch of meteorology, knowledge to which Dr. Lorenz himself has made many notable contributions. At the same time it shows the directions in which further research should now be pursued, so that those concerned directly or indirectly with research in this important field will find the monograph a stimulating and encouraging contribution to the literature.

It seems appropriate in this foreword to place on record the circumstances which led up to this publication. As will be seen from the following paragraphs, the events in question demonstrate that WMO enjoys, in more ways than one, the benefits of the long tradition of international collaboration in meteorology built up over the past century.

The first international conference in meteorology was held in 1853 and the International Meteorological Organization (IMO), a non-governmental body, was created twenty years later. In 1951 WMO began its activities as a governmental organization and a specialized agency of the United Nations. In so doing it took over the functions of the IMO and accepted many new additional responsibilities appropriate to its new status. In taking over the responsibilities of IMO, WMO also took over its modest financial resources.

It was agreed that the surplus left over after discharging various liabilities and obligations should be used to commemorate the old IMO in appropriate ways. To this end the annual IMO Prize was established, to be awarded on an international basis to an outstanding meteorologist. In addition, it was decided to institute an IMO Lecture which would be delivered at each of the four-yearly sessions of the Congress of the Organization, and would take the form of a review of progress in some branch of meteorology or an account of some new advanced theory. An acknowledged expert in the chosen field would be invited to prepare the review, which would then be published by the Organization. The actual lecture would be a condensed version of the review.

The first IMO Lecture was delivered at the Fifth Congress of the Organization, held in Geneva in April 1967. For reasons explained above, the subject *The nature and theory of the general circulation of the atmosphere* was selected, and Dr. Lorenz was invited to prepare the review and deliver the lecture. The present monograph constitutes the full text of Dr. Lorenz's review. The lecture was based on a summary of the review which is included in this present volume in the four official languages of the Organization — English, French, Russian and Spanish.

It would be difficult to imagine a more fitting commencement to the series of IMO Lectures, and, as already explained, it is confidently felt that this contribution to scientific literature will be warmly welcomed on all sides.

In conclusion, I have great pleasure in acknowledging here, on behalf of the World Meteorological Organization, our deep appreciation of the very high scientific standard of the work which Dr. Lorenz has produced and the very friendly and full collaboration which he has extended in all matters relating to the preparation and publication of the monograph.

D. A. Davies

D. A. DAVIES
Secretary-General

SUMMARY

I think the causes of the General Trade-Winds have not been fully explained by any of those who have wrote on that subject...

George Hadley, 1735

The opening words of Hadley's classical paper afford an apt description of the state of the same subject today. Despite many excellent studies performed since Hadley's time no complete explanation of the general circulation of the atmosphere has been produced.

The physical laws upon which an explanation would have to be based are very complicated and not perfectly known. Many theoretical studies have therefore treated only an idealized atmosphere — usually one of uniform composition, enveloping an Earth with a level homogeneous surface, and driven by a heat source not varying with time or longitude. A rigorous treatment of an idealized atmosphere sometimes affords a qualitatively correct although non-rigorous account of the real atmosphere.

The problem of explaining the circulation of even an idealized atmosphere is rendered difficult by the presence of advection — the displacement of the fields of motion and temperature by the field of motion itself. Because the motion is not uniform, different portions of the advected fields undergo different displacements, and the fields become distorted. The variety of patterns which the circulation may assume is therefore far greater than it would be if advection were not present, and the circulation shows little tendency to repeat its past history.

Mathematically the process of advection is manifested by the non-linearity of the governing equations. Because the general solution is non-periodic, it cannot be expressed explicitly with a finite number of symbols. Many theoretical studies have therefore aimed to determine only the characteristic properties or statistics of the general solution.

Closed systems of auxiliary equations whose unknowns are the desired statistics cannot be established, because the original equations are non-linear. The possibility of establishing closed systems consisting of equations and ordered inequalities has not been sufficiently explored.

The only presently feasible procedure for estimating the statistics consists of determining particular time-dependent solutions by numerical means, and evaluating the statistics of these solutions in the manner in which climatological statistics are evaluated from real weather data. The results often appear realistic, but the particular solutions are not always representative, and the procedure does not reveal the relative importance of the separate physical processes.

When averaged with respect to longitude, the advective processes appear as cross-latitude transports of angular momentum and energy. The atmosphere must carry sufficient amounts of these quantities poleward across middle latitudes to balance the amounts which it receives from its environment in low latitudes and gives to its environment in higher latitudes. The required amounts may be carried by a meridional circulation (net equatorward flow at some levels accompanied by net poleward flow at others) or superposed large-scale eddies (cyclones and anticyclones, troughs and ridges). A direct meridional

cell, with equatorward flow below and poleward flow aloft, would carry angular momentum and energy poleward.

Hadley explained the trade winds and prevailing westerlies by noting that heating should produce a direct meridional cell in each hemisphere. The equatorward current at low levels should be deflected by the Earth's rotation to become the trade winds. The returning poleward current aloft should be deflected to become the upper-level westerlies, which upon sinking should become the surface westerlies. In its time Hadley's paper appeared to offer a satisfactory explanation.

Early nineteenth century observations indicated that the surface westerlies drifted poleward rather than equatorward. James Thomson and William Ferrel introduced schemes in which shallow, frictionally-induced, indirect cells occurred in middle and higher latitudes, underneath the larger direct cells. Their explanations also appeared sufficient in their time.

Late nineteenth century observations of cloud motions, culminating in the international cloud observations instigated by the International Meteorological Organization, indicated that the supposed upper-level poleward currents across middle latitudes did not exist. No scheme of meridional cells consistent with the observations could be found which would transport the required angular momentum and energy. Ultimately the zonally symmetric schemes of the circulation had to be abandoned.

Modern observations reveal that large-scale eddies exert a dominating influence upon the zonally averaged circulation by transporting angular momentum and energy poleward across most latitudes. The transport of angular momentum by the eddies is concentrated near the tropopause, and it attains its maximum values near the thirtieth parallels. To complete the balance there must be direct meridional cells in low latitudes, stronger than Hadley's theory would have demanded, and indirect cells in middle latitudes. These cells must extend through the depth of the troposphere.

Since the meridional cells alone do not transport the proper amounts of angular momentum and energy to satisfy the balance requirements, the zonally averaged circulation does not by itself satisfy the dynamic equations. The problem of obtaining pertinent solutions of the equations is therefore much more difficult than it had appeared to be when zonally symmetric solutions were considered sufficient. Any complete explanation of the zonally averaged motion must include an explanation of the configuration of the eddies.

The eddies gain their energy from the zonally averaged circulation in the form of available potential energy, by transporting energy toward latitudes of lower temperature. They supply kinetic energy to the zonally averaged motion by transporting angular momentum toward latitudes of higher angular velocity. To deduce the latter result by treating the eddies as a form of turbulence, one would have to assume a negative coefficient of turbulent viscosity.

Circulations produced in rotating containers of fluid in the laboratory sometimes possess eddies similar in structure to atmospheric eddies. It is thus implied that the physical factors responsible for the presence and structure of the eddies are those which are shared by the atmosphere and the laboratory models. Particular solutions of the dynamic equations obtained numerically also reveal eddies with the proper structure. It is thus implied that the most important physical processes have been incorporated into the equations as they are generally formulated.

For an idealized atmosphere certain specific features of the circulation can be readily explained. First, there must be a circulation, since a state of no motion would be incompatible with the poleward temperature gradient which radiative processes alone would demand. Next, since the kinetic energy of the circulation is dissipated by friction, the poleward temperature gradient must be somewhat less than that demanded by radiation alone, in order that available potential energy may be generated by heating. The poleward pressure gradient must then increase with altitude in agreement with the hydrostatic equation. To balance the pressure gradients the westerly wind component must increase with elevation in approximate agreement with the thermal wind relation, or else there must be a strong downward transfer of northward momentum across middle levels; there appears to be no mechanism for maintaining the latter process. At low levels there must be easterlies at some latitude and westerlies at others, or else no systematic easterlies and westerlies at all; otherwise there would be a net frictional torque which would progressively alter the rotation of the Earth.

One circulation fulfilling these requirements is Hadley's circulation, possibly with Thomson's or Ferrel's modification. This circulation must possess a direct meridional cell to transport the required amount of energy poleward. This cell also transports angular momentum poleward, whence there must be easterly surface winds in low latitudes and westerlies in higher latitudes.

Hadley's circulation and any other zonally symmetric circulations are not observed, because they are unstable with respect to small-amplitude wavelike disturbances of large scale. The observed circulation must therefore possess eddies. The transport of angular momentum by these eddies largely determines the distribution of surface easterlies and westerlies. The structure of the eddies constitutes one of the outstanding aspects of the general circulation not yet theoretically explained.

One approach to the problem is based upon classical turbulence theory. The eddies are assumed to transport angular momentum and energy toward latitudes of lower angular velocity and temperature. There is no physical basis for applying this theory to large-scale eddies, and in any case it yields incorrect results.

Another approach is based upon the theory of baroclinic stability. The large-amplitude eddies are assumed to be similar in shape to the small-amplitude eddies which would amplify most rapidly when superimposed upon the existing zonally averaged circulation. The results are more realistic than those given by classical turbulence theory but they are not in complete agreement with observations, and the physical basis is somewhat uncertain.

The eddies appear to be less irregular than the turbulence approach would suggest, and less regular than the stability approach would suggest. Both approaches assume that the eddies acquire some sort of equilibrium configuration determined by the zonally averaged circulation. It is likely that the eddies cannot be described in this manner since, while attempting to reach any equilibrium configuration, they will produce a new zonally averaged circulation which will in turn demand a new equilibrium configuration for the eddies.

It appears possible that, for an idealized atmosphere, some closed system of equations and ordered inequalities whose unknowns are statistics may be derived; this system might then be solved rigorously for upper and lower bounds for the transport of angular momentum by the eddies across middle latitudes. From such a solution it may be possible to formulate a comprehensible qualitative argument, explaining why the eddies must transport angular momentum poleward, and hence why the trade winds and prevailing westerlies appear where they do.

RÉSUMÉ

Je pense qu'aucun des auteurs qui ont écrit sur les alizés n'a complètement expliqué les causes de ces vents...

Georges Hadley, 1735

La phrase par laquelle débute le mémoire classique de Hadley décrit encore parfaitement l'état actuel de nos connaissances en la matière. En dépit des nombreuses et excellentes études effectuées depuis celle de Hadley, personne n'a pu donner jusqu'à present une explication complète de la circulation générale de l'atmosphère.

Les lois physiques sur lesquelles toute explication devrait se fonder sont très compliquées et encore imparfaitement connues. En conséquence, de nombreuses études théoriques se sont bornées à considérer une atmosphère idéalisée — généralement une atmosphère de composition uniforme, qui entoure une terre présentant une surface unie et homogène, et qui est mue par une source de chaleur dont l'intensité n'est pas soumise à des variations en fonction du temps ou de la longitude. L'atmosphère idéalisée, lorsqu'elle est traitée en toute rigueur, peut parfois donner une description qualitativement correcte, mais qui ne saurait être parfaitement exacte, de l'atmosphère réelle.

Il est difficile d'expliquer la circulation, même dans le cas d'une atmosphère idéalisée, du fait de l'advection — qui consiste dans le déplacement des champs de mouvement et de température sous l'effet du champ de mouvement lui-même. Comme le mouvement n'est pas uniforme, diverses portions des champs soumis à l'advection subissent des déplacements différents, ce qui provoque des distorsions de ces champs. De ce fait, les configurations que peut présenter la circulation sont beaucoup plus diversifiées que s'il n'y avait pas d'advection, et l'histoire de la circulation n'a guère tendance à se répéter.

Mathématiquement, le processus d'advection se traduit par le fait que les équations qui régissent ces mouvements ne sont pas linéaires. Etant donné que la solution générale est non périodique, elle ne peut être exprimée explicitement par un nombre fini de symboles. En conséquence, de nombreuses études théoriques n'ont eu pour objectif que de déterminer les propriétés ou les éléments statistiques caractéristiques de la solution générale.

Les équations originelles étant non linéaires, il n'est pas possible d'établir des systèmes fermés d'équations auxiliaires dont les inconnues seraient les valeurs statistiques qu'on cherche à obtenir. La possibilité d'établir des systèmes fermés composés d'équations et d'inégalités ordonnées n'a pas été suffisamment explorée.

La seule méthode qui soit applicable à présent pour estimer les éléments statistiques consiste à déterminer, par des moyens numériques, des solutions particulières correspondant à des instants différents, et à établir les valeurs statistiques de ces solutions, de la même façon qu'on établit les statistiques climatologiques à partir des données météorologiques réelles. Les résultats obtenus semblent souvent proches de la réalité, mais les solutions particulières ne sont pas toujours représentatives, et la méthode ne met pas en lumière l'importance relative des divers processus physiques qui sont en jeu.

Lorsqu'on en établit la moyenne en fonction de la longitude, les processus advectifs apparaissent comme des transports de moment cinétique et d'énergie s'effectuant perpendiculairement aux parallèles. L'atmosphère doit véhiculer des quantités suffisantes de ces grandeurs à travers les latitudes moyennes, en direction du pôle, pour équilibrer les quantités qu'elle reçoit du milieu qui l'entoure, aux basses latitudes, et celles qu'elle cède à ce milieu, aux latitudes élevées. Les quantités requises peuvent être entraînées par une circulation méridienne — un courant résultant en direction de l'équateur à certains niveaux étant associé à un courant résultant en direction du pôle à d'autres niveaux — ou par des tourbillons de grande échelle se superposant au courant — cyclones et anticyclones, thalwegs et dorsales. Une cellule méridienne directe, dont le courant se dirige vers l'équateur à la base et vers le pôle en altitude, déplacerait des quantités de moment cinétique et d'énergie vers le pôle.

Hadley a expliqué les alizés et les vents d'ouest dominants en notant que l'échauffement devrait provoquer une cellule méridienne directe dans chaque hémisphère. Le courant dirigé vers l'équateur dans les basses couches, étant dévié par la rotation de la terre, donnerait naissance aux alizés. Le courant de retour en altitude, dirigé vers le pôle, étant lui-même dévié, donnerait naissance aux contre-alizés qui, en s'infléchissant vers le sol, deviennent les vents d'ouest en surface. A l'époque de sa parution, il semblait que le mémoire de Hadley apportait une explication satisfaisante du phénomène.

Les observations effectuées au début du dix-neuvième siècle ont révélé que les vents d'ouest en surface dérivaient davantage vers le pôle que vers l'équateur. James Thomson et William Ferrel ont proposé des systèmes dans lesquels des cellules indirectes de faible épaisseur, induites par frottement, apparaissent aux latitudes moyennes et élevées, en dessous des cellules directes plus étendues. A l'époque, leurs explications, elles aussi, semblèrent suffisantes.

A la fin du dix-neuvième siècle, les observations du mouvement des nuages, qui devaient aboutir aux programmes internationaux d'observation des nuages entrepris à l'instigation de l'Organisation météorologique internationale, montrèrent qu'aux latitudes moyennes il n'existe pas, contrairement à ce que l'on supposait, de courant en altitude en direction du pôle. Il ne fut pas possible de concevoir un système, basé sur des cellules méridiennes, qui soit en accord avec les observations et qui rende compte du transport des quantités requises de moment cinétique et d'énergie. Finalement, il fallut abandonner les systèmes de circulation comportant une symétrie zonale.

Les observations modernes révèlent que les tourbillons de grande échelle exercent une influence prépondérante sur la circulation moyenne dans chaque zone, en transportant des quantités de moment cinétique et d'énergie vers le pôle à travers la plupart des latitudes. Le transport de moment cinétique par les tourbillons est concentré au voisinage de la tropopause et il atteint sa plus grande intensité dans les parages des trentièmes parallèles. Pour qu'un équilibre puisse s'établir, il doit exister des cellules méridiennes directes aux basses latitudes, plus intenses que ne l'exigeait la théorie de Hadley, et des cellules indirectes aux latitudes moyennes. Ces cellules doivent s'étendre à toute l'épaisseur de la troposphère.

Etant donné que les cellules méridiennes ne transportent pas, à elles seules, les quantités appropriées de moment cinétique et d'énergie pour satisfaire aux conditions d'équilibre, la circulation moyenne en chaque zone ne satisfait pas, par elle-même, les équations de la dynamique. Il se révèle, par conséquent, beaucoup plus difficile de trouver des solutions pertinentes à ces équations que cela n'était le cas lorsque l'on considérait comme suffisantes les solutions comportant une symétrie zonale. Toute explication complète du mouvement moyen en chaque zone doit comporter une explication de la configuration des tourbillons.

Les tourbillons tirent leur énergie de la circulation moyenne en chaque zone, sous forme d'énergie potentielle disponible, en transportant de l'énergie vers les latitudes où règnent des températures plus basses. Ils fournissent de l'énergie cinétique à la circulation moyenne en chaque zone, en transportant des quantités de moment cinétique vers les latitudes où la vitesse angulaire est plus élevée. Pour obtenir ce dernier résultat en considérant les tourbillons comme une forme de turbulence, il faudrait supposer un coefficient de viscosité turbulente négatif.

Les circulations engendrées en laboratoire, au sein d'un fluide contenu dans un récipient animé d'un mouvement de rotation, présentent quelquefois des tourbillons dont la structure est similaire à celle des tourbillons de l'atmosphère. Il en découle implicitement que les facteurs physiques responsables de la présence et de la structure des tourbillons sont communs à l'atmosphère et aux modèles de laboratoire. Des solutions particulières des équations de la dynamique, obtenues numériquement, révèlent également des tourbillons de structure appropriée. Ceci implique que les processus physiques les plus importants se trouvent incorporés dans les équations, telles que celles-ci sont généralement formulées.

Dans le cas d'une atmosphère idéalisée, certaines caractéristiques spécifiques de la circulation peuvent être facilement expliquées. En premier lieu, il doit exister une circulation, puisque l'absence de mouvement serait incompatible avec le gradient de température entre l'équateur et les pôles que les processus radiatifs impliquent à eux seuls. Ensuite, puisque l'énergie cinétique de la circulation se dissipe par frottement, le gradient de température en direction du pôle doit être quelque peu inférieur à celui qui correspondrait au seul rayonnement, afin que l'énergie potentielle disponible puisse être produite par l'échauffement. Le gradient de pression en direction du pôle doit donc augmenter avec l'altitude, conformément à l'équation de l'hydrostatique. Pour équilibrer les gradients de pression, la composante ouest du vent doit augmenter avec l'altitude, conformément à l'équation de l'hydrostatique. Pour équilibrer les gradients de pression, la composante ouest du vent doit augmenter avec l'altitude, de façon à correspondre approximativement à la relation du vent thermique. Sinon, il doit se produire, aux altitudes moyennes, un fort transfert vers le bas de quantités de mouvement en direction du nord ; il ne semble pas qu'il existe de mécanisme pour entretenir ce dernier processus. Dans les basses couches, il doit y avoir des vents d'est à certaines latitudes et des vents d'ouest à d'autres, ou bien aucun régime systématique de vents d'est et d'ouest. Si ce n'était pas le cas, il apparaîtrait un couple de friction résultant qui modifierait progressivement la rotation de la terre.

La circulation de Hadley, éventuellement avec les modifications de Thomson ou Ferrel, répond à ces critères. Cette circulation doit comporter une cellule méridienne directe pour transporter la quantité d'énergie requise en direction du pôle. Cette cellule transporte également des quantités de moment cinétique, vers le pôle et il doit donc y avoir des vents d'est en surface aux basses latitudes et des vents d'ouest à des latitudes plus élevées.

La circulation de Hadley, pas plus qu'aucune autre circulation à symétrie zonale, ne peut être observée, du fait qu'elle est instable en ce qui concerne les perturbations de faible amplitude en forme d'onde qui se produisent à grande échelle. La circulation observée doit, par conséquent, comporter des tourbillons. Le transport de quantités de moment cinétique par ces tourbillons détermine dans une large mesure la distribution des vents d'est et d'ouest en surface. La structure des tourbillons constitue l'un des aspects marquants de la circulation générale qui n'ont pas encore été expliqués théoriquement.

On peut aborder le problème en se fondant sur la théorie classique de la turbulence. Pour ce faire, on suppose que les tourbillons transportent des quantités de moment cinétique et d'énergie vers les latitudes où règnent une vitesse angulaire et une température inférieure. Il n'existe pas de base physique qui permette d'appliquer cette théorie aux tourbillons de grande échelle et, de toute façon, celle-ci fournit des résultats incorrects.

Une autre méthode d'approche est fondée sur la théorie de la stabilité barocline. On suppose que les tourbillons de grande amplitude ont une forme similaire à celle des tourbillons de faible amplitude, lesquels s'amplifient très rapidement lorsqu'ils sont superposés à la circulation moyenne existant en chaque zone. Les résultats obtenus sont beaucoup plus conformes à la réalité que ceux produits par la théorie classique de la turbulence, mais ils ne s'accordent pas parfaitement avec les observations, et le fondement physique de cette théorie est assez mal assuré.

Les tourbillons semblent être moins irréguliers que la méthode de la turbulence tendrait à le faire croire, et moins réguliers que la théorie de la stabilité le ferait penser. Ces deux théories supposent que les tourbillons atteignent une sorte de configuration d'équilibre, déterminée par la circulation moyenne en chaque zone. Il est probable qu'on ne parviendra pas à décrire les tourbillons de cette manière, étant donné que ceux-ci, en cherchant à parvenir à une configuration d'équilibre, quelle qu'elle soit, provoqueront une nouvelle circulation moyenne en chaque zone qui, à son tour, exigera une nouvelle configuration d'équilibre des tourbillons.

Il semble que, dans le cas d'une atmosphère idéalisée, on parviendra à établir un système fermé d'équations et d'inégalités ordonnées dont les inconnues soient des valeurs statistiques, et à le résoudre ensuite de manière rigoureuse aux limites supérieure et inférieure en ce qui concerne le transport de quantités de moment cinétique par les tourbillons, à travers les latitudes moyennes. A partir d'une telle solution, il sera peut-être possible de formuler un argument qualitatif qui explique pourquoi les tourbillons doivent transporter des quantités de moment cinétique en direction du pôle, et pourquoi les alizés et les vents d'ouest dominants se manifestent là où ils le font.

РЕЗЮМЕ

Я думаю, что никому из писавших на эту тему не удалось полностью объяснить причины возникновения пассатов...

Джордж Хэдли, 1735 г.

Эти слова, которыми открывается классический труд Хэдли, точно характеризуют состояние наших настоящих познаний в этой области. Несмотря на многие замечательные исследования, выполненные после Хэдли, исчерпывающее объяснение общей циркуляции атмосферы до сих пор не найдено.

Физические законы, на основе которых может быть дано это объяснение, чрезвычайно сложны и до конца не выяснены. Поэтому многие теоретические исследования до сих пор имели дело только с идеализированной атмосферой, т. е. атмосферой, однородной по своему составу, обволакивающей земной шар с гладкой однородной поверхностью и приводимой в движение источником тепла, не изменяющимся во времени и пространстве. Строгая трактовка идеализированной атмосферы дает иногда качественно верную, но далеко не точную характеристику состояния реальной атмосферы.

Проблема объяснения процессов циркуляции даже идеализированной атмосферы осложняется наличием адвекции — перемещений полей движения и температуры самим полем движения. Поскольку это движение неупорядоченное, различные участки адвективных полей перемещаются по-разному, и общая картина полей искажается. Поэтому циркуляция принимает гораздо более разнообразные формы, чем это было бы при отсутствии адвекции, и редко обнаруживает тенденцию к повторению.

С математической точки зрения процесс адвекции проявляется в нелинейности основных уравнений. Поскольку общее решение является непериодическим, его невозможно точно выразить при помощи ограниченного числа символов. Поэтому во многих теоретических исследованиях ставилась цель дать определение только главных особенностей или статистических хакактеристик, вытекающих из общего решения.

Замкнутые системы уравнений, где неизвестными величинами являются статистические данные, построить невозможно ввиду нелинейности исходных уравнений. Вопрос о возможности построения замкнутой системы уравнений и упорядоченных неравенств изучен еще недостаточно.

Единственный возможный в настоящее время способ оценки статистических данных заключается в определении численными методами частных решений, связанных временной зависимостью, и оценке статистических величин, вытекающих из этих решений, таким путем, чтобы климатологические статистические данные оценивались на основе данных о реальной погоде. Результаты часто оказываются реалистическими, но частные решения не всегда репрезентативны, и эта процедура не раскрывает сравнительной роли отдельных физических процессов.

При осреднении по долготе адвективные процессы представляются как перенос углового момента и энергии в меридиональном направлении. Этот перенос через средние широты в направлении к полюсу должен компенсировать энергию, получаемую атмосферой в низких широтах и отдаваемую ею в высоких широтах. Необходимое количество энергии может переноситься посредством меридиональной циркуляции, т. е. потоком, направленным к экватору на некоторых уровнях, сопровождающимся потоком в направлении к полюсу на дургих уровнях, или наложенными на них крупномасштабными турбулентными вихрями — циклонами и антициклонами, ложбинами и гребнями.

Непосредственная меридиональная ячейка с экваториальным потоком внизу и полярным наверху приводит к переносу углового момента и энергии к полюсу.

Хэдли объяснял природу пассатов и преобладающего западного переноса таким образом, что нагревание должно вызывать образование непосредственной меридиональной ячейки в каждом полушарии. Направленный к экватору поток в нижних слоях под влиянием отклоняющей силы вращений земли становится пассатом. Обратный поток, движущийся в полярном направлении, отклоняясь, становится западным переносом в верхних слоях, который опускаясь к поверхности земли становится приземным западным ветром. В то время казалось, что теория Хэдли дает удовлетворительное объяснение этих процессов.

Исследования, проведенные в начале девятнадцатого столетия, показали, что приземные западные ветры направлены скорее к полюсу, чем к экватору. Джеймс Томсон и Уильям Феррел предложили схемы, согласно которым под крупными непосредственными ячейками в средних и высоких широтах располагаются мелкие косвенные ячейки циркуляции, вызываемые трением. Это объяснение также представлялось в свое время достаточным.

Проводившиеся в конце девятнадцатого столетия наблюдения за движением облаков, кульминационной точкой которых были международные наблюдения над облачностью, организованные Международной Метеорологической Организацией, показали, что предполагаемых потоков в верхних слоях атмосферы в направлении полюсов не существует.

Схема, объясняющая перенос углового момента и энергии меридиональными ячейками в достаточно больших масштабах, не подтвердилась проведенными наблюдениями. В конечном счете зонально-симметричные схемы циркуляции пришлось отбросить.

Современные наблюдения показывают, что крупномасштабные турбулентные вихри играют доминирующую роль в зонально-осредненной циркуляции, перенося угловой момент и энергию в полярном направлении в большинстве широтных поясов. Перенос углового момента турбулентными вихрями сосредоточен у тропопаузы и достигает максимума в тридцатых широтах. Равновесие должно обеспечиваться наличием непосредственных меридиональных ячеек в низких широтах, более мощных, чем указывал Хэдли, и косвенных ячеек в средних широтах. Эти ячейки, повидимому, захватывают всю толщу тропосферы.

Поскольку меридиональные ячейки не обеспечивают переноса углового момента и энергии в достаточных размерах для достижения равновесия, схема зонально-осредненной циркуляции сама по себе не удовлетворяет требованиям динамических уравнений. Поэтому проблема нахождения соответствующих решений этих уравнений гораздо более сложная, чем это представлялось тогда, когда считались достаточными зонально-симметричные решения. Для того, чтобы дать исчерпывающее объяснение зонально-осредненного движения, необходимо объяснить конфигурацию турбулентных вихрей.

Турбулентные вихри получают энергию из зонально-осредненной циркуляции в форме потенциальной энергии, которая переносится в направлении широт с более низкими температурами. Они питают кинетической энергией зонально-осредненный поток, перенося угловой момент в широты, характеризующиеся более высокой угловой скоростью. Чтобы придти к этому выводу, рассматривая вихри как форму турбулентности, пришлось бы исходить из предположения о том, что турбулентная вязкость имеет отрицательный коэффициент.

Циркуляция, образующаяся при вращении сосудов с жидкостью в лаборатории, иногда порождает турбулентные токи, сходные по структуре с вихрями в атмосфере. Это приводит к мысли, что физические факторы, определяющие наличие и структуру этих вихрей действуют одинаково в атмосфере и в лабораторной модели. Частные решения динамических уравнений, полученные численными методами, также выявляют вихри аналогичной структуры. Таким образом, можно предполагать, что общепринятые уравнения отражают наиболее важные физические процессы.

Некоторые специфические черты циркуляции в идеализированной атмосфере могут быть легко объяснены. Во-первых, циркуляция должна иметь место, поскольку неподвижность атмосферы противоречила бы наличию порождаемого, хотя бы только процессами радиации, температурного градиента, направленного к полюсам. Во-вторых, поскольку кинетическая энергия циркуляции гасится силами трения, температурный градиент в полярном направлении должен быть несколько меньше, чем он должен был бы быть при воздействии одной лишь радиации. Направленный к полюсам барический градиент согласно гидростатическому уравнению в таком случае должен увеличиваться с высотой. Для уравновешивания барического градиента составляющая западного ветра должна увеличиваться с высотой в примерном соответствии с зависимостью термического ветра; в противном случае должен наблюдаться сильный нисходящий перенос момента в северном направлении через средние уровни. Каких-либо данных о наличии факторов, которые вызывали бы этот последний процесс, у нас нет. В нижних слоях в некоторых широтах должны наблюдаться восточные ветры, а в других широтах — западные, или же вообще отсутствие систематических восточных и западных ветров; в противном случае возник бы фрикционный вращающий момент, который прогрессивно изменял бы характер вращения Земли.

Единственный тип циркуляции, удовлетворяющий этим требованиям, — это циркуляция Хэдли, возможно с некоторыми изменениями, предложенными Томсоном и Феррелом. Этот тип циркуляции предусматривает наличие непосредственной меридиональной ячейки, переносящей достаточное количество энергии в полярном направлении. Эта ячейка переносит в полярном направлении также угловой момент; таким образом в низких широтах должны быть восточные приземные, а в верхних широтах — западные ветры.

Циркуляция Хэдли и какие-либо другие зонально-симметричные циркуляции не наблюдаются, поскольку они являются неустойчивыми в отношении крупномасштабных возмущений, имеющих волнообразный характер с малой амплитудой. Поэтому для наблюдаемой циркуляции должны быть характерны турбулентные вихри. Перенос углового момента этими вихрями в значительной мере определяет распределение приземных восточных и западных ветров. Структура вихрей является одной из важнейших проблем общей циркуляции, до сих пор не нашедшей теоретического объяснения.

Один из подходов к решению этой проблемы основывается на классической теории турбулентности. Предполагается, что турбулентные вихри переносят угловой момент и энергию в широты, характеризующиеся более низкими угловой скоростью и температурой.

Физического обоснования для применения этой теории к крупномасштабным вихрям не существует, и в любом случае она дает неверные результаты.

Другой подход основывается на теории бароклинной устойчивости. При этом исходят из предположения, что вихри с большой амплитудой аналогичны по форме вихрям с малой амплитудой, которые быстрее всего развиваются при наложении на существующую зонально-осредненную циркуляцию. Получаемые результаты больше соответствуют реальной действительности, чем результаты, которые дает классическая теория турбулентности, но и они не согласуются с наблюдениями, и их физическая основа несколько неясна.

Турбулентные вихри носят менее нерегулярный характер, чем это должно было быть, исходя из теории турбулентности, и менее регулярный характер, чем это предусматривает теория бароклинной устойчивости. Оба эти подхода исходят из того, что турбулентные вихри приобретают некую уравновешенную конфигурацию, обусловленную зонально-осредненной циркуляцией. Такое описание вихрей вряд ли правильно, так как, принимая уравновешенную конфигурацию, они вызовут новую зонально-осредненную циркуляцию, которая в свою очередь приведет к образованию новой уравновешенной конфигурации вихрей.

Возможно, что для идеализированной атмосферы можно построить некоторую замкнутую систему уравнений и упорядоченных неравенств, в которых неизвестными величинами будут статистические данные; возможно, что при помощи этой системы можно будет дать точные решения в отношении верхней и нижней границ переноса вихрями углового момента через средние широты. На основе такого решения возможно удастся сформулировать исчерпывающий качественный аргумент, объясняющий, почему турбулентные вихри должны переносить угловой момент в направлении к полюсам и, следовательно, почему пассаты и преобладающие западные ветры наблюдаются там, где они есть.

RESUMEN

Yo creo que ninguno de los que han escrito sobre la circulación general de los vientos alisios ha explicado completamente sus causas...

George Hadley, 1735

La frase de Hadley que encabeza su clásico trabajo constituye también una descripción de la situación actual con respecto al mismo tema. A pesar de los numerosos y excelentes estudios que se han llevado a cabo desde la época de Hadley, no se ha conseguido hallar una explicación completa de la circulación general de la atmósfera.

Las leyes físicas en las que tendría que fundarse la explicación de este fenómeno son muy complicadas y no completamente conocidas. En consecuencia, se han hecho muchos estudios en los que se considera una atmósfera ideal, habitualmente de composición uniforme, que envuelve una tierra cuya superficie es homogénea y plana, regida por una fuente calorífica que no varía con el tiempo ni con la longitud. El estudio completo y detallado de una atmósfera ideal permite algunas veces obtener una representación correcta aunque no rigurosa de la atmósfera real.

El problema de explicar la circulación de incluso una atmósfera ideal resulta difícil por la presencia del fenómeno de advección, que consiste en el desplazamiento de los valores de velocidad y temperatura originado por el movimiento de la misma atmósfera. Como el movimiento no es uniforme, los valores sometidos a la advección experimentan desplazamientos distintos según la zona en que se hallen y, en consecuencia, se produce una distorsión en su distribución. La variedad de las estructuras que la circulación puede adoptar es, por lo tanto, mucho mayor que si no existiera la advección. Por otra parte, existe poca tendencia a que los procesos de la circulación se repitan.

Matemáticamente, el proceso de advección se manifiesta por el hecho de que las ecuaciones que lo rigen no son lineales. Debido a que la solución general no es periódica, no puede expresarse explícitamente con un número finito de símbolos. En consecuencia, el objeto de muchos estudios teóricos ha sido determinar únicamente las propiedades características o estadísticas de la solución general.

No se pueden establecer sistemas muy aproximados de ecuaciones auxiliares cuyas incógnitas sean los datos estadísticos que se buscan, debido a que las ecuaciones originales no son lineales. No se ha estudiado suficientemente la posibilidad de establecer sistemas muy aproximados constituidos de ecuaciones y desigualdades ordenadas.

El único procedimiento posible en la actualidad para estimar los datos estadísticos consiste en determinar cada una de las soluciones que dependen del tiempo por métodos numéricos y evaluar los datos estadísticos de estas soluciones de la misma manera que se evalúan los datos estadísticos climatológicos a partir de los datos meteorológicos reales. Los resultados así obtenidos parecen con frecuencia bastante reales, pero las soluciones en cada caso no son siempre representativas y el procedimiento no pone de manifiesto la relativa importancia de cada uno de los procesos físicos independientes.

Cuando las corrientes de advección mantienen una longitud geográfica constante, el proceso de advección se presenta como un transporte de momento angular y energía de una latitud a otra. La atmósfera ha

de transportar cantidades suficientes de energía en dirección al polo y a través de las latitudes medias para compensar la cantidad de energía que recibe del medio que le rodea en las bajas latitudes y que libera en las latitudes altas. Las cantidades de energía necesarias pueden ser transportadas por una circulación a lo largo de los meridianos cuyo movimiento resultante estará dirigido hacia el ecuador en algunos niveles, acompañado otras veces de una corriente dirigida al polo, o por medio de grandes remolinos superpuestos constituidos de ciclones y anticiclones, surcos y cuñas. Una circulación directa a lo largo de los meridianos, constituida de una corriente inferior dirigida al ecuador y otra corriente superior en dirección del polo, transportaría momento angular y energía al polo.

Hadley explicó los vientos alisios y los vientos dominantes del oeste haciendo notar que el calentamiento debe producir una circulación directa a lo largo de los meridianos en cada hemisferio. La corriente dirigida al ecuador a niveles bajos debe ser desviada por la rotación de la tierra para convertirse en los vientos alisios. La corriente superior que retorna en dirección al polo debe ser desviada para transformarse en los vientos superiores del oeste que, al descender, deben constituir los vientos del oeste en superficie. En su época, el razonamiento de Hadley pareció ofrecer una explicación satisfactoria.

A principios del siglo XIX, las observaciones realizadas indicaron que los vientos del oeste en superficie derivaban hacia el polo y no hacia el ecuador. James Thomson y William Ferrel establecieron esquemas que mostraban la existencia de circulaciones indirectas poco profundas e inducidas por fricción, originadas en las latitudes medias y altas, por debajo de las circulaciones directas más amplias. Sus explicaciones parecieron también suficientes en su época.

A finales del siglo XIX, las observaciones del movimiento de las nubes, que culminaron en los programas de observación internacional fomentados por la Organización Meteorológica Internacional, indicaron que las supuestas corrientes en altitud dirigidas hacia el polo y situadas en las latitudes medias no existían. No pudo establecerse un esquema de circulaciones a lo largo de los meridianos que, estando de acuerdo con las observaciones, pudiera transportar el momento angular y la energía necesarios. Ultimamente se abandonó el esquema simétrico zonal de la circulación.

Las modernas observaciones ponen de manifiesto que los remolinos de grandes dimensiones ejercen una influencia dominante en la circulación zonal, transportando momento angular y energía hacia los polos a través de casi todas las latitudes. El transporte de momento angular por los remolinos se concentra cerca de la tropopausa y alcanza sus más altos valores cerca de los paralelos treinta. Para completar el equilibrio, han de haber necesariamente circulaciones directas a lo largo de los meridianos en las latitudes bajas, más fuertes de lo que requería la teoría de Hadley, y circulaciones indirectas en las latitudes medias. Estas circulaciones han de ampliarse necesariamente hasta cruzar por completo el espesor de la troposfera.

En vista de que las circulaciones a lo largo de los meridianos no transportan las cantidades adecuadas de momento angular y de energía para satisfacer las necesidades de equilibrio, la circulación zonal no satisface por sí misma las ecuaciones dinámicas. En consecuencia, el problema de hallar soluciones adecuadas de las ecuaciones resulta mucho más difícil de lo que parecía cuando se consideraban suficientes las soluciones zonales simétricas. Cualquier esquema completo que se haga del movimiento zonal, ha de incluir necesariamente la explicación de la configuración de los remolinos.

Los remolinos obtienen su energía de la circulación zonal en forma de energía potencial libre, al transportar energía hacia latitudes de temperatura inferior. Suministran energía cinética a los movimientos zonales, transportando momento angular hacia latitudes de velocidad angular superior. Para deducir este último resultado considerando a los remolinos como una forma de turbulencia, sería preciso asumir que existe un coeficiente negativo de viscosidad turbulenta.

Las circulaciones producidas en el laboratorio en depósitos giratorios de fluido presentan algunas veces remolinos similares en estructura a los atmosféricos. En consecuencia, se deduce que los factores físicos responsables de la presencia y estructura de los remolinos son los que existen tanto en la atmósfera como en los modelos de laboratorio. Las soluciones particulares de las ecuaciones dinámicas obtenidas numéricamente revelan también la existencia de remolinos con su estructura característica. Esto quiere decir que en las ecuaciones, tal como se formulan en general, han sido incorporados los procesos físicos más importantes.

En una atmósfera ideal pueden explicarse fácilmente ciertas características específicas de la circulación. En primer lugar, ha de haber necesariamente una circulación, ya que la atmósfera estática sería incompatible con el gradiente de temperatura que se observa hacia el polo y que tendría que existir forzosamente como consecuencia de los procesos de radiación solamente. En segundo lugar, como la energía cinética de la circulación se disipa por fricción, el gradiente de temperatura que existe en dirección al polo debe ser algo menor del que exige la radiación sola, con el fin de que se pueda crear energía potencial libre por calentamiento. El gradiente de presión en dirección al polo ha de aumentar entonces con la altitud, de acuerdo con la ecuación hidroestática. Para equilibrar los gradientes de presión, el viento de componente oeste debe aumentar con la altitud, de acuerdo aproximadamente con la relación del viento térmico o, si no es así, ha de haber necesariamente una fuerte transferencia hacia abajo del momento de inercia en dirección norte, a través de los niveles medios ; al parecer no existe ningún mecanismo que explique este último proceso. A los niveles inferiores ha de existir necesariamente viento del este en algunas latitudes y del oeste en otras ; de no ser así, no pueden existir vientos del este o del oeste a ninguna latitud. De no ocurrir así los hechos, tendría que existir un par de fricción resultante que alteraría progresivamente la rotación de la tierra.

Una de las circulaciones que satisfacen estas características es la circulación de Hadley, posiblemente con las modificaciones de Thomson o Ferrel. Esta circulación ha de poseer necesariamente un ciclo directo en la dirección de los meridianos para transportar la cantidad necesaria de energía en dirección al polo. Este ciclo transporta también momento angular hacia el polo y, por lo tanto, deben existir vientos de superficie del este en las latitudes bajas y vientos del oeste en las latitudes altas.

La circulación de Hadley y cualquier otra circulación zonal simétrica no pueden ser observadas debido a que son inestables con respecto a las perturbaciones ondulatorias de pequeña amplitud que se producen en gran escala. En consecuencia, la circulación observada ha de poseer necesariamente remolinos. El transporte de momento angular por medio de estos remolinos determina en gran parte la distribución de los vientos de superficie del este y del oeste. La estructura de los remolinos constituye uno de los aspectos más notables de la circulación general que no han sido aún explicados teóricamente.

Uno de los planteamientos del problema se funda en la teoría clásica de la turbulencia. Se supone que los remolinos transportan momento angular y energía hacia latitudes de menor velocidad angular y temperatura. No existen bases físicas para poder aplicar esta teoría a los remolinos que se producen en gran escala, y en todo caso los resultados son incorrectos.

Otro planteamiento se funda en la teoría de la estabilidad baroclínica. Se supone que los remolinos de gran amplitud tienen una forma similar a los de pequeña amplitud, los cuales se agrandarían más rápidamente cuando estuvieran superpuestos a la circulación zonal ya existente. Los resultados de esta teoría son más realistas que los obtenidos por la teoría clásica de la turbulencia pero no están de completo acuerdo con las observaciones y el fundamento físico no resulta muy claro.

Parece que los remolinos son menos irregulares de lo que sugiere el planteamiento fundado en la turbulencia y menos regulares de lo que se deduce según el planteamiento fundado en la estabilidad. Ambos planteamientos suponen que los remolinos adquieren una especie de configuración de equilibrio determinada por la circulación zonal. Es posible que los remolinos no puedan ser descritos de esta manera, ya que al intentar alcanzar una configuración de equilibrio producirán una nueva configuración zonal que a su vez requerirá una nueva configuración de equilibrio de los remolinos.

Parece posible que, en una atmósfera ideal, se puedan establecer algunos sistemas de ecuaciones aproximados y desigualdades ordenadas cuyas incógnitas sean los datos estadísticos. Este sistema podría entonces ser resuelto rigurosamente en los límites superiores e inferiores, por lo que se refiere al transporte de momento angular, por los remolinos a través de las latitudes medias. A partir de esta solución, quizás sea posible formular un razonamiento cualitativo que explique por qué los remolinos han de transportar necesariamente momento angular hacia el polo y, por lo tanto, por qué los vientos alisios y los vientos dominantes del oeste aparecen en las zonas en que se les observa.

INTRODUCTION

The atmosphere is a fluid whose circulation possesses a highly complex structure. The circulation is governed by a set of laws which are known to a fair degree of precision, and in principle it should be possible to use these laws to deduce the circulation. Nevertheless, the problem of deducing the behaviour of the atmosphere presents many obstacles which have not yet been overcome, and the greater portion of our knowledge of the atmosphere has been the result of direct observation. As a consequence, many of the major advances in our understanding of the atmosphere have followed major improvements in the process of observing it.

The atmosphere recognizes no political boundaries. The weather above one nation is inevitably coupled with the weather above others. The circulation which must be observed if a satisfactory understanding of the atmosphere is to be gained is truly global in extent. Yet, at least in the past, it has not been possible to observe in any detail the weather above one nation except from within that nation. Thus it is that advances in meteorology, perhaps more than in any other science, have been dependent upon a certain degree of international co-operation.

The recognition of the need for co-operation led to a number of international conferences in the middle nineteenth century, and finally to the creation of the International Meteorological Organization in 1873. In its earliest days the IMO was concerned with such basic needs as the exchange of weather information on a routine basis — a prerequisite for the construction of adequate daily weather maps — and the establishment of sufficient uniformity in weather observations to enable the information from different nations to serve a common purpose. Subsequently the IMO fostered such enterprises as the International Cloud Observations of 1896-1897, which played a role in overthrowing the accepted theories of the general circulation of the atmosphere, and in directing the thoughts of meteorologists toward some of the newer ideas.

The International Meteorological Organization was superseded by the World Meteorological Organization in 1951. At the Fourth Congress of the WMO in 1963, it was decided to institute a lecture to be delivered at each session of the World Meteorological Congress. This lecture was to be known as the "IMO Lecture" in commemoration of the International Meteorological Organization.

In consideration of the effort currently being devoted by the WMO to the development of a global observation system, it was decided that the first IMO Lecture should be concerned with the subject of the general circulation of the atmosphere. The present monograph is the result of this decision; the lecture, presented before the Fifth Congress in 1967, was based upon the material contained herein.

The general circulation of the atmosphere means many things to many persons. To some it is the time-averaged state of the atmosphere, with all of its local geographical details. To some it is the instantaneous world-wide state of the atmosphere, whose extended-period fluctuations are responsible for the vicissitudes of the weather. To some it is the collection of permanent and semipermanent synoptic features of the atmospheric circulation, including the intertropical convergence zone, the jet streams, the major semipermanent cyclonic and anticyclonic centres, and the summer and winter moonsoons. To some it is the collection of all quantitative statistical properties of the circulation.

In a monograph of this length it would be possible to consider every aspect of the circulation in a brief and perhaps perfunctory manner, or to treat a few aspects in a more thorough fashion. I have chosen the latter course. Accordingly, a considerable share of the discussion is centred about the nature and cause of the fields of motion, temperature, and moisture, averaged with respect to longitude and time.

It should not be inferred on this account that these fields constitute my own concept of the "general circulation", or that they are necessarily the most important aspects. Possibly they have received the greatest amount of theoretical attention. In reality this choice of emphasis is not so restrictive as it might appear to be. The long-term zonally averaged wind, temperature, and humidity fields are not by any means a closed set of properties, to be accounted for independently of the remaining properties of the atmosphere. Indeed, it has become increasingly apparent that a complete explanation of these features requires a consideration of many if not all of the principal features of the circulation. Accordingly, in presenting a detailed account of some of the time-and-longitude averaged fields, I have necessarily touched upon most of the remaining aspects.

Nevertheless, in order to hold the size of this monograph within reasonable limits, I have found it necessary to omit all but passing reference to several aspects which logically belong in any complete treatment. Three of these are of sufficient importance to merit a word of mention now.

First there is the high atmosphere. The circulation of the atmosphere is global in its vertical as well as its horizontal extent. The effect of what takes place at high levels upon what takes place lower down is however at best difficult to assess, and it is not certain that the tropospheric circulation would be greatly modified if the circulation in the three per cent of the mass of the atmosphere above 25 kilometres could somehow be forced to behave in a different manner. Accordingly, I have restricted the scope of this study by confining attention to the troposphere and lower stratosphere.

Second, I have not gone into detail concerning the fluctuations of the general circulation, which range in duration from the familiar index cycle to the glacial and interglacial periods. An appreciation of these changes is prerequisite to any rational system of extended-range or long-range weather forecasting. Studies of the circulation at different phases of the various oscillations can be a partial substitute for the controlled experiments which we are unable to perform, and they are capable of yielding considerable information concerning the mechanism through which the circulation operates.

Finally, I have not attempted to go into any detail regarding radiation, the process which is ultimately responsible for the existence of the circulation. Here I feel that the mutual interaction between the field of motion and the field of radiation is so complicated that we are only beginning to appreciate its true importance. The frequently heard statement that the circulation would remain nearly the same if only the grossest features of the radiation field were retained receives some support from the laboratory model experiments, where the field of heating is only the crudest approximation to the heating in the atmosphere, but the statement is still only a hypothesis, and it is in need of much careful study. Possibly it is only the grossest features of the circulation which would be nearly the same.

Throughout most of this study the qualitative nature and theory of the circulation have been stressed, even though quantitative statistics are presented, and the reader who wishes to pass over the mathematical equations will find that in most instances, with the exception of Chapter II where the equations themselves are the principal topic, he can still follow the text. It would have been possible to make the discussion completely qualitative, and omit the equations altogether. Nevertheless, I do not feel that this monograph would serve its purpose in the best manner if this had been done.

Although this work is addressed largely to the meteorological world, it is my hope that it may find an audience among those engaged in other fields of study. Accordingly, I have devoted some space to the discussion of such basic meteorological concepts as the definition of the geostrophic wind, which could have been omitted altogether if the work had been addressed to meteorologists alone.

In the course of preparing this monograph I have received assistance in so many forms from so many persons that it is impossible to acknowledge every individual contribution. I have been especially fortunate in having the opportunity to engage in almost daily discussions with my colleague Professor Victor P. Starr, whose ideas concerning the general circulation have always been a source of inspiration. I am also greatly indebted to my colleague Dr. Robert E. Dickinson for reviewing the manuscript in a most critical fashion, and offering numerous suggestions.

I also wish to express my appreciation to the following persons for the assistance in one form or another which they have provided: Professor José P. Peixoto of the University of Lisbon; Dr. Robert M. White, Mr. Jay S. Winston, and Mr. John P. Webber of the Environmental Science Services Administration; Dr. Ralph Shapiro of the Air Force Cambridge Research Laboratories; Dr. Walter O. Roberts, Dr. Chester W. Newton, and Mr. Harry van Loon of the National Center for Atmospheric Research; Dr. Barry Saltzman of the Travelers Research Center; Professors Dave Fultz and George W. Platzman of the University of Chicago; and Professor Reginald E. Newell and Miss Madeleine Heyman of the Massachusetts Institute of Technology. My sincerest thanks go to Mrs. Marie L. Gabbe for the arduous task of preparing the manuscript, and to Miss Isabel Kole for the preparation of the numerous charts and diagrams. Finally, I wish to thank the World Meteorological Organization for making the publication of this monograph a reality.

Edward N. Lorenz
Massachusetts Institute of Technology
February, 1967

CHAPTER I

THE PROBLEM

I think the causes of the General Trade-Winds have not been fully explained by any of those who have wrote on that Subject...

George Hadley (1735)

We have chosen the opening words of Hadley's famous paper for the opening words of this monograph because they seem to afford an apt description of the state of the same subject today. We have no desire to imply that tremendous progress has not been made, because, in the light of today's knowledge, Hadley's remark appears to be a considerable understatement. Yet not in any of the thousand or more excellent works which have appeared since that time, nor in any combination of these works, is a full explanation of the distribution of easterly and westerly winds to be found.

It is evident that the validity of this claim depends very much upon what constitutes a full explanation. It is not to be expected that there will ever be complete agreement on this matter. At this point we shall simply express the opinion that the requisites for a complete answer to a qualitative question differ considerably from those for a complete answer to a quantitative question. Before considering this matter in greater detail, we shall present an account of Hadley's paper, which will serve to illustrate some of the points involved.

Prior to Hadley's time there had been sporadic attempts to account for the trade winds, and one of these which pictured the winds as exhalations from the sargassum weed in the subtropical seas nevertheless found its way into a scholarly journal. In sharp contrast was the notable work of the astronomer Edmund Halley (1686), who presented a detailed and methodical account of the trade winds as observed in three separate oceans, and sought a common cause for them. He rejected an earlier notion that the air by reason of its lightness simply could not keep up with the Earth's surface in its diurnal rotation, and ascribed the north-easterly trades on the north side of the Equator and the south-easterly trades on the south side to the tendency of the air to converge toward the most strongly heated region, as this region progressed about the equatorial belt. For reasons which are not clear he assumed that the cumulative effect of the afternoon tendency to move toward the western sun would outweigh that of the morning tendency to move toward the east.

In concordance with Halley, Hadley concluded that the distribution of solar heating would lead to a general rising motion in lower latitudes and a sinking motion in higher latitudes, the circuit being completed by equatorward motion at low levels and poleward motion aloft, but he rejected the idea that motion toward the sun would lead to any average westward or eastward movement. He then noted that in the absolute sense the Earth's surface moves most rapidly eastward at the lowest latitudes, and he maintained that if the air were initially moving equatorward with no relative eastward or westward motion it would, in attempting to converse its absolute velocity, arrive at lower latitudes moving westward relative to the earth. He found, in fact, that air travelling considerable distances would acquire a much greater westward velocity than any ever observed, and assumed that the frictional drag of the

Earth's surface would in the course of a few days reduce the velocities to those actually found — thus the trade winds.

He next noted that the required counter-drag of the air upon the Earth would continually slow down the Earth's rotation unless opposed by an opposite drag in other regions; this he assumed to occur in the belt of prevailing westerlies in middle latitudes. To account for the westerlies he maintained that the air initially moving directly poleward at high levels would soon acquire an eastward relative velocity, and upon reaching higher latitudes and being cooled would sink and become the prevailing westerlies.

Although Hadley's remarkable paper contains scarcely a thousand words, many hundred thousand words have since been written about it, and it is not surprising to find that some of these have criticized it adversely. One fault requires immediate correction: in the absence of eastward or westward forces, air moving equatorward or poleward conserves its absolute angular momentum rather than its absolute velocity. This tendency to conserve angular momentum is identical with what is now designated as the east-west component of the deflective force, whose proper formulation has been credited to the nineteenth-century mathematician Coriolis among others. But Hadley preceded Coriolis by a century, and perhaps he deserves credit for being as nearly correct as he was. Hadley's error caused him to underestimate the Coriolis force by a factor of two, but since the remainder of his argument was entirely qualitative, his error did not influence it.

Far more significant are his positive contributions. Hadley realized what today seems fairly obvious, that, by reason of continuity of mass, general equatorward motion at one level requires general poleward motion at some other level; and, what is less obvious, that, by reason of conservation of total angular momentum, general westward motion dragging upon the Earth's surface at one latitude requires general eastward motion at some other latitude. His ideas embody the concept of a global circulation, no one of whose major branches can be explained independently of the remaining branches.

Hadley stated that he felt it unnecessary to consider the changes in solar heating with the seasons, and he rejected the diurnal variations of heating, which had played a dominant role in Halley's hypothesis, as having any important effect. He did not consider the presence of oceans and continents, whose contrasting thermal capacities could have destroyed the symmetry of the heating, nor the mountains and other obstacles which could have distorted the flow. He did not consider the presence of water vapour, whose thermodynamic properties were in any event not known in his day. Had he been questioned on these omissions, he might have maintained that these influences would alter the flow to some extent, but not so greatly as to render his arguments invalid.

Many theoreticians today would take a different attitude. They would maintain that what they were studying was not the Earth's atmosphere at all, but an idealized atmosphere, consisting of a gas of uniform composition enveloping a planet with a level homogeneous surface, and driven by an external heat source not varying with longitude or time. They would regard the Earth's atmosphere as only one of many conceivable planetary atmospheres, which in turn comprise but one type of many conceivable types of thermally driven rotating fluid systems. Certainly the general theory of planetary atmospheric circulations is as suitable a subject for theoretical study as the specialized theory of the circulation of the Earth's atmosphere. Moreover, although one cannot deny that simplifications are often made solely to facilitate theoretical treatments, it would appear that, within the collection of possible planetary atmospheres, one which is devoid of irregularities occupies a more central and fundamental position than one with any specific arrangement of irregularities.

It is noteworthy that Hadley adopted an approach which has characterized numerous subsequent attempts to account for the atmospheric circulation, not to mention many other natural phenomena.

He attempted to describe how the final steady circulation which he envisioned would have developed from a previous simpler circulation which lacked the specific features whose development he wished to account for. In his case the simpler circulation was the one which he assumed would prevail in the absence of rotation. In many subsequent studies it has been a state of rest.

Hadley has been criticized for disregarding the north-south component of the Coriolis force altogether, and it is unlikely that he was aware of its existence. Consideration of this force would have been useless, in any case, in an argument making no reference to pressure. As a consequence he apparently supposed that the vertical and meridional (north-south) motion would not change during the development of the zonal (east-west) motion, and his task of describing the development was relatively simple. In reality, as soon as zonal motion has been produced by the deflection of the initial meridional motion, additional meridional motion will be produced by the deflection of the zonal motion, whereupon additional zonal motion will be produced by the deflection of the additional meridional motion while additional north-south pressure gradients will concurrently be produced by the convergence and divergence of the additional meridional motion. Both the additional pressure gradients and the deflection of the additional zonal motion will produce further additional meridional motion, etc., and it is reasonable to conclude that Hadley would have had a difficult time in carrying his argument to completion. Indeed, it is difficult to see how any argument of this sort, involving two or more processes whose effects may alternately combine and cancel, and requiring more than two or three steps, can be carried to a successful conclusion unless it is made quantitative, so that the accumulated changes of each quantity can be properly recorded. In this event the argument is converted into a stepwise numerical integration. Recently such integrations have been widely used with excellent results, but they often require hundreds of steps for completion.

A modern theoretician attempting to reproduce Hadley's description of the development of the trade winds in a rigorous quantitative fashion would in fact find that many years would be needed for the circulation to become nearly steady if he represented the effect of friction through a coefficient of molecular viscosity. To achieve a steady circulation within a few days he would be forced to introduce the much larger coefficient of turbulent viscosity. Use of this coefficient can be justified only in combination with a further idealization.

It is utterly impracticable to describe every gust of wind or even every cumulus cloud occurring at a particular time, even if the description is to appear only in the memory of the largest existing digital computer. It is therefore customary in problems of global scale to define the circulation as a smoothed circulation, from which motion systems of thunderstorm size or less have been subtracted. Meanwhile the effects of these systems cannot be disregarded. Ordinarily it is postulated that the statistical properties of the small-scale motions can be described in terms of the smoothed circulation, although really suitable formulae which accomplish this have yet to be established. The simplest way to represent these properties is through the use of coefficients of turbulent viscosity and conductivity, which may exceed the corresponding molecular coefficients by a factor of 10^5 or more. Qualitatively, this idealization treats the atmosphere as a highly viscous, highly thermally conductive fluid.

Evidently Hadley unknowingly used this idealization in his argument, since he assumed that the trade winds would be reduced to their observed velocities within a few days. It is interesting to speculate as to whether, in an atmosphere with very high molecular viscosity and conductivity but otherwise like the Earth's atmosphere, the troublesome small-scale motions would actually fail to develop. If this is the case, the present idealization, like the ones previously described, replaces the Earth's atmosphere by a physically conceivable system.

In any event, in a comprehensive study of what is known about the global atmospheric circulation, it is necessary to recognize both the real and the idealized atmospheres. The idealized atmosphere has formed the subject of the great majority of theoretical studies. The observations needed to confirm the results of these studies have of necessity been restricted mainly to the real atmosphere. Since the two atmospheres are not the same, certain discrepancies between theory and observation are inevitable.

It is remarkable that a few changes in wording, entailing, however, a considerable change in approach, would have eliminated all the shortcomings of Hadley's work thus far mentioned. Hadley sought a steady-state circulation, independent of longitude. In such a circulation there must be at least one latitude, separating low-level easterlies from low-level westerlies, where the flow is directly toward the Equator. If Hadley had referred to a particular parcel of air crossing this latitude at some initial time, instead of referring to an initial circulation where all the air flowed directly equatorward or poleward, his ensuing sentences would have formed a qualitatively acceptable account of the nature and maintenance of the steady circulation which he envisioned.

Hadley's only fault which cannot be remedied by a slight rewording of his arguments is less obvious, and it lies in his original assumption about the vertical motions. It can be shown that in a thermally forced system the temperature and the upward motion are positively correlated, but the correlation need not be perfect nor even very high. Hadley assumed in essence that all of the air would rise in low latitudes and sink in high latitudes. From this point on, barring further errors in reasoning, he was forced to obtain the picture of the circulation which he did. Observations which were unavailable in the eighteenth century but have since become superabundant reveal that this picture is incorrect. Yet it is within the realm of possibility that there somewhere exists a planet whose circulation conforms by and large to Hadley's picture. Such a circulation, whether real or hypothetical, is now known as a Hadley circulation.

If such a planet exists, Hadley's work, with the indicated changes in wording, is not only a description of the circulation there but also an essentially correct account of the basic reasons why this circulation occurs. Yet it is in no way a demonstration that the envisioned circulation must take place in preference to some other one. It lacks quantitative considerations, and on a qualitatively similar but quantitatively different planet there are alternative possibilities, one of the more obvious being the type of circulation which actually occurs on the Earth. Stated otherwise, Hadley's work lacks mathematical rigour. For this reason, we cannot look upon it as a full explanation.

A demand for mathematical rigour is not a demand for mathematical symbols and formulae. It is perfectly possible for a purely verbal argument to be mathematically rigorous. But, particularly when the argument is very complicated, a non-rigorous qualitative approach offers numerous opportunities for errors in reasoning. One of the best ways to avoid such errors is to formulate the problem in mathematical symbols, and manipulate these symbols according to established procedures.

What, then, constitutes a full or complete explanation ? This depends upon whether the question being answered is qualitative or quantitative.

Consider, for example, the problem of explaining why the average surface wind at latitude 20°N is directed from 15°N of E at 5 metres per second (or whatever the exact direction and speed may be). The wind is influenced by the field of pressure, which in turn is influenced by the field of temperature. Certainly then the precise wind velocity depends upon the precise amount of energy received from the sun, and upon the precise values of the physical constants which characterize the Earth and its atmosphere. Water vapour and liquid water ultimately affect the wind velocity by altering the thermodynamic properties of the atmosphere, and perhaps to an even greater extent by influencing the fields of incoming and outgoing radiation. The distribution of water in the atmosphere is in turn affected by the locations of oceans and continents, and of course by the field of motion itself. If all the relevant physical factors could be properly incorporated into the governing equations, and if the equations could be solved in a rigorous fashion, the proper numerical values would be found. The observed wind velocities would then seem to be completely explained, whereas no simpler procedure could be expected to give the correct result.

A correct answer to the quantitative question of why the wind blows from 15°N of E at 5 metres per second is of necessity an answer to the qualitative question of why the wind blows from a general easterly direction, but it may not be a very satisfactory answer. It may not indicate which of the many physical factors involved are needed to bring about the easterly wind, and which are mere modifying influences. In short, it may fail to answer the more general question as to why planetary atmospheres sharing certain features with the Earth's atmosphere possess easterly surface winds at low latitudes.

This difficulty need not arise if an analytic expression for the wind velocity in terms of the various physical constants has been found, but analytic solutions of meteorological equations are rather rare. If the solution has been obtained by numerical means, it would have to be repeated many times, with different values of the constants, in order to apply to the general case. At best this would be an extremely roundabout way of obtaining a desired answer which is not quantitative at all.

Moreover, even if the irrelevant physical factors are all eliminated, and a rigorous solution of the resulting simplified equations is obtained, the reader who has followed the demonstration from beginning to end may still gain little physical insight as to why easterly winds must exist, particularly if the demonstration is complicated or lengthy, or depends upon mathematical theorems whose proofs he does not recall or understand. Whereas a lack of rigour may lead to incorrect results, rigour alone does not guarantee understanding. An argument of the type presented by Hadley, if correct, may well prove more satisfying. Thus an acceptable answer to a qualitative question may well be more difficult to produce than an acceptable answer to a quantitative one.

Both quantitative and qualitative questions concerning the global circulation frequently arise. The most complete answer to the problem should therefore consist of a rigorous quantitative solution of the governing equations, yielding the observed circulation, together with a qualitative and possibly verbal explanation of the basic reasons why the principal qualitative features occur. In this event the qualitative explanation need not be rigorous, but it should be correct, and it must certainly be consistent in every respect with the quantitative solution which it accompanies.

From what has been said it appears that the motion of the atmosphere cannot be explained without full consideration of the accompanying fields of pressure, temperature, and moisture, and that these fields in turn cannot be explained independently of the field of motion. Such a statement cannot be made for all fluid systems. The future motion of a homogeneous incompressible fluid, for example, is completely determined by the present field of motion together with the external mechanical forces, and the circulation of such a system may be regarded as synonymous with the field of motion. In the case of the atmosphere it is more logical and certainly more convenient to regard the circulation as consisting of the field of motion together with the accompanying fields of the remaining meteorological variables.

The question naturally arises as to why no complete explanation of the global circulation has yet been produced. As already noted, the laws governing the real atmosphere are very complex, and are not perfectly known. We shall attempt to show now why the circulation of even the idealized atmosphere has yet to be fully explained.

The equations governing the idealized atmosphere appear to possess a steady-state solution which is also independent of longitude; this solution describes the Hadley circulation. If all other particular solutions could be shown to converge toward this solution, the problem of determining the circulation would be simply the problem of finding this solution. The determination of steady-state solutions of various systems of equations is one of the more frequently encountered problems in fluid dynamics.

When the general solution does not approach the Hadley solution asymptotically, the equations are likely to possess periodic solutions. Again, if all particular solutions, excluding those exceptional ones which converge toward the Hadley solution, could be shown to converge toward the periodic solutions, the problem of determining the circulation would reduce to the problem of finding these solutions.

Observations reveal, however, that the behaviour of the real atmosphere is neither steady nor periodic. Theoretical studies imply that the idealized atmosphere is likewise non-periodic; indeed, if the atmosphere has been idealized to the extent that it is forced to behave periodically, it has probably been over-idealized for the present purposes. Except in special instances it is not possible to express a complete non-periodic solution, even approximately, with a finite number of symbols, and the goal of determining the complete life-history of the idealized atmosphere must be abandoned.

This state of affairs is brought about by the non-linearity of the equations. Among the non-linear terms are those representing advection — the displacement of the field of motion, temperature, water vapour, or some other quantity, by means of the field of motion itself. In a sufficiently idealized atmosphere with crudely represented heating and friction, advection is the only non-linear process. Since the motion which brings about the displacement is generally not uniform, different portions of the field of each displaced variable undergo different displacements, and the field as a whole is distorted as well as displaced. Under continual distortion it may soon acquire a shape bearing little resemblance to its earlier configurations, and possessing much fine detail. With such an infinite variety of shapes there is no need for a pattern ever to repeat itself in all its features simultaneously, and the circulation need not vary periodically.

Yet non-linearity does not assure non-periodicity. The number of possible circulation patterns, none of which bears any resemblance to any of the others, is limited, and ultimately a pattern must occur which resembles a previous pattern rather closely, particularly in its coarser features. If the further evolution of the pattern is stable, in the sense that small differences between separate solutions of the equations will not amplify, the previous history will tend to repeat itself and the pattern will continue to recur at regular intervals, at least in an idealized atmosphere where the external conditions are steady. If, instead, the behaviour is unstable, approximate repetitions of previous history will ordinarily be only temporary, and periodicity need not develop.

Since it is not feasible to determine the complete history of the circulation theoretically, we must turn our attention to slightly less ambitious problems. One of these is the problem of explaining each pattern in a long but finite succession of circulation patterns; this in essence is the problem of long-range forecasting. A different problem, and the one with which this monograph is concerned, is that of explaining the characteristic properties, or statistics, of the collection of all circulation patterns which ever occur.

The equations governing the circulation are most readily written in a form expressing the time-derivative of each atmospheric variable — velocity, temperature, water-vapour content, etc. — in terms of the current values of the same set of variables. They do not directly account for any particular circulation pattern, except in terms of some other pattern which has just occurred or is just about to occur. It is as though the laws had been created for the convenience of the weather forecaster.

But the problem of determining long-term statistics is not the problem of weather forecasting, even though the governing equations may be the same. The latter is strictly an initial-value problem; the former does not *a priori* involve any initial values, even though initial-value procedures sometimes offer the only tractable means of solution. Whereas the latter is strictly a problem in differential equations, the former is a problem in ergodic theory, which is concerned with long-term statistical properties of solutions of equations.

The results of ergodic theory do not assure us even of the existence of long-term statistics, since there are systems of equations for which the average values of particular solutions over long intervals do not converge to any limit as the period of averaging becomes infinite. Assuming that the atmospheric equations are not of this peculiar and possibly exceptional type, each particular solution possesses its own long-term statistics, but there is no assurance that different solutions possess the same statistics. For a large class of systems of equations, however, there is only one set of statistics which a randomly chosen particular solution has a greater-than-zero probability of possessing. Such systems are called *transitive*. A transitive system may possess in addition any number of particular solutions having different sets of statistics, but the probability that a randomly chosen solution possesses one of these sets of statistics is zero (in the same sense that the probability is zero that a number chosen at random from the set of real numbers between zero and one will be a rational fraction). For example, in an atmosphere whose general solution is unsteady, the probability of choosing at random a solution which asymptotically approaches the Hadley solution is zero. If two or more sets of statistics have greater-than-zero probabilities of being chosen at random, the system is called *intransitive*.

Ergodic theory has not yet provided us with a general rule for determining whether a given system is transitive or intransitive. We therefore do not know whether the atmosphere is capable of possessing more than one set of statistics. Lest it appear implausible that the atmosphere could actually behave in an essentially different manner from what is observed, let us note that certain laboratory systems designed to simulate the atmosphere have proven to be intransitive. Unfortunately for our understanding of the atmosphere, but perhaps fortunately for the continuation of the human race, we cannot halt the atmospheric circulation and then see whether it will redevelop in a different manner.

Assuming that the atmosphere is transitive, we must then decide which statistics ought to be determined. There is no hard-and-fast rule, but the long-term time-averaged circulation, or more specifically the limiting form of the time-averaged circulation as the period of averaging approaches infinity, might be regarded as a minimum requirement. Undoubtedly this average circulation has received the most theoretical attention in recent years.

Yet time averages *per se* are not necessarily the statistics of greatest interest. Perhaps the average circulation is of more interest as a first approximation to the particular circulation to be expected at any given time. The trade winds, for example, are so persistent that an explanation of the time-averaged trades might be considered tantamount to an explanation of the time-variable trades. The upper-level westerly flow in middle latitudes, while less persistent, is still far more than a mere statistical residual.

Other regularly occurring features are poorly represented by time averages of the meteorological variables. Nothing indicates the frequency or even the presence of migratory cyclones and anticyclones. The jet stream appears only in attenuated form, and most of its familiar meanders are lacking.

All of these features are indicated by suitably chosen statistics, and hence by the collection of all long-term statistics. This collection includes such quantities as joint probability distributions, and it is of course impossible in practice to explain all of these, simply because an infinite amount of labour would be needed. Conceivably it might be possible to explain any particular statistic. Yet not even the long-term time-averaged circulation has thus far been fully explained.

The underlying difficulty is again the non-linearity. By rendering the general solution of the equations non-periodic, non-linearity makes it impossible to solve the equations by analytic methods and then obtain statistics by integrating with respect to time.

The most feasible method of solving non-linear equations with non-periodic solutions is as an initial-value problem by numerical means. This method yields finite segments of particular solutions. Statistics may be easily evaluated from these.

Such segments possess one of the principal disadvantages which characterize real meteorological data; they are finite samples from a population, and are not necessarily representative. The best method of assuring reasonably representative results is to extend the solution over a long time-interval, but this method may entail a prohibitive amount of computation.

More than any other theoretical procedure, numerical integration is also subject to the criticism that it yields little insight into the problem. The computed numbers are not only processed like data but they look like data, and a study of them may be no more enlightening than a study of real meteorological observations.

An alternative procedure which does not suffer this disadvantage consists of deriving a new system of equations whose unknowns are the statistics themselves. This procedure can be very effective for problems where the original equations are linear, but, in the case of non-linear equations, the new system will inevitably contain more unknowns than equations, and can therefore not be solved, unless additional postulates are introduced.

Moreover, even if the new system of equations could be solved, it would not necessarily yield the desired result. The separate solutions of the new system would include the statistics of all solutions of the original system. The statistics of the Hadley solution could perhaps be recognized as such and eliminated, but there would remain the statistics of an infinity of periodic and otherwise special solutions.

The separate solutions of a system of equations whose unknowns are statistics will therefore show nearly as wide a variety as the statistics evaluated from separate finite segments of solutions of the original equations. For example, there are presumably special periodic solutions representing circulations which are permanently of the "high-index" or "low-index" type, with well developed or poorly developed middle-latitude westerlies; there are presumably a great many more special periodic solutions which oscillate between high-index and low-index régimes, but do not divide their time between the régimes in the same proportion as does the general solution. The statistics of these special solutions are included among the solutions of the new system of equations.

In short, the only presently feasible procedure for determining quantitative statistics consists of evaluating them directly from particular time-dependent solutions of the original equations, and the only known procedures for solving these equations are numerical. Even these procedures are feasible only because of high-speed computing machines. Ultimately with the development of much larger and faster computers it may become possible to estimate the statistics of the general solution with a high degree of precision, even for the real atmosphere, although the proper representation of the effects of small-scale systems may prove to be a stumbling block. At present the procedure is limited to a rather idealized atmosphere. Moreover this procedure, being numerical, is of the type which contributes least to a qualitative understanding of the circulation.

There remains the possibility of rigorous procedures which are not quantitative at all. Any qualitative statement about the circulation may be formulated as a mathematical inequality; for example, the statement that the trade winds blow from a general easterly direction is equivalent to the statement that the eastward wind component in these latitudes is less than zero. There is no difficulty in deriving various incomplete systems — systems with more unknowns than equations — whose unknowns are statistics. Sometimes enough inequalities connecting the statistics may be established to complete the system. In this event it may be possible to solve the system of equations and inequalities for upper and lower bounds of the statistics, and thereby obtain qualitative descriptions of certain features of the circulation.

Possibly the relevant systems derivable in this manner are intractable. We feel, however, that the current failure to have obtained a qualitative explanation through this procedure must be attributed mainly to failure to have exploited the procedure.

If the causes of the circulation have not been fully explained, what can be the nature of the thousand or more excellent studies previously alluded to ? Some of these have dealt principally with observations, thereby providing a better picture of the phenomena to be accounted for. Some have sought to reproduce the circulation or some of its features by means of laboratory models, or with the aid of electronic digital computers, thereby making it possible to perform controlled experiments. Some have aimed to establish relationships between various features of the circulation by analytical means. Some have presented comprehensive assessments of the current state of progress. In the following chapters we shall examine some of these studies, and attempt to identify the contributions which they have made to our present understanding of the problem.

CHAPTER II

THE DYNAMIC EQUATIONS

Before one can make any serious attempt to explain the circulation of the atmosphere, he must become familiar with the circulation which he wishes to explain, and with the physical laws which govern it. One might argue that familiarity with the physical laws should be sufficient; from these one should be able to deduce all the properties of the circulation. Certainly there are physical systems whose behaviour can be inferred from the relevant laws, particularly when the non-linear terms in the equations representing these laws are of secondary importance. Yet experience suggests that the investigator who attempts to deduce the atmospheric circulation without first observing it is placing himself at a considerable disadvantage; to date we have not even accomplished the supposedly simpler task of explaining the circulation after observing it.

Indeed, we are continually encountering new features whose existence we had not anticipated from years of familiarity with the governing laws. One of the more spectacular of these is the recently discovered 26-month or quasi-biennial oscillation, whose outstanding feature is the appearance of persistent easterly and westerly winds in alternate years, in low latitudes in the stratosphere. There now exists an extensive literature on the subject (see Reed, 1965), but we are still awaiting a satisfactory explanation, which is not surprising when we recall that even the trade winds and the prevailing westerlies at sea-level are not completely explained.

The problem of formulating usable equations cannot be completely separated from that of observing the atmosphere. In any nearly exact form the equations cannot be satisfactorily solved by any known procedure. Certain approximations must be introduced. The possible approximations are so numerous that a suitable choice among them can be anticipated only if it is guided by observations. Lack of familiarity with the atmosphere has led to such incongruities as attempts to study the circulation with the equations for irrotational flow.

In this chapter we shall first present the system of governing equations in a fairly precise form. We shall then introduce some of the more frequently used approximations. The approximate systems have formed the basis for most of the attempts to account for the circulation in recent years. In the following chapter we shall describe the circulation as it has been observed. Necessarily, however, some of the observed properties of the circulation must be introduced in this chapter, while some of the equations must be examined in the next. The reader who is already familiar with the dynamic equations or who wishes to pass over the mathematical formalism may prefer to proceed immediately to the next chapter at this point.

The exact equations

It is convenient to group the laws governing the atmosphere into two categories. First there are the basic hydrodynamic and thermodynamic laws which apply to all or a large class of fluid systems. These include the law of conservation of mass, Newton's second law of motion, and the first law of thermodynamics, which state that matter can neither be created nor destroyed, momentum can be altered only by a force, and internal energy can be altered only by the performance of work or the addition or removal of heat. The ideal gas law also belongs in this category, although it is less general than the other laws. At great depths in Jupiter's atmosphere, for example, where the density may be comparable to that of a liquid, the ideal gas law is presumably not valid, while at extreme heights in our own atmosphere it is also inapplicable.

The remaining laws are the ones needed to express the forces and the heating in terms of the current state of the atmosphere and its environment. This category includes the laws governing the absorption, reflection, and scattering of solar radiation, and the absorption, emission, and transfer of infra-red radiation, by the various atmospheric constituents, notably carbon dioxide, ozone, and the various phases of water. It includes the laws of turbulent viscosity and conductivity, i.e. the laws governing the transfer of momentum and sensible heat by turbulent eddies. In principle these laws could perhaps be derived from the basic laws of hydrodynamics and thermodynamics, but no one has yet succeeded in accomplishing this task. Finally, it includes the laws governing the evaporation and condensation of water, and the conversion of cloud droplets into raindrops and snow crystals. The list is by no means exhaustive.

The equations representing the basic laws may be written in vector form, in terms of the independent variables

t: time,
$\mathbf{r}$: position, with respect to Earth's centre;

the dependent variables

$\mathbf{V}$: velocity, relative to rotating Earth,
p: pressure,
α: specific volume,
T: temperature;

the vectors characterizing the Earth

$\mathbf{\Omega}$: Earth's angular velocity,
$\mathbf{g}$: apparent acceleration of Earth's gravity;

the physical constants characterizing the atmosphere

c_v: specific heat of air at constant volume,
c_p: specific heat of air at constant pressure,
R: $c_p - c_v$, gas constant for air,
γ: c_p/c_v, approximately 7/5;

and the friction and heating

$\mathbf{F}$: frictional force per unit mass,
Q: net heating per unit mass.

A complete alphabetical list of symbols used in this work appears as an appendix. The symbols are for the most part the standard or most frequently used ones in current meteorological practice. In some instances it has been necessary to choose among several commonly used symbols, while a few less familiar symbols have been introduced to avoid using the same symbol for two quantities. We prefer the symbol Q for the rate of heating to the expression dQ/dt sometimes used in thermodynamics, since the latter expression tends to imply that there is some quantity "heat" whose time-derivative is the rate of heating.

The basic hydrodynamic and thermodynamic laws may be represented, with some redundancy, by the equations

$$d\mathbf{V}/dt = -2\mathbf{\Omega} \times \mathbf{V} - \alpha\nabla p + \mathbf{g} + \mathbf{F}, \tag{1}$$

$$d\alpha/dt = \alpha\nabla\cdot\mathbf{V}, \tag{2}$$

$$dT/dt = -(\gamma - 1)T\nabla\cdot\mathbf{V} + Q/c_v, \tag{3}$$

$$dp/dt = -\gamma p\nabla\cdot\mathbf{V} + (\gamma - 1)Q/\alpha, \tag{4}$$

$$p\alpha = RT, \tag{5}$$

or by other equations exactly equivalent to these. The time-derivatives in equations (1)-(4) are individual time-derivatives, referring to the rate of change at a point which moves with the flow.

The equation of motion (1) and the equation of continuity (2) represent Newton's second law of motion and the law of conservation of mass. As written they apply equally well to a gas or a liquid. The equation of motion is written for a frame of reference which rotates with angular velocity $\mathbf{\Omega}$. The true acceleration differs from the apparent acceleration $d\mathbf{V}/dt$ by the Coriolis acceleration $2\mathbf{\Omega}\times\mathbf{V}$ and the centripetal acceleration $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r})$. The rotation of the system is therefore fully taken into account by introducing the "Coriolis force" $-2\mathbf{\Omega}\times\mathbf{V}$, and "apparent gravity" $\mathbf{g}$ which differs from true gravity by $-\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r})$, and otherwise regarding the system as if it were not rotating. Once this has been accomplished, it is permissible for most purposes to treat the Earth (except for topographic features) as a sphere instead of an ellipsoid, with a gravitational force of constant magnitude directed toward the centre, since, within the lowest 25 kilometres of the atmosphere, the maximum angle between $-\mathbf{g}$ and $\mathbf{r}$ is only 0.2 degrees, while the magnitude of $\mathbf{g}$ varies by only slightly more than one per cent.

The thermodynamic equation (3) represents the first law of thermodynamics, while (5) is the equation of state. As written, they apply to an ideal gas. Certain modifications are needed to make them apply to an atmosphere where water can appear in different phases or in varying amounts. In formulating equation (3) we have noted that the internal energy per unit mass is $c_v T$, and we have used the customary assumption that the work done upon a unit mass in compressing it is given by $-p\, d\alpha/dt$; we shall presently consider the implications of this assumption. We have then used (2) and (5) to express the work as $-RT\nabla\cdot\mathbf{V}$, after which (3) follows. Equation (4) may be derived from (2) and (3) with the aid of (5). It is often more convenient to use the density ρ as a dependent variable in place of its reciprocal α.

Equations (1)-(4) are prognostic, i.e. they express the time-derivatives of the dependent variables in terms of the current values of these variables. Equation (5) is diagnostic, i.e. it contains no time-derivatives. The diagnostic equation may be used to eliminate any two of the three variables α, T, p from the system of equations (1)-(4). In each case the system then contains one vector and two scalar prognostic equations, or equivalently five scalar prognostic equations, with the same number of dependent variables. It is therefore a closed system, i.e. it is sufficient to determine the future values of the dependent variables in terms of the present, provided that the net frictional force $\mathbf{F}$ and the net heating Q are regarded as known functions of the independent and dependent variables.

For practical reasons it is often desirable to express the equation of motion (1) in scalar form. The Earth is sufficiently spherical in shape to justify the use of a spherical co-ordinate system. The equations may then be written in terms of the additional independent variables

λ: longitude, measured eastward,
φ: latitude, measured northward,
z: elevation, measured upward,
r: magnitude of $\mathbf{r}$, distance from Earth's centre ;

the dependent variables

u: $r\cos\varphi\, d\lambda/dt$, eastward component of $\mathbf{V}$,
v: $r\, d\varphi/dt$, northward component of $\mathbf{V}$,
w: dz/dt, upward component of $\mathbf{V}$;

and the constants

a: Earth's mean radius,
Ω: magnitude of $\mathbf{\Omega}$,
g: mean magnitude of $\mathbf{g}$.

The velocity components u, v, w are the scalar products of $\mathbf{V}$ with the unit vectors

$\mathbf{i}$: $(\mathbf{\Omega}\times\mathbf{r})/|\mathbf{\Omega}\times\mathbf{r}|$,
$\mathbf{j}$: $\mathbf{k}\times\mathbf{i}$,
$\mathbf{k}$: $\mathbf{r}/r$.

Because of the curvature of the spherical co-ordinate system, the components of the acceleration $d\mathbf{V}/dt$ are not the time-derivatives of the components of $\mathbf{V}$. Additional terms involving the time-derivatives of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ occur. Thus the equations of motion become

$$\frac{du}{dt} = \frac{\tan\varphi}{r}uv - \frac{1}{r}uw + 2\Omega\sin\varphi\, v - 2\Omega\cos\varphi\, w - \frac{\alpha}{r\cos\varphi}\frac{\partial p}{\partial\lambda} + F_\lambda, \tag{6}$$

$$\frac{dv}{dt} = -\frac{\tan\varphi}{r}u^2 - \frac{1}{r}vw - 2\Omega\sin\varphi\, u - \frac{\alpha}{r}\frac{\partial p}{\partial\varphi} + F_\varphi, \tag{7}$$

$$\frac{dw}{dt} = \frac{1}{r}u^2 + \frac{1}{r}v^2 + 2\Omega\cos\varphi\, u - g - \alpha\frac{\partial p}{\partial z} + F_z, \tag{8}$$

where F_λ, F_φ, F_z are the components of $\mathbf{F}$.

The individual and the local time-derivatives of an arbitrary scalar quantity X are related by the formula

$$dX/dt = \partial X/\partial t + \mathbf{V}\cdot\nabla X, \tag{9}$$

whence, in view of the equation of continuity (2),

$$\rho dX/dt = \partial(\rho X)/\partial t + \nabla\cdot\rho X\mathbf{V}. \tag{10}$$

The latter form is especially convenient when the equations are to be integrated over a volume. In the curvilinear co-ordinate system,

$$\nabla\cdot\mathbf{V} = \frac{1}{r^2\cos\varphi}\left(\frac{\partial}{\partial\lambda}ru + \frac{\partial}{\partial\varphi}r\cos\varphi\, v + \frac{\partial}{\partial z}r^2\cos\varphi\, w\right), \tag{11}$$

while an analogous expression holds for $\nabla\cdot\rho X\mathbf{V}$. An element of volume is given by $r^2\cos\varphi\, d\lambda\, d\varphi\, dz$.

It is often advantageous to write the equations in terms of the potential temperature

$$\theta = p_{00}{}^{\kappa}\, T/p^{\kappa} \tag{12}$$

or the related specific entropy (of an ideal gas)

$$s = c_p \ln\theta \tag{13}$$

where $\kappa = R/c_p$ is about 2/7, and the constant $p_{00} = 1000$ mb has been introduced to make θ and T dimensionally similar. It follows from (3) and (4) that

$$d\theta/dt = (p_{00}{}^{\kappa}/c_p)\, Q/p^{\kappa}, \tag{14}$$

$$ds/dt = Q/T. \tag{15}$$

Equation (15) reveals the nature of the thermodynamic assumptions which occur in the usual formulation of the governing equations.

According to (15), the entropy change equals the ratio of the heating to the temperature. It is a fundamental principle of thermodynamics that this is so during a reversible process, but not necessarily during an irreversible process. Yet (15) has been derived from (3), and hence ostensibly from the first law of thermodynamics, which holds equally well for reversible and irreversible processes.

Since the entropy of the atmosphere must increase during any irreversible process not involving the environment, we must somewhere have introduced the assumption that all processes of this sort involve some heating, and hence that none of these processes involves a performance of work alone. This we did in formulating (3), when we assumed that the work was always given by $- p d\alpha/dt$. In order to render (3) valid despite this assumption we must therefore, when the state of the atmosphere is altered by an irreversible process, define Q as the heating which would occur in a reversible process which would alter the state of the atmosphere in a similar manner.

One of the most important irreversible processes in the atmosphere is the mixing of different masses of air. For convenience we may distinguish between the mixing of masses of different temperature, i.e. turbulent conduction, and the mixing of masses of different velocity, i.e. turbulent friction. In the former process there must also be some difference in velocity to accomplish the mixing, but this may be assumed negligibly small.

The former process does not *per se* involve any net performance of work. There is also no net gain of internal energy, and hence no net heating, but (3) is valid provided that the originally colder air is assumed to be brought to its new temperature by heating, and the originally warmer air is assumed to undergo an equal amount of cooling. Exchange of energy by radiation may be treated similarly.

In the latter process the total kinetic energy decreases. Since the total energy is not altered, the internal energy increases by a similar amount. (We may for the sake of this discussion neglect the presence of gravity, so that potential energy need not be considered.) One might be tempted to assume that the increase in internal energy could result entirely from a performance of work, as given by the work term in (3). In that case there would be no heating. If (15) is accepted, there would then be no entropy change. The assumption would then lead to the absurdity that the mixing process is reversible.

It follows, then, that if (3) and (15) are to be retained, the system must be assumed to gain *by heating* as much internal energy as the kinetic energy which it loses. This so-called frictional heating must be included in Q in order that (3) may be valid.

Equations (1)-(11), together with suitable expressions for $\mathbf{F}$ and Q, are in principle sufficient for a mathematical study of the circulation. Qualitative arguments are nevertheless often more readily presented in terms of angular momentum and energy.

Per unit mass, the absolute angular momentum about the Earth's axis is given by the formula

$$M = \Omega r^2 \cos^2 \varphi + r \cos \varphi \, u. \tag{16}$$

The first term on the right-hand side of (16) represents the so-called Ω-momentum, the absolute angular momentum which would be present if the atmosphere were in solid rotation with the Earth. The second term is the relative angular momentum, associated with the motion relative to the Earth. The terms in (6) containing $1/r$, depending upon the curvature of the co-ordinate system, and the terms containing Ω, depending upon the rotation, drop out in the angular-momentum equation

$$dM/dt = - \alpha \, dp/d\lambda + r \cos \varphi \, F_{\lambda}, \tag{17}$$

which states that absolute angular momentum is altered only by a torque. An equivalent statement would be that relative angular momentum is altered only by a torque, provided that the Coriolis torque is included. Equation (17) could of course have been used to derive (6).

Likewise, per unit mass the kinetic energy, potential energy, and internal energy (of an ideal gas) are given by

$$K = \tfrac{1}{2}\,\mathbf{V}\cdot\mathbf{V}, \tag{18}$$

$$\Phi = gz, \tag{19}$$

$$I = c_v T. \tag{20}$$

The terms in (6)-(8) containing $1/r$ and Ω also drop out in the kinetic energy equation

$$dK/dt = -\alpha\mathbf{V}\cdot\nabla p + \mathbf{V}\cdot\mathbf{F}. \tag{21}$$

Since obviously

$$d\Phi/dt = gw\,, \tag{22}$$

while (3) may be written

$$dI/dt = -\alpha\, p\, \nabla\cdot\mathbf{V} + Q\,, \tag{23}$$

we obtain the equation of total energy

$$d(K + \Phi + I)/dt = -\alpha\,\nabla\cdot p\mathbf{V} + \mathbf{V}\cdot\mathbf{F} + Q\,. \tag{24}$$

When integrated over any region with a fixed boundary, the term $-\alpha\,\nabla\cdot p\mathbf{V}$ represents the work done on this region by the pressure force on the boundary; thus in general it describes a transfer of energy from one region to another.

The angular-momentum and energy principles are fundamental in any treatment of the circulation. If in some approximate formulation of the equations they are not retained, the results are likely to be unrealistic. A spurious energy source may, for example, cause the wind to increase without limit.

The usual mathematical formulation of friction and heating is much less precise than that of the processes which we have so far considered. Friction seems to act mainly to transfer horizontal momentum in the vertical direction, so that, to a good approximation,

$$\mathbf{F} = \alpha\,\partial\boldsymbol{\tau}/\partial z, \tag{25}$$

where $\boldsymbol{\tau}$ is a horizontally directed vector representing the drag of the air above a given level upon the air below. The drag is often expressed in terms of the vertical shear of the wind through a coefficient of turbulent viscosity μ; thus

$$\boldsymbol{\tau} = \mu\partial\mathbf{U}/\partial z, \tag{26}$$

where $\mathbf{U}$ denotes the horizontal velocity $u\mathbf{i} + v\mathbf{j}$, as distinguished from $\mathbf{V}$. The value of μ should preferably depend upon the intensity of the turbulence, but in an idealized atmosphere it is frequently taken to be a constant.

Likewise, in an idealized atmosphere Q may be taken as the difference between a function of latitude and height alone, representing incoming radiation, and a function of temperature alone, representing outgoing radiation. For the real atmosphere the many equations governing radiation, turbulence, phase changes of water, and other processes, are required. It is beyond the scope of this discussion to present all of the relevant equations. We shall, however, indicate the modifications of equations (1)-(5) required by the presence of water.

The hydrodynamic equations (1) and (2) appear to remain virtually unaltered. In the equation of state (5), the gas constant R must be replaced by the slightly greater variable gas "constant" appropriate to a mixture of air and water, or, alternatively, the temperature T may be replaced by the slightly higher virtual temperature

$$T_v = (1 - q)T + (R_w/R)qT, \tag{27}$$

where R_w is the gas constant for water and q is the specific humidity. Throughout much of the atmosphere T_v and T differ by less than a degree, but near the surface in the tropics the difference may exceed 4°C.

The more important effects of water vapour appear in the thermodynamic equation (3) and the derived equation (4). The internal energy must be replaced by the internal energy of moist air, given by

$$I = c_v(1 - q)T + (c - R_w)qT + Lq, \tag{28}$$

where c is the specific heat of water and L is the latent heat of condensation at temperature T. Alternatively, the release of latent heat, given approximately by $-L\, dq/dt$, may be included as part of the heating Q. In either event the specific humidity q must be included as an additional dependent variable.

A common simplification is the assumption that liquid water falls out immediately upon forming from condensation. In this case q may be considered to remain constant, except in ascending saturated air, where it retains its saturation value, and near the Earth, where it may increase as a result of turbulent diffusion. Thus

$$dq/dt = \begin{cases} -\alpha \partial E/\partial z & \text{if} \quad q < q_s \quad \text{or} \quad dq_s/dt \geqq 0, \\ dq_s/dt & \text{if} \quad q = q_s \quad \text{and} \quad dq_s/dt < 0, \end{cases} \tag{29}$$

where E is the upward turbulent transfer of water vapour per unit horizontal area, and $q_s(T, p)$ is the value of q which saturated air at temperature T and pressure p would possess. The limiting value E_0 of E as the surface of the Earth is approached is simply the rate of evaporation from the surface.

It would be more realistic to retain the liquid water content as another dependent variable, in which case q would retain its saturation value in descending air containing liquid water. If the solid water content is retained as still another variable, the possibility of supercooled water clouds in place of ice-crystal clouds must be recognized.

The hydrostatic equation and the primitive equations

Equations (1)-(5) are the so-called exact equations, although they evidently contain a number of approximations. In a sense they are too exact. Examination reveals that they possess certain properties which render them somewhat awkward for a study of the global circulation.

One of the most prominent features of the circulation is hydrostatic equilibrium — the approximate balance between gravity and the vertical pressure gradient force. The familiar hydrostatic equation

$$\partial p/\partial z = - g\rho \tag{30}$$

describing this equilibrium is obtained by equating the appropriate terms in the vertical equation of motion (8).

If systems of thunderstorm size or less are eliminated, the remaining circulation possesses vertical motions with typical speeds of a few centimetres per second. These motions may develop during the course of a day or less. Vertical accelerations of about 10^{-4} cm sec^{-2}, or about 10^{-7} times that of gravity, are therefore of considerable interest.

Reference to equation (8) for the vertical acceleration reveals the term $-g$. Since this term is almost exactly balanced by the term $-\alpha\partial p/\partial z$, accelerations comparable to that of gravity do not occur. It is evident however that rather minute disruptions of the field of pressure or density will upset the hydrostatic balance sufficiently to cause accelerations far in excess of 10^{-4} cm sec^{-2}. One may therefore ask why such vertical accelerations do not appear.

What happens is that these accelerations do occur temporarily, but the ensuing vertical motions alter the pressure and density fields in such a way as to reverse the sign of the acceleration a few minutes or even a few seconds later. What develops is therefore not a strong vertical current, but oscillations about some mean state. These oscillations are simply vertically travelling sound waves. They do not appear to have much significance for the global circulation, but their possible presence greatly complicates the mathematics.

It would be awkward to try to describe the effect of heating or some other disturbing influence by tracing the evolution of the atmosphere through each sound-wave oscillation, when one is interested only in the state about which the oscillations occur. It is more satisfactory to replace equation (8) by the hydrostatic equation (30). This equation almost exactly describes the mean state without describing the oscillations about it. In most theoretical studies of the circulation except those dealing specifically with motions of smaller scale, the system of governing equations has been modified by substituting (30) for (8).

Since the hydrostatic equation is diagnostic, its introduction leaves the new system with no prognostic equation for w. There are, however, two prognostic equations for p, namely the thermodynamic pressure-tendency equation (4) and the hydrostatic pressure-tendency equation

$$\partial p/\partial t = -g\int_z^\infty \mathbf{V}\cdot\rho\mathbf{V}\, dz, \tag{31}$$

obtained by integrating (30) with the upper boundary condition $p = 0$ at $z = \infty$. Elimination of dp/dt and $\partial p/\partial t$ from (4) and (31) yields an additional diagnostic equation, which may be solved for w in terms of the remaining variables, using the lower boundary condition $w = 0$ at $z = 0$. In effect the horizontal motions alone tend to alter the pressure and density fields in such a manner as to upset the existing hydrostatic equilibrium. The field of vertical motion is assumed to be that field required to maintain hydrostatic equilibrium by compensating for the effects of the horizontal motions. With w itself defined in terms of the other variables there is no need for an explicit expression for dw/dt, and with the aid of the diagnostic equations the system reduces to a closed system of three equations in the three dependent variables u, v, p.

There are certain objections to this system as it stands. It is desirable to retain the angular-momentum and energy principles. With equation (8) replaced by (30) the kinetic energy equation (21) no longer holds. It may be restored, however, provided first that kinetic energy is redefined to exclude the energy of the vertical motion, so that $K = \mathbf{U}\cdot\mathbf{U}/2$, and second that the terms in the horizontal equations of motions (6) and (7) containing w are discarded. These approximations seem to be as acceptable as the hydrostatic approximation, in view of the general smallness of w. However, the angular momentum equation (17) now no longer holds. It may also be restored by replacing r by the Earth's mean radius a

in the definition (16) of absolute angular momentum M, and in the equation of eastward motion (6). The energy principle is now upset again, but it may again be restored by replacing r by a in (7). In essence, in replacing r by a, the diverging of the Earth's radii as they extend upward from the surface is completely disregarded.

We present the new system of equations in two forms. The first form uses the co-ordinate system of equations (6)-(11). In the second form pressure p instead of elevation z is used as the vertical co-ordinate.

With z as the vertical co-ordinate, the new system may be written

$$d\mathbf{U}/dt = -f\mathbf{k}\times\mathbf{U} - (1/\rho)\nabla p + \mathbf{F}, \tag{32}$$

$$dp/dt = -\gamma p\nabla\cdot\mathbf{U} - \gamma p\,\partial w/\partial z + (\gamma - 1)\rho Q, \tag{33}$$

$$\gamma p\partial w/\partial z = -\gamma p\nabla\cdot\mathbf{U} - \mathbf{U}\cdot\nabla p + g\int_z^\infty \nabla\cdot\rho\mathbf{U}dz + (\gamma - 1)\rho Q, \tag{34}$$

$$\rho = -(1/g)\partial p/\partial z, \tag{35}$$

where $f = 2\Omega\sin\varphi$ is the Coriolis parameter. The hydrostatic pressure tendency equation (31) could have been used instead of (33). In this system it is to be understood that all vectors (except $\mathbf{k}$) are two-component horizontal vectors; ∇ is a horizontal differential operator. Wherever $1/r$ would ordinarily occur it is to be replaced by $1/a$; thus the components of the equation of motion become

$$\frac{du}{dt} = \frac{\tan\varphi}{a}uv + fv - \frac{1}{\rho}\frac{1}{a\cos\varphi}\frac{\partial p}{\partial\lambda} + F_\lambda, \tag{36}$$

$$\frac{dv}{dt} = -\frac{\tan\varphi}{a}u^2 - fu - \frac{1}{\rho}\frac{1}{a}\frac{\partial p}{\partial\varphi} + F_\varphi. \tag{37}$$

The individual and local time derivatives of a scalar X are connected by the relation

$$\rho dX/dt = \partial(\rho X)/\partial t + \nabla\cdot\rho X\mathbf{U} + \partial(\rho X w)/\partial z, \tag{38}$$

while the horizontal divergence is

$$\nabla\cdot\mathbf{U} = \frac{1}{a\cos\varphi}\left(\frac{\partial}{\partial\lambda}u + \frac{\partial}{\partial\varphi}v\cos\varphi\right), \tag{39}$$

with an analogous expression for $\nabla\cdot\rho X\mathbf{U}$. An element of volume is assumed to be $a^2\cos\varphi\,d\lambda\,d\varphi\,dz$.

For many purposes this new system is suitable. For other purposes it is far more convenient to introduce pressure p as a new vertical co-ordinate; thus p becomes an independent variable while z becomes a dependent variable, and $\omega = dp/dt$ replaces w as a further dependent variable. In this system the equation of continuity becomes the diagnostic equation (42), and the complete system may be written

$$d\mathbf{U}/dt = -f\mathbf{k}\times\mathbf{U} - g\nabla z + \mathbf{F}, \tag{40}$$

$$dT/dt = \kappa T\omega/p + Q/c_p, \tag{41}$$

$$\nabla\cdot\mathbf{U} + \partial\omega/\partial p = 0, \tag{42}$$

$$\partial z/\partial p = -RT/(gp). \tag{43}$$

It is equally possible to use α or θ instead of T as a dependent variable in the thermodynamic equation (41). The components of the equation of motion are

$$\frac{du}{dt} = \frac{\tan\varphi}{a} uv + fv - \frac{g}{a\cos\varphi}\frac{\partial z}{\partial\lambda} + F_\lambda, \tag{44}$$

$$\frac{dv}{dt} = -\frac{\tan\varphi}{a} u^2 - fu - \frac{g}{a}\frac{\partial z}{\partial\varphi} + F_\varphi. \tag{45}$$

The individual and local time derivatives are related by the equation

$$dX/dt = \partial X/\partial t + \mathbf{U}\cdot\nabla X + \omega\,\partial X/\partial p, \tag{46}$$

or with the aid of the equation of continuity (42)

$$dX/dt = \partial X/\partial t + \nabla\cdot X\mathbf{U} + \partial(X\omega)/\partial p, \tag{47}$$

where

$$\nabla\cdot\mathbf{U} = \frac{1}{a\cos\varphi}\left(\frac{\partial}{\partial\lambda}u + \frac{\partial}{\partial\varphi}v\cos\varphi\right), \tag{48}$$

and an analogous expression holds for $\nabla\cdot X\mathbf{U}$. It is understood that the partial derivatives $\partial/\partial t$, $\partial/\partial\lambda$, $\partial/\partial\varphi$ and ∇ are now to be interpreted as derivatives with p held constant, so that their meaning is not the same as in (32)-(35). Formally (48) is identical with (39), but the partial derivatives have their altered meaning. An element of mass is assumed to be $(1/g)a^2\cos\varphi\,d\lambda\,d\varphi\,dp$.

This considerably simpler system of equations is obtained only at the expense of a more complicated lower boundary condition. The condition $w = 0$ must now be written $dz/dt = 0$, while the lower boundary $p = p_0$ is no longer a co-ordinate surface.

For some purposes a satisfactory approximation is obtained by assuming as a lower boundary the co-ordinate surface $p = p_{00} =$ constant, with $\omega = 0$ as a lower boundary condition. The height of the lower boundary is then considered variable. In particular this approximation does not introduce spurious sources of angular momentum and energy. It has the effect of eliminating the so-called external gravity waves, whose propagation involves oscillations of the total mass within a vertical column.

Equations (32)-(35) or their equivalent forms (40)-(43) are the so-called primitive equations. This designation has arisen from their use in numerical weather prediction, where they have been taken as the starting point for the derivation of the simpler geostrophic model which we shall presently consider. Apparently it was thought improbable that anyone would attempt to use the exact equations, which are more primitive than the primitive equations.

Vorticity and divergence

For many purposes it is advantageous to express the horizontal wind field $\mathbf{U}$ in terms of its vorticity ζ and its divergence δ:

$$\zeta = \nabla\cdot\mathbf{U}\times\mathbf{k}, \tag{49}$$

$$\delta = \nabla\cdot\mathbf{U}. \tag{50}$$

Here ∇ will denote differentiation along an isobaric (i.e. constant-pressure) surface, although the slightly different vorticity and divergence fields defined by the same formulas, with ∇ denoting differentiation along a horizontal surface, have also been used. The vorticity might more properly be termed the component of the vorticity vector $\nabla\times\mathbf{V}$ normal to an isobaric surface.

If the stream function ψ and the velocity potential χ are defined by the equations

$$\nabla^2\psi = \zeta, \tag{51}$$

$$\nabla^2\chi = -\delta, \tag{52}$$

the rotational non-divergent wind field $\mathbf{U}_r$ and the divergent irrotational wind field $\mathbf{U}_d$ defined as

$$\mathbf{U}_r = \mathbf{k}\times\nabla\psi, \tag{53}$$

$$\mathbf{U}_d = -\nabla\chi, \tag{54}$$

satisfy the relation

$$\mathbf{U}_r + \mathbf{U}_d = \mathbf{U}. \tag{55}$$

If $\mathbf{U}$ and hence ζ and δ are defined over a complete spherical surface, ψ and χ (except for additive constants) and hence $\mathbf{U}_r$ and $\mathbf{U}_d$ are uniquely determined.

It should be observed that in a circulation which is symmetric with respect to the Earth's axis, such as Hadley's circulation, the zonal motion u is completely determined by $\mathbf{U}_r$, while the meridional motion v is completely determined by $\mathbf{U}_d$. In the more general case, the eastward and northward motion, averaged about a latitude circle, are determined respectively by $\mathbf{U}_r$ and $\mathbf{U}_d$.

A form of the equation of horizontal motion (40) which is exactly equivalent but more convenient for many purposes is

$$\partial\mathbf{U}/\partial t = -(\zeta + f)\mathbf{k}\times\mathbf{U} - \omega\,\partial\mathbf{U}/\partial p - \nabla(gz + \mathbf{U}\cdot\mathbf{U}/2) + \mathbf{F}. \tag{56}$$

From equation (56) one may easily derive the vorticity equation

$$\partial\zeta/\partial t = -\mathbf{U}\cdot\nabla(\zeta + f) - \omega\partial\zeta/\partial p - (\zeta + f)\delta - \nabla\omega\cdot\partial\mathbf{U}/\partial p\times\mathbf{k} + \nabla\cdot\mathbf{F}\times\mathbf{k}, \tag{57}$$

and the divergence equation

$$\partial\delta/\partial t = -\mathbf{U}\cdot\nabla(\zeta + f)\times\mathbf{k} - \omega\partial\delta/\partial p + (\zeta + f)\zeta - \nabla\omega\cdot\partial\mathbf{U}/\partial p - \nabla^2(gz + \mathbf{U}\cdot\mathbf{U}/2) + \nabla\cdot\mathbf{F}. \tag{58}$$

Equations (57) and (58) may appear at first glance to be more clumsy than the equations of motion (44) and (45). The advantages to be gained from using them stem from a combination of two circumstances.

First, it is a matter of observation that the vorticity ζ is ordinarily considerably larger than the divergence δ, except in low latitudes. Thus the rotational field $\mathbf{U}_r$ tends to be stronger than the divergent field $\mathbf{U}_d$, so much so that $\mathbf{U}_r$ affords a fair approximation to $\mathbf{U}$.

Second, the height z is completely absent in the vorticity equation (57). The equation therefore specifies the time-derivative of one feature of the wind field in terms of the wind field alone.

In dealing with certain features of the circulation, rather than the total circulation, one may neglect the weaker field $\mathbf{U}_d$ and hence δ and ω altogether. The vorticity equation by itself then becomes a closed system, provided that the friction $\mathbf{F}$ can be expressed in terms of $\mathbf{U}_r$. If $\mathbf{F}$ is also neglected, the vorticity equation reduces to

$$\partial\zeta/\partial t = -\nabla\psi\cdot\nabla(\zeta + f)\times\mathbf{k}, \tag{59}$$

or, equivalently,

$$d(\zeta + f)/dt = 0. \tag{60}$$

The sum $\zeta + f$ is the absolute vorticity, since the Coriolis parameter f equals the absolute vorticity which a fluid at rest with respect to the rotating Earth would possess. Equation (60) expresses the conservation of absolute vorticity, and is the equation used by Rossby (1939) in his famous study of the propagation of large-scale waves (now known as Rossby waves) in the upper-level westerly-wind belt.

Equation (59) contains no sources nor sinks for absolute vorticity. It strictly conserves the total kinetic energy, and also the total absolute angular momentum, at each level, and hence allows no conversion between kinetic and other forms of energy. It therefore cannot be used to explain the existing amounts of kinetic energy and absolute angular momentum, or the statistical distribution of absolute vorticity. Inclusion of friction would merely lead to a dissipation of all the kinetic energy, with an ultimate state of solid rotation. In dealing with the total circulation it is therefore necessary to retain the divergence. Substantial simplifications are nevertheless possible.

The geostrophic equation and the geostrophic model

Although the troublesome vertically travelling sound waves have been effectively filtered out of the primitive equations, there remain other modes of motion which are of questionable importance for the global circulation. These may also be eliminated by further approximations.

A feature of the circulation in middle and higher latitudes which is almost as prominent as hydrostatic equilibrium is geostrophic equilibrium — the approximate balance between the Coriolis force and the horizontal pressure gradient force. The familiar geostrophic equation

$$\mathbf{U} = (g/f)\mathbf{k} \times \nabla z \tag{61}$$

describing this balance is obtained by equating the appropriate terms in (40). The right hand side of (61) is often regarded as a definition of the geostrophic wind $\mathbf{U}_g$.

Just as temporary departures from hydrostatic equilibrium lead to oscillations about a mean state with periods of minutes or less, departures from geostrophic equilibrium lead to oscillations with periods of several hours or less. These oscillations are gravity waves, of which the previously mentioned external gravity waves are a special type. Like the vertically travelling sound waves, they are often assumed to have little significance for the global circulation, although it is less certain that this assumption is valid.

In any event it is inconvenient to trace the development of the circulation through each gravity-wave oscillation, and the substitution of the geostrophic equation for the equation of motion suggests itself. It would be possible to replace the eastward equation of motion (44) by the northward component of (61), or the northward equation of motion (45) by the eastward component of (61), and in either case obtain a closed system containing one prognostic equation, but this procedure does not appear particularly appropriate. In the former case K would have to be defined as $v^2/2$, and in the latter case as $u^2/2$, in order to preserve the energy principle. Since both horizontal components of the wind contain an important fraction of the total kinetic energy, it is to be expected that neither procedure would lead to realistic results. It would also be possible to replace both components of the equation of motion (40) by (61), and retain the thermodynamic equation as a single prognostic equation, but in that case the effects of vertical motion on the temperature field would not appear to be very well represented.

However, since the wind $\mathbf{U}$ is expressible as the sum of $\mathbf{U}_r$ and a smaller residual $\mathbf{U}_d$, and also as the sum of $\mathbf{U}_g$ and a smaller residual $\mathbf{U} - \mathbf{U}_g$, it follows that $\mathbf{U}_r$ is the sum of $\mathbf{U}_g$ and a reasonably small residual $(\mathbf{U} - \mathbf{U}_g) - \mathbf{U}_d$. The geostrophic vorticity $\nabla \cdot \mathbf{U}_g \times \mathbf{k}$ is generally a fair approximation to the vorticity ζ, although it is a considerable overestimate in intense cyclones. The geostrophic divergence $\nabla \cdot \mathbf{U}_g$, on the other hand, is always positive in poleward flow and negative in equatorward flow, and bears little resemblance to the divergence δ as observed in the atmosphere.

Just as the vertical equation of motion (8) may be replaced by the hydrostatic equation (30) obtained by retaining the most significant terms in (8), so the divergence equation (58) may be replaced by a variant of the geostrophic equation

$$\nabla \cdot (f \nabla \psi) = g \nabla^2 z\,, \tag{62}$$

obtained by retaining the linear terms in (58) not involving $\mathbf{U}_d$. Just as the substitution of (30) for (8) reduces the number of prognostic equations from five to three, so the substitution of (62) for (58) effectively reduces the number from three to one. The system no longer contains a prognostic equation for δ, but an additional diagnostic equation may be obtained by differentiating (62) with respect to p to obtain the relation

$$\nabla \cdot (f \nabla\, \partial\psi/\partial p) = -(R/p)\nabla^2 T, \tag{63}$$

and then differentiating (63) with respect to t and substituting from the vorticity equation (57) and the thermodynamic equation (41). The new equation, the so-called ω-equation, may in principle be solved for ω (or δ or χ) in terms of the remaining variables.

In effect, the rotational non-divergent motions alone tend to alter the wind and temperature fields in such a way as to upset the existing geostrophic equilibrium. The divergent irrotational wind and its accompanying field of ω are assumed to be those fields needed to maintain geostrophic equilibrium by compensating for the effects of the rotational wind.

Further modifications are now needed to retain the energy principle. The kinetic energy must be redefined as $K = \mathbf{U}_r \cdot \mathbf{U}_r/2$, and all the quadratic terms in the vorticity equation except those involving $\mathbf{U}_r$ only must be discarded. The vorticity equation and the thermodynamic equation then assume the form

$$\partial\zeta/\partial t = -\nabla\psi \cdot \nabla(\zeta + f) \times \mathbf{k} + \nabla \cdot (f \nabla \chi) + \nabla \cdot \mathbf{F} \times \mathbf{k}\,, \tag{64}$$

$$\partial T/\partial t = -\nabla\psi \cdot \nabla T \times \mathbf{k} + \nabla T \cdot \nabla \chi + \sigma\omega + Q/c_p\,, \tag{65}$$

where

$$\sigma = -(\partial T/\partial p - \kappa\, T/p)\,. \tag{66}$$

Together with (63) and the ω-equation, either (64) or (65) forms a closed system if suitable boundary conditions are given.

The atmosphere is said to be statically stable or unstable according to whether θ increases or decreases with elevation. From the definition of θ and the hydrostatic equation it follows that

$$\sigma = -(p/p_{00})^{\kappa} \partial\theta/\partial p = -(1/c_p)\, \partial(c_p T + gz)/\partial p\,, \tag{67}$$

so that σ is a measure of the static stability. Throughout most of the atmosphere σ is positive. Static instability favours the development of small-scale convective motions, which ordinarily act to stabilize the stratification. The quantity $c_p T$, which plays an important role in the atmospheric energy balance, is the so-called sensible heat per unit mass. It follows that the stratification is stable or unstable according to whether the sensible heat plus potential energy increases or decreases with elevation.

We shall not present the ω-equation explicitly. Suffice it to say that it is extremely awkward to use. Much of the awkwardness results from the variability of f in (64) and (65).

Equations (63)-(65) describe the so-called geostrophic model, used extensively in numerical weather prediction, usually with additional simplifications. Although it is convenient to be rid of most of the quadratic terms in the vorticity equation, this simplification is too extreme for studying many aspects of the general circulation. An approximation which is less drastic than (62) is the equation of balance

$$\boldsymbol{\nabla}\cdot(\zeta + f)\boldsymbol{\nabla}\psi - \boldsymbol{\nabla}^2(\mathbf{U}_r\cdot\mathbf{U}_r/2) = g\boldsymbol{\nabla}^2 z, \tag{68}$$

obtained by eliminating from the divergence equation all terms which involve the divergence, and hence retaining some of the important quadratic terms. It may be noted that in the ideal case of a stationary circular cyclone or anticyclone the familiar gradient wind formula, which is obtained by equating the pressure-gradient, Coriolis, and centrifugal forces, and which is often used as a refinement of the geostrophic formula for the purpose of estimating winds from pressure data, satisfies the equation of balance.

When equation (68) is used, the quadratic terms in the vorticity equation (57), excepting those involving $\mathbf{U}_d$ alone, must be retained if the energy principle is to be preserved. Thus the equation becomes

$$\partial\zeta/\partial t = -\boldsymbol{\nabla}\psi\cdot\boldsymbol{\nabla}(\zeta + f)\times\mathbf{k} + \boldsymbol{\nabla}\cdot(\zeta + f)\boldsymbol{\nabla}\chi - \boldsymbol{\nabla}\cdot\omega\boldsymbol{\nabla}\,\partial\psi/\partial p + \boldsymbol{\nabla}\ \mathbf{F}\times\mathbf{k}. \tag{69}$$

The appropriate form of the ω-equation is correspondingly more awkward. It is doubtful that it has ever been put to use without numerous further simplifications.

The beta plane

In low latitudes the geostrophic wind is generally regarded as a poor approximation to the actual wind, and at the Equator it becomes infinite. Equation (62) involves only the products of f with the wind and the geostrophic wind, and need not lead to mathematical impossibilities at the Equator, but it is doubtful that it can yield realistic results at low latitudes. The new system of equations may therefore be used to best advantage in problems where the circulation in middle and higher latitudes is of primary concern. For such problems the beta-plane approximation, first introduced by Rossby (1939) in the previously cited paper, greatly simplifies the mathematics. A similar approximation could be used in conjunction with the primitive equations.

In the beta-plane approximation, the spherical surface of the Earth is replaced by a plane in which rectangular Cartesian co-ordinates (x, y) are introduced. The lines $y =$ constant and $x =$ constant are identified with the parallels and meridians. In Rossby's original work the plane was of infinite horizontal extent, but in many subsequent applications it has been restricted to the area between two parallel lines, which are identified with latitude circles. In the x-direction all dependent variables are commonly assumed to vary periodically, acquiring their original values after a distance which is identified with the circumference of the Earth.

In the divergence equation, or in the geostrophic equation which replaces it, the Coriolis parameter f is assigned a constant value. It is also taken as a constant in the vorticity equation, except in the term $-\boldsymbol{\nabla}\psi\cdot\boldsymbol{\nabla}f\times\mathbf{k}$ where its northward derivative $\partial f/\partial y$ is assigned a second constant value β. Thus the term reduces to $-\beta\partial\psi/\partial x$.

The remaining awkward features of the ω-equation result from the variability of σ and the presence of the term $\boldsymbol{\nabla}T\cdot\boldsymbol{\nabla}\chi$ in (65).

The latter term represents the advection of temperature by $\mathbf{U}_d$, and in practice it is usually discarded. The static stability σ is also frequently replaced by $\tilde{\sigma}$, where the tilde ($\sim$) denotes an average over an isobaric surface. Both of these approximations upset the energy principle, but this may be restored by adding a suitable term depending upon p and t alone in the thermodynamic equation. The system of equations may then be written

$$\partial\zeta/\partial t = -\nabla\psi\cdot\nabla\zeta \times \mathbf{k} - \beta\partial\psi/\partial\chi - f\delta + \nabla\cdot\mathbf{F} \times \mathbf{k} \tag{70}$$

$$\partial T/\partial t = -\nabla\psi\cdot\nabla T \times \mathbf{k} + \widetilde{\sigma}\omega + \kappa\,\widetilde{\omega T}/p + Q/c_p \tag{71}$$

$$f\partial\psi/\partial p = -RT/p \tag{72}$$

$$f\partial^2\omega/\partial p^2 + (R\widetilde{\sigma}/fp)\,\nabla^2\omega = \frac{\partial}{\partial p}(\nabla\psi\cdot\nabla\zeta \times \mathbf{k}) - \nabla^2\left(\nabla\psi\cdot\nabla\frac{\partial\psi}{\partial p} \times \mathbf{k}\right) + \beta\frac{\partial^2\psi}{\partial\chi\partial p} - \nabla\frac{\partial\mathbf{F}}{\partial p} \times \mathbf{k} - \frac{\kappa}{fp}\nabla^2 Q. \tag{73}$$

The term containing $\widetilde{\omega T}$ may be omitted in applications where time variations of $\widetilde{T}$ are irrelevant. Usually the variations of $\widetilde{\sigma}$ are also suppressed; if they are to be included, the appropriate equation is obtained from (71).

The greatly simplified ω-equation is now seen to be an elliptic differential equation in ω, since $\widetilde{\sigma}$ is almost invariably positive. In applications involving specific features of the circulation, the terms containing $\mathbf{F}$ and Q are often omitted.

Much effort has been devoted to justifying the use of the beta plane. The general conclusion is that it should yield qualitatively realistic results if its application is restricted to middle and higher latitudes. Certainly it has rendered some problems tractable when they could not otherwise have been handled by analytic procedures.

CHAPTER III

THE OBSERVED CIRCULATION

The characteristic features of the circulation of the Earth's atmosphere form a rather heterogeneous set. They include such familiar qualitative properties as approximate hydrostatic equilibrium, which prevails throughout the atmosphere, and approximate geostrophic equilibrium, which prevails in middle and higher latitudes. They include the quantitative spatial distributions of simple statistics such as the time-averaged wind velocity, and more complicated statistics such as joint probability distributions. Finally they include the existence of such entities as fronts and migratory cyclones, whose presence may not be apparent from an inspection of the quantitative time averages. All of these features serve to distinguish the circulation of our atmosphere from the circulations which may be found in other fluid systems.

In this chapter we shall consider the various aspects of the observed circulation. We shall give particular attention to the fields of motion, temperature, and moisture, averaged with respect to longitude and time. We shall see in the following chapters that even a partial explanation of these fields requires careful consideration of many of the remaining aspects, including the structure of cyclonic and anticyclonic disturbances.

Measurement of the circulation

The meteorologist who wishes to observe the circulation of the atmosphere cannot follow the customary procedures of the laboratory scientist. He cannot design his own experiment in such a manner as to isolate the effects of specific influences, and thereby perhaps disprove certain hypotheses and lend support to others. He can only accept the circulation as it exists. Moreover, since the circulation is global in extent, he cannot even make his measurements singlehandedly, but must rely for the most part upon those which have been made by other persons, in most instances for different purposes.

If the circulation were completely steady, the task of measuring it could be a straightforward matter of geographical exploration. Expeditions could travel to various points of the globe, in the manner of the famous Challenger Expedition of nearly a century ago (see Buchan, 1889), to measure the weather elements at various elevations, and the data gathered by these expeditions could be assembled into a three-dimensional picture of the atmosphere. If the circulation pattern varied with the time of the day and the season of the year, but not otherwise, the task would be prolonged but not too greatly complicated.

But perhaps the most easily observed characteristic of the circulation is its unsteadiness. Fluctuations occur on all space scales and all time scales, and include simple gusts and lulls in the local winds, the development and decay of individual thunderstorms, the passage of migratory cyclones and anticyclones, the extended-period oscillations between high-index and low-index circulation patterns, and the world-wide changes which presumably took place as the prehistoric continental ice-sheets advanced and retreated. If attention is confined to features of larger horizontal scale, the fluctuations of shorter time-scale largely disappear, but the long-term fluctuations remained undiminished.

In Chapter I we noted that the unsteadiness of the atmosphere offered us a choice of theoretical problems; we could consider the sequence of instantaneous weather patterns or confine our attention to the long-term statistical properties of these patterns. We chose the latter problem. We noted that it would be convenient to explain the long-term properties without first deducing the instantaneous patterns, but the possibility did not seem promising. It would likewise be convenient to measure the long-term properties without first observing the instantaneous patterns, but this possibility is also precluded. We are forced by our standard instruments to measure instantaneous quantities, or, more precisely, very short-term averages, since the instruments do not respond to the most rapid fluctuations. The desired long-term statistics must then be estimated by processing the instantaneous measurements.

Ideally the long-term average circulation is the limiting value of the average over a long interval, as the interval becomes infinite. For an idealized atmosphere which is governed by specified mathematical equations, the existence of such an average is sometimes assured. For the real atmosphere such an average, even if it does exist, is not likely to be the one which we desire, or which we can readily estimate from our available data. Over sufficiently long intervals the geographical features will change, and the atmospheric circulation will presumably change in response to the changed geography. Perhaps the most that we can hope to estimate from modern data is the long-term average circulation which would prevail if the geographical features of the Earth, and the output from the sun, could be prevented from undergoing any further changes.

We can never precisely determine long-term averages by computing averages over shorter intervals. One-week averages vary within a season, seasonal or annual averages vary within a decade, and ten-year averages vary within a century. All the weather data ever collected form no more than a statistical sample. The saving feature is that averages over separate intervals of a season or longer are often near enough alike so that either one affords an acceptable estimate of the other for many purposes; thus it is likely that both are acceptable estimates of longer-term averages.

It would therefore appear that averages over complete seasons are a minimum requirement, while considerably longer-term averages are preferable. It would clearly be impossible for one investigator or one small organization to embark upon a programme of performing all the needed measurements. Fortunately, weather data have been collected on a routine basis for a century or more at many locations, and the set of upper-level wind, temperature, and humidity observations which have been made daily or several times a day at each of several hundred stations during the past ten to twenty years forms one of the most remarkable data collections existing in any science.

These observations have been made primarily for the purpose of weather forecasting, and it is doubtful that any other potential use would have brought forth the necessary funds. The needs of weather forecasting have dictated to a considerable extent the locations of the observing stations, which are most abundant in densely populated regions. Over much of the ocean they are virtually absent; even during the International Geophysical Year there was but one southern hemisphere station reporting upper-level winds in the 50 degrees of longitude immediately west of South America. A more regular distribution of stations, perhaps at the intersections of standard parallels and meridians, would better serve the interests of global-circulation research. Nevertheless, the study of the circulation owes a great debt to the practice of weather forecasting, for without these observations our understanding could not have approached its present level.

Yet certain gaps will continue to exist in our knowledge of the circulation as long as extensive regions without regular observations remain. If we acknowledge that the density of observations in the more heavily populated regions is adequate, we must conclude that the present situation could in principle be remedied simply by establishing more weather stations in sparsely populated areas, and maintaining a sufficient number of weather ships in those parts of the oceans where there is little commercial shipping. But even a single weather ship is a costly affair, always vulnerable to discontinuation by an economy-minded government, while even a hundred weather ships would not provide sufficient coverage. If truly global weather information is to be obtained in the foreseeable future, some other procedure must be adopted.

Current plans call for international participation in the World Weather Watch — a truly global observation system in which the World Meteorological Organization will play a co-ordinating role. The primary aim of this system will be the further improvement of weather forecasts, particularly at a range of several days, but it is also intended that the system should enhance our understanding of the global circulation, especially since this understanding seems to be a prerequisite for successful extended-range forecasting. Certain feasibility studies are already in progress. In addition to the expansion of the existing observational network to cover less densely populated land areas, three techniques of observation which are not presently in routine use are under consideration.

First, under the proposed scheme several thousand balloons are to be kept aloft, drifting with the wind at various constant levels. Successive observations of their positions will make it possible to compute the wind fields at these levels. On an experimental basis a number of these balloons have been released in the southern hemisphere. Several balloons drifting at 200 mb have made one or more complete circuits about the globe, and, on 31 December 1966, one balloon had been aloft for 220 days, and had completed 18 circuits, while its latitude oscillated between subtropical and subpolar regions. At 500 mb the balloons have thus far been prevented from remaining aloft for long intervals by the accumulation of ice.

Second, the balloon observations are to be supplemented by more conventional measurements from a large number of floating buoys. These may be anchored or allowed to drift. Like the balloons, the buoys may be considered expendable, to be replaced as needed.

Finally, artificial satellites are to be used for remote sensing of the atmosphere. A wealth of information may be gathered by measuring the radiation received from the atmosphere or the underlying Earth over a wide portion of the spectrum. Measurements of infra-red radiation in several wavelengths are expected to yield reasonably reliable vertical profiles of temperature and humidity above the cloud-tops. Observations in visible light will continue to disclose the cloud systems of various sizes. Measurements in the ultra-violet are expected to reveal the distribution of ozone. It has even been proposed that the satellite, in addition to measuring natural radiation, may carry a downward-directed laser, and gain further information by measuring the back-scattered light.

The other essential role of the satellite will be one of data-collection and transmission. Many of the balloons and buoys will be in remote regions where conventional means of transmitting their measurements will be impracticable. It is planned that a balloon or buoy will store its information until such a time as one of a set of communication satellites passes overhead. At this time the information will be transmitted to the satellite. Later, when the satellite passes over a receiving station, it will retransmit its information, which can then be dispatched by conventional procedures.

Despite the well-known high cost of satellites, such a system is expected to be far less expensive than a large fleet of weather ships, and the necessary funds can reasonably be expected to be forthcoming, at least for a long enough time to determine whether the system is feasible. One can only speculate now as to the eventual benefits to be gained from such a system, but it may well mark the beginning of a new era in the measurement of global weather.

Hydrostatic and geostrophic equilibrium

The most prominent qualitative characteristic of the circulation is the presence of hydrostatic equilibrium, i. e. the approximate balance between gravity and the vertical pressure force. Closely associated with hydrostatic equilibrium is the tendency of the motion to be nearly horizontal. It may be noted that in small-scale turbulent flow or intense cumulus convection the vertical velocities are not small compared to the horizontal velocities, while the smallest-scale features need not even be hydrostatic. It is these very features which we prefer not to regard as part of the circulation. The smoothed circulation, which remains when the small-scale features have been subtracted out, is quasi-hydrostatic and quasi-horizontal.

The familiar hydrostatic equation expressing hydrostatic equilibrium (equation 30) may be written in a form relating pressure, temperature and elevation thus:

$$\partial\,(\ln p)/\partial z = -\,g/(RT). \tag{74}$$

For a moist atmosphere the slightly higher virtual temperature should replace the temperature.

It is perhaps somewhat odd that hydrostatic equilibrium should head a list of observed properties of the circulation, because it is not actually observed on a day-by-day basis; it is taken for granted. Ever since the barometer was invented the supposition that it measures the total weight of a column of air has seldom been seriously questioned. Nearly all routine upper-level observations measure temperature and humidity as functions of pressure, and the elevations at which the specific measurements are made are then calculated with the aid of the hydrostatic equation. Such measurements can neither confirm nor deny hydrostatic equilibrium. Nevertheless, the barometer has been used as an altimeter on countless occasions other than routine weather observations, even to measure small differences in elevation. The general agreement obtained whenever these measurements are compared with measurements by other methods indicates that approximate hydrostatic equilibrium is practically always present.

Moreover, appreciable departures from hydrostatic equilibrium, other than those of short duration which are associated with small-scale motions, would soon lead to large vertical motions, which are not observed. Although the average upward velocity over a large area cannot be directly measured by any existing technique, it is reasonable to assume that it could easily be measured if it were comparable in magnitude to a typical horizontal velocity.

The importance of hydrostatic equilibrium lies in the restrictions which it places on the form which the circulation may assume. The measurement, description, and explanation of the circulation are thereby greatly facilitated. As we have noted, once hydrostatic equilibrium is accepted, it is not necessary to measure both pressure and temperature as functions of elevation; it is sufficient to measure temperature as a function of pressure. The fields of pressure and temperature (or virtual temperature) become manifestations of the same field — the field of mass. The pressure field completely determines the temperature field, while the temperature field, together with the distribution of pressure at sea-level or any other single level determines the pressure field. Contrary to what has often been assumed, however, there is no method of inferring the sea-level pressure field hydrostatically from the three-dimensional temperature field alone.

Hydrostatic equilibrium also renders it unnecessary to measure the nearly unmeasurable vertical-velocity field. The field of vertical motion is assumed to be that field which is needed to maintain hydrostatic equilibrium by offsetting the tendency of the horizontal motions to disrupt it. Moreover, if hydrostatic equilibrium itself can be explained, it is no longer necessary to explain the vertical motion in terms of unbalanced vertical forces; the vertical motion must be that which is required to maintain the equilibrium. This motion is ordinarily found to be very weak compared to the horizontal motion.

In the previous chapter we observed that the presence of hydrostatic equilibrium rendered it convenient to use pressure as the vertical co-ordinate in the system of dynamic equations, and to let height become a dependent variable. It is equally convenient to use pressure as the vertical co-ordinate in presenting the observations. Since World War II routine weather reports have been transmitted in the form of heights, temperatures, humidities, and winds at standard pressures, and constant-pressure weather maps have replaced the previously used constant-level maps.

In this co-ordinate system the hydrostatic equation becomes

$$\partial z/\partial\,(\ln p) = -RT/g\,. \tag{75}$$

Our earlier remarks concerning the fields of pressure and temperature now apply to the fields of height and temperature.

Further qualitative properties which are somewhat analogous to hydrostatic equilibrium and quasi-horizontal motion and nearly as prominent are geostrophic equilibrium, i. e. the approximate balance between the horizontal Coriolis force and the horizontal pressure force, and quasi-non-divergent motion. Again, these properties are not characteristic of small-scale motions.

The geostrophic equation expressing geostrophic equilibrium may be written (see equation 61)

$$\mathbf{U} = (g/f)\,\mathbf{k} \times \nabla z\,. \tag{76}$$

Thus it relates the wind to the slope of a constant-pressure surface. Combined with the hydrostatic equation, it becomes the thermal wind equation

$$\partial\mathbf{U}/\partial p = -(R/fp)\,\mathbf{k} \times \nabla T \tag{77}$$

which relates the vertical shear of the wind to the gradient of temperature along a constant-pressure surface.

Whereas the hydrostatic equation is often treated as exact, the geostrophic equation must be treated as only a fair approximation for many purposes. It is especially unreliable in low latitudes. Thus it has become customary to distinguish between the wind $\mathbf{U}$ and the geostrophic wind $\mathbf{U}_g$, the latter being the wind which would have to accompany the existing height (or pressure) field in order to render equation (76) exact, and therefore being defined by the right side of (76). It would be equally logical to speak of the geostrophic height gradient (or pressure gradient) in referring to the gradient which would have to accompany the existing wind field to render (76) exact, but these expressions do not seem to be in common use. The unfortunate result has been a frequent tendency to assume that the wind field is somehow produced geostrophically by the height field, and to overlook the possibility that the fields determine one another through mutual effects.

Unlike hydrostatic equilibrium, geostrophic equilibrium is directly revealed by modern routine observations. Historically it was not always recognized. During the early nineteenth century there was considerable debate as to whether the wind tended to blow around or into a cyclone. By the middle of the century observations had become sufficient to decide in favour of the former alternative. Near the Earth's surface there is also a noticeable "frictional" component toward low pressure, which undoubtedly made the interpretation of the earlier observations more difficult.

Geostrophic equilibrium, like hydrostatic equilibrium, is important in that its presence facilitates the measurement, description and explanation of the circulation. Where wind measurements are absent, the geostrophically measured wind may often be used with fair confidence. Until fairly recently much of our knowledge of upper-level winds was derived in this manner. To the extent that geostrophic equilibrium prevails, the wind, height (or pressure), and temperature fields become separate manifestations of a single field.

The field of horizontal divergence is not quite strong enough to be reliably determined directly from the wind observations. It is often assumed to be that field which is needed to maintain geostrophic equilibrium by offsetting the disruptive effects of the rotational part of the wind. If geostrophic equilibrium can be explained, the field of divergence need not be explained in terms of unbalanced forces; it must be the field needed to maintain equilibrium. This field is generally found to be weaker than the field of vorticity. It must be emphasized, however, that the geostrophic approximation does not compare in accuracy with the hydrostatic approximation, and for many purposes separate measurements of wind and height (or pressure) and direct measurements of the divergence are desirable; this is especially true in tropical regions.

Resolution of the circulation

The observed fields of motion, temperature, and moisture in the atmosphere cannot be represented by any simple analytic formulae, and quantitative statistics of the circulation are most easily presented in the form of tables or graphs. The three-dimensional spatial distribution of any particular statistic, such as the time-averaged wind velocity, can be reasonably well represented by a set of two-dimensional charts, which may be horizontal maps or vertical cross-sections. However, a collection of charts consisting of a separate set of maps or cross-sections for every statistic of interest would be altogether unwieldy, and it would be quite incomplete if the statistics were limited to such familiar quantities as means and standard deviations. While a map of the time-averaged field of motion might afford a good description of the trade winds, it would not reveal the prevalence of migratory cyclones in higher latitudes. Maps of covariances at suitable time-lags and space-lags might imply the existence of cyclones, but a map of cyclone frequency would serve the purpose more readily.

It is evident that nothing short of a comprehensive atlas containing hundreds (or more likely thousands) of charts could present nearly all of the quantitative statistics of possible importance. We shall therefore limit our quantitative presentation to a few statistics which will have a special bearing on the remainder of this work, recognizing that these may not be the statistics of greatest interest to one who is pursuing a slightly different problem. We shall follow this account with a qualitative description of some of the remaining features of special significance.

It will be convenient to classify the principal features of the circulation into four categories, as follows:

(1) *Features which appear when the variables are averaged with respect to time and longitude.* These are typified by the familiar trade winds. A time average may mean an average over all time, or over all years at a particular time of the year. Some writers prefer to restrict the term "general circulation" to features in this category, and it is these features which, directly or indirectly, will receive the major attention in this monograph.

(2) *Features in addition to those of the first category which appear when the variables are averaged with respect to time alone.* These are typified by the Asiatic summer and winter monsoons. Most writers include these as features of the general circulation.

(3) *Features in addition to those of the first category which appear when the variables are averaged with respect to longitude alone.* These are typified by the familiar fluctuations of the zonal index. Studies of these features are ordinarily regarded as general-circulation studies by those engaged in them.

(4) *Features in addition to those of the first three categories which appear when the variables are not averaged.* These are typified by migratory cyclones. Many of these features are ordinarily regarded as secondary circulations; some of their over-all statistical properties are frequently considered to be characteristics of the general circulation.

Following the modified notation of Starr and White (1954), we shall let a bar (—) denote the time average of any quantity, and a prime (′) the departure of a quantity from its time average. Likewise we shall let brackets ([ ]) denote the average of any quantity with respect to longitude, and a star (*) the departure of a quantity from its longitudinal average. It is evident that the operators —, ′, [ ] and * are commutative.

The wind field $\mathbf{U}$ may now be resolved according to the formulae

$$\mathbf{U} = \overline{\mathbf{U}} + \mathbf{U}' \tag{78}$$

$$\mathbf{U} = [\mathbf{U}] + \mathbf{U}^* \tag{79}$$

and thus, in greater detail,

$$\mathbf{U} = [\overline{\mathbf{U}}] + \overline{\mathbf{U}}^* + [\mathbf{U}]' + \mathbf{U}^{*\prime} \tag{80}$$

Resolutions of the fields of temperature T, specific humidity q, and other quantities may be similarly performed.

There remains some ambiguity in the definitions of the averages, which must be removed. An average over time t may be an average for fixed values of λ, φ, z, or for fixed values of λ, φ, p, or for some other choice of independent variables. One type of averaging is not identical with another. Since our data consist mainly of observations at standard pressure levels, we shall let $\overline{\mathbf{U}}$ represent the time average for fixed λ, φ, p. Likewise, $[\mathbf{U}]$ will be the average with respect to longitude λ for fixed t, φ, p, i. e. the average along a latitude circle on an instantaneous isobaric surface.

Strictly speaking the features of $\mathbf{U}$ in the four categories previously enumerated are respectively those features appearing in the field of $[\overline{\mathbf{U}}]$, in $\overline{\mathbf{U}}$ but not $[\overline{\mathbf{U}}]$, in $[\mathbf{U}]$ but not $[\overline{\mathbf{U}}]$, and in $\mathbf{U}$ but neither $\overline{\mathbf{U}}'$ nor $[\mathbf{U}]$. To a considerable extent these are the features appearing respectively in the fields of $[\overline{\mathbf{U}}]$, $\overline{\mathbf{U}}^*$, $[\mathbf{U}]'$ and $\mathbf{U}^{*\prime}$. Similar remarks apply to the features of T and q. Some features, such as the jet stream, do not clearly fall into any one category.

Although the notation used in (78)-(80) has been adopted by a number of writers, there seems to be less uniformity in the accompanying terminology. We shall refer to the fields of $\overline{\mathbf{U}}$ and $\mathbf{U}'$ in (78) as the long-term or time-averaged or standing motion and the transient motion. We shall refer to the components $[u]$ and $[v]$ of $[\mathbf{U}]$ in (79) as the zonal circulation and the meridional circulation, and to the components of $\mathbf{U}^*$ as the eddies. We shall also include the field of $[\omega]$ demanded by continuity as part of the meridional circulation. Thus the terms in (80) become respectively the time-averaged or standing zonal and meridional circulations, the time-averaged or standing eddies, the transient zonal and meridional circulations, and the transient eddies.

The customary use of the terms "zonal" and "meridional" has led to some ambiguity. A "zone" generally means a latitude circle or a region extending along a latitude circle. "Zonal motion" generally means motion *parallel* to the zones, and is synonymous with u, while "meridional motion" means motions parallel to the meridians or meridional planes, and may be synonymous with v, or with v and ω. A "zonal average" generally means an average *within* zones, or with respect to longitude. "Zonal symmetry" generally denotes invariability within zones. We shall adhere to this usage.

The ambiguity arises in connection with the term "zonal circulation", which is sometimes used to mean the zonal motion u, sometimes the zonally-averaged motion $[\mathbf{U}]$, and sometimes zonally-averaged zonal motion $[u]$. We shall use the term only in the last sense. Likewise we shall use "meridional circulation" only to denote zonally-averaged meridional motion. The term "mean meridional circulation" has been used for the latter purpose, but it has also been used for the time-averaged meridional circulation. The frequently used term "mean motion" is not specific enough when both time averages and zonal averages are being considered.

The long-term zonally averaged circulation

Within this chapter we shall limit the quantitative statistics to the fields of $[\overline{\mathbf{U}}]$, $[\overline{T}]$ and $[\overline{q}]$. We do not imply by this limitation that the long-term zonally averaged fields are the only ones of importance. It does appear that these fields have received the greatest amount of theoretical attention. We shall presently see that a proper explanation of them also involves the transient motions and the eddies.

We consider first the distribution of $[\bar{u}]$, the long-term zonal circulation. In view of the copious data which is continually accumulating in ever greater amounts, it might seem that this field should be rather precisely known by now. This does not appear to be the case.

The logistics of weather data are rather intricate. The fact that an observation has been performed in the prescribed manner is no assurance that it will find its way into any particular data collection; it is even less certain that it will arrive without errors. Those who make immediate use of the data — the weather forecasters — are generally not the ones who are charged with storing them for possible future use. So many data are now stored in various collections that the mere process of extracting the portion needed for a particular study is a formidable task. The use of large digital computing machines has made it possible to handle sets of data which would otherwise be completely unwieldy, but it has also made it easy to overlook the type of error which would have been immediately evident in the days when the processing was done by hand. One or two wind speeds recorded on punched cards or magnetic tape as 500 m sec^{-1} instead of 50 m sec^{-1}, for example, will render a computed statistic quite worthless.

In any event, there appear to be no estimates of $[\bar{u}]$ based upon a major portion of the upper-level wind observations which have been collected since World War II. A number of studies have been based upon smaller portions of the data.

Figures 1 and 2 present the distribution of $[\bar{u}]$ for northern winter and southern summer (October-March) and southern winter and northern summer (April-September), as computed in separate studies by Buch (1954) and Obasi (1963). The data samples for these studies were somewhat limited. Buch used northern-hemisphere data for 1950 only, while Obasi used southern-hemisphere data for 1958 only. However, the studies are distinguished in that they are based entirely on wind observations, whereas the more extensive studies thus far completed have depended partly upon winds estimated from pressure observations. Moreover, Buch and Obasi used the same computational procedure, and the number of stations (145) available in the southern hemisphere in 1958 was comparable to the number (81) available in the northern hemisphere in 1950.

The method of computation consisted of determining the time average $\bar{u}$ at the 850-, 700-, 500-, 300-, 200- and 100-mb levels at each station, using all available data. These averages were then entered on hemispheric maps, and isopleths were drawn. From the isopleths, values of $\bar{u}$ at the intersections of standard parallels and meridians were recorded, and these values were averaged to obtain estimates of $[\bar{u}]$.

The most complete compilation of northern-hemisphere winds seems to be that of Crutcher (1959, 1961), who has prepared horizontal maps of $\bar{u}$ and other statistics at the 850- ,800-, 500-, 300-, 200- and 100-mb levels, and vertical cross-sections of these statistics for each ten degrees of longitude. It is a simple matter to average Crutcher's values; the resulting values of $[\bar{u}]$ are shown for winter (December-February) and summer (June-August) in Figures 3 and 4.

Crutcher's values are based upon at least five years of data in most regions, and in this respect are superior to Buch's. We prefer, however, not to combine Crutcher's values and Obasi's in the diagrams, since they were computed by different procedures and probably do not afford a reliable comparison of the hemispheres. In particular, in regions where observed-wind data were scarce, Crutcher used winds which had been estimated by the gradient-wind formula from constant-pressure charts, although he avoided the simpler geostrophic-wind formula.

Another thorough study covering both hemispheres is the one by Heastie and Stephenson (1958), also based upon five years of data. Cross-sections for January and July appear in Figures 5 and 6. Here no attempt has been made to utilize observed winds except in the tropics. North of 25°N and south of 25°S the winds are geostrophically estimated from the contours on constant pressure charts.

Finally, we mention a detailed and frequently quoted study by Mintz (1954), based essentially upon all available appropriate data prior to 1950. Again, the winds are geostrophically estimated, except between 20°N and 20°S. Except near Antarctica, Mintz's southern-hemisphere data are restricted to the longitudes of Australia and New Zealand. We therefore present only his northern hemisphere values of $[\bar{u}]$; these are shown in Figures 7 and 8.

Returning to Figure 1, we note that Buch finds a winter westerly wind maximum of 23 m sec^{-1} slightly below the 200-mb level at about 35°N. In Figure 3, however, the maximum has increased to 34 m sec^{-1}, and it is found slightly farther south. In Figure 5, it is similarly located, but it has attained a strength of 37 m sec^{-1}. Finally, in Figure 7, Mintz locates the maximum south of 30°N, with a strength of 42 m sec^{-1}. In view of this great diversity of estimates, we find it difficult to maintain that the average zonal westerly wind field is quantitatively known.

Some of the disagreement among the estimates can be easily explained. Buch found it necessary to combine the six months October-March for his winter computations, to obtain a reasonably large sample. Crutcher and Mintz, having more years of data at their disposal, used December-February, when the winds are stronger than in October and November, while Heastie and Stephenson used January alone.

Perhaps equally important is the bias toward light winds which is ordinarily present in observed-wind data. One of the principal causes of missing upper-level reports is the occurrence of excessively strong winds, which carry the balloon beyond the range of the receiving instrument before the highest elevations are reached. Virtually all collections of upper-level wind data, but particularly the less recent ones, are therefore biased in favour of light winds. Geostrophically estimated winds suffer only slightly from this bias for, although strong winds will cause a pressure observation as well as a wind observation to be missing, it is the pressure on either side of a strong current rather than the pressure within the current which is used in estimating the strength of the current.

Moreover, even without missing data the geostrophic approximation systematically overestimates the winds in the zones of strong westerlies. In the free atmosphere a wind equal to the geostrophic wind would have no horizontal acceleration, and would therefore tend to follow a great-circle trajectory. But, on the average, trajectories in the zone of westerlies are curved at least as greatly as the latitude circles; the average winds are therefore subgeostrophic. From (45), (47) and (48) we find upon averaging that

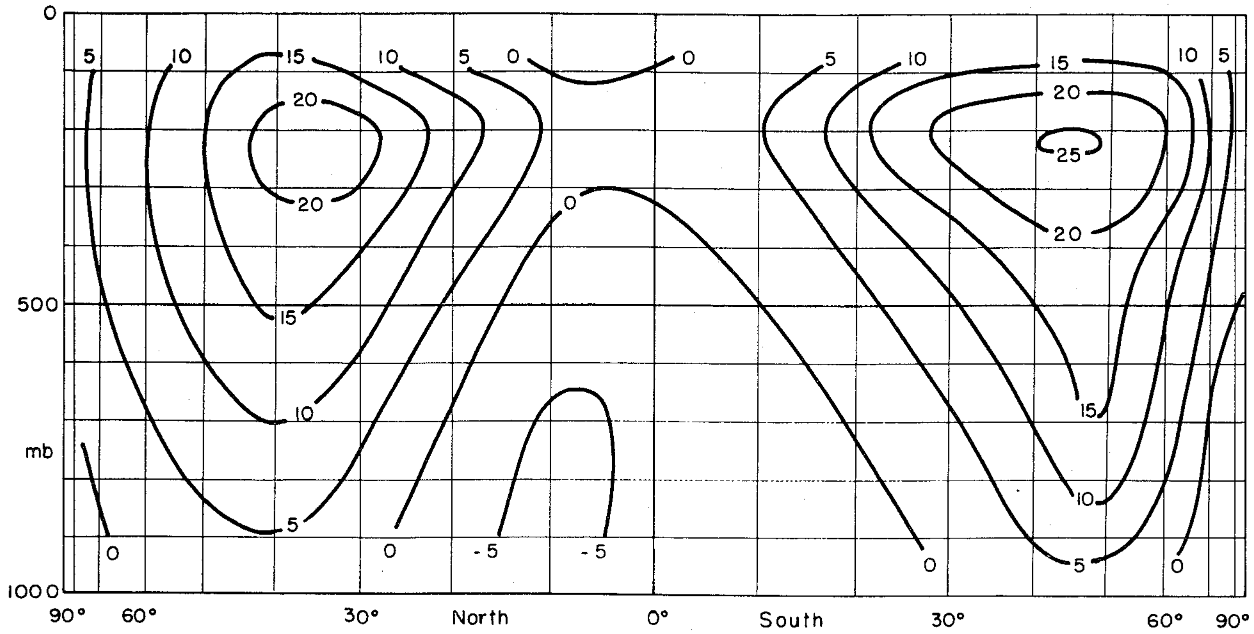


Figure 1. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in northern winter and southern summer (October-March) as estimated by Buch (1954) and Obasi (1963). Values are in m sec^{-1}

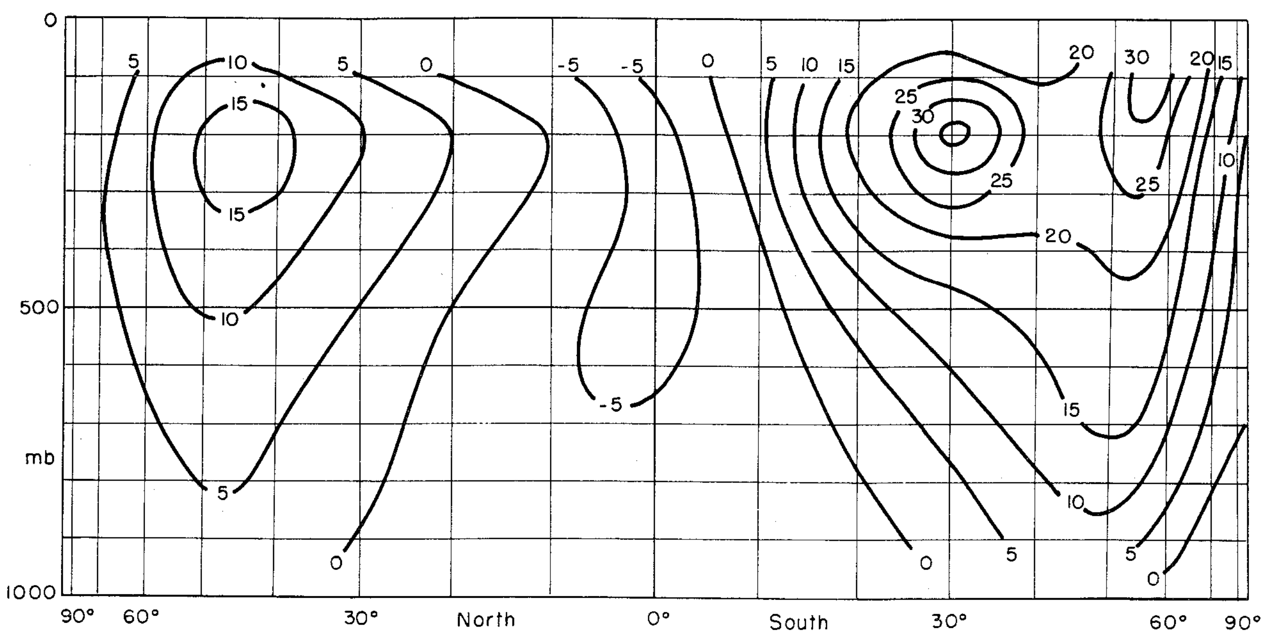
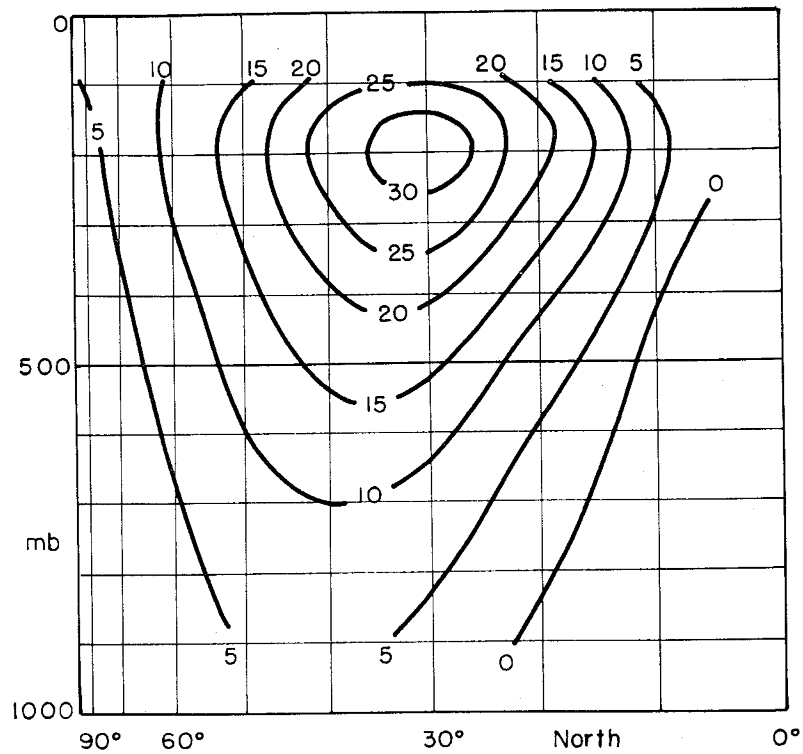


Figure 2. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in northern summer and southern winter (April-September) as estimated by Buch (1954) and Obasi (1963). Values are in m sec^{-1}

Figure 3. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in northern winter (December-February) as estimated from charts compiled by Crutcher (1959, 1961). Values are in m sec^{-1}



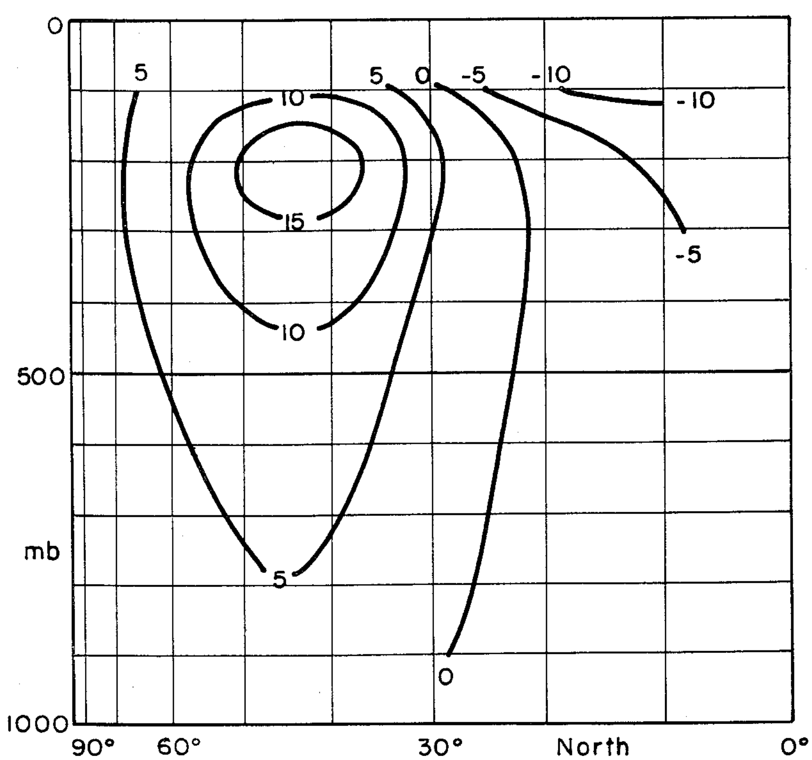


Figure 4. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in northern summer (June-August) as estimated from charts compiled by Crutcher (1959, 1961). Values are in m sec^{-1}

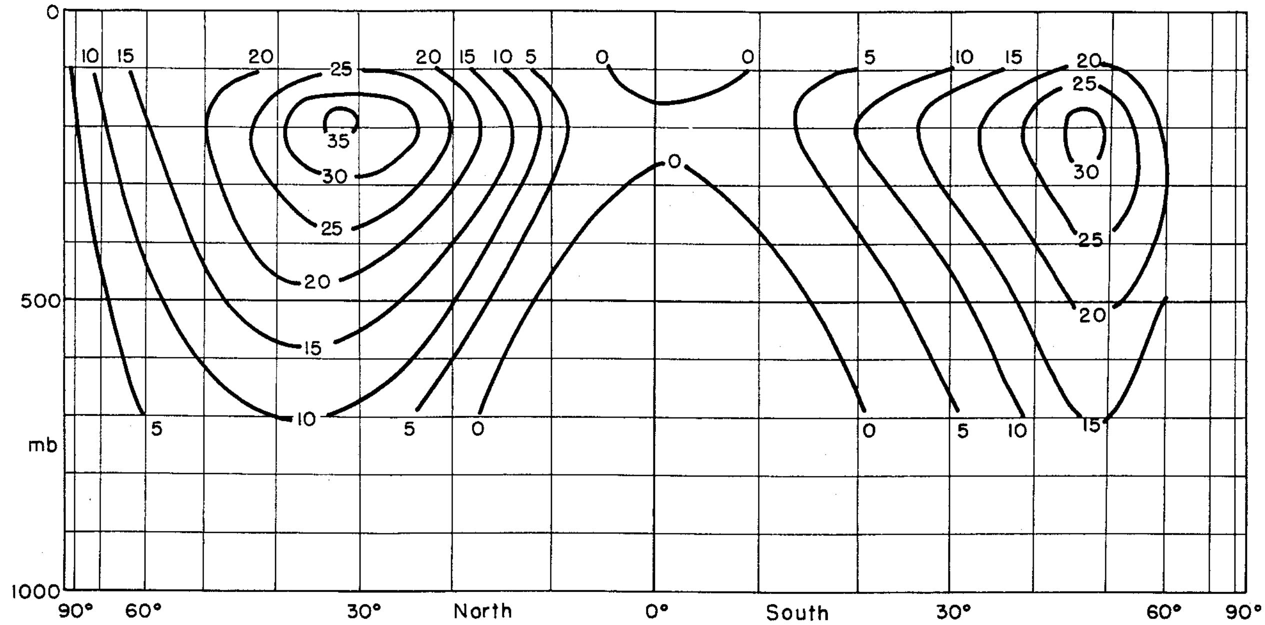


Figure 5. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in January as estimated by Heastie and Stephenson (1958). Values are in m sec^{-1}

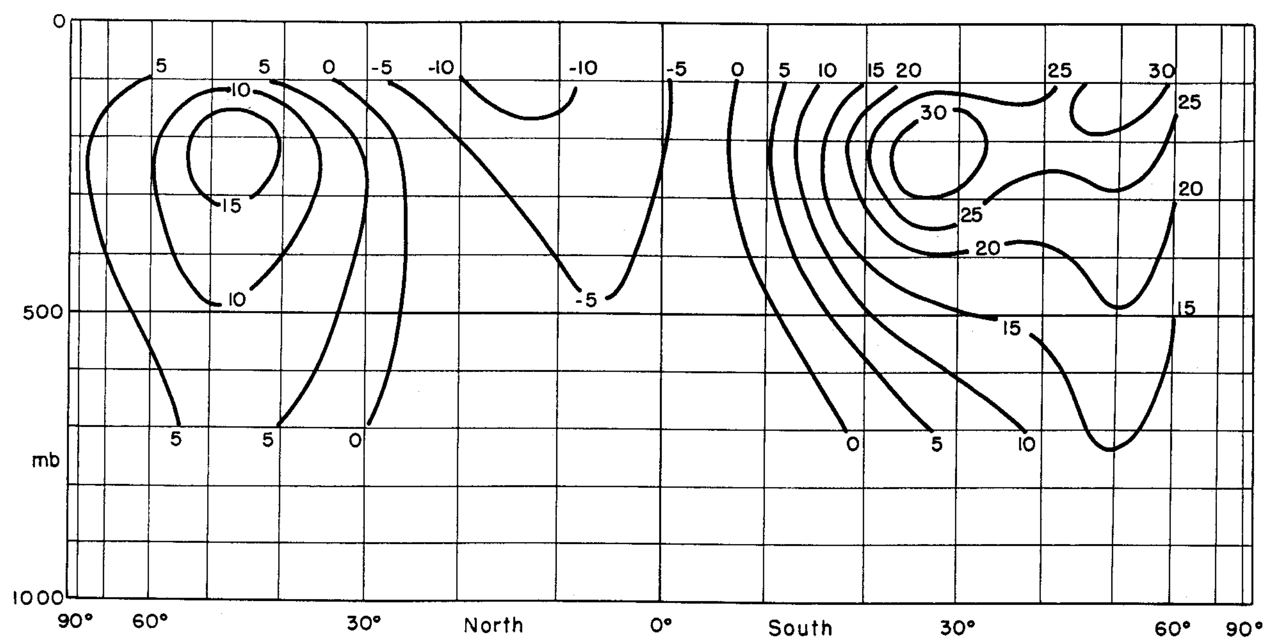
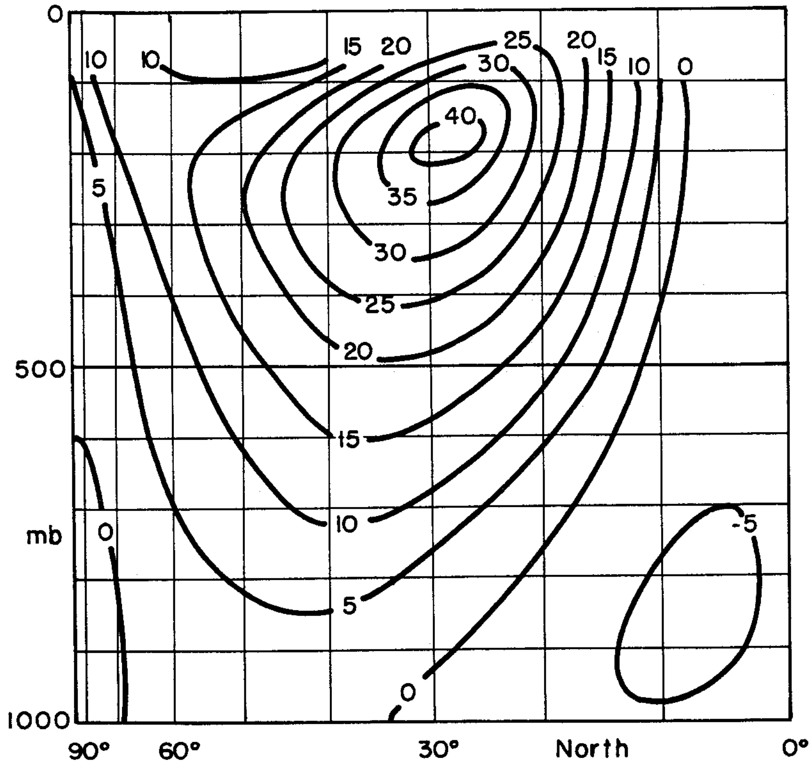


Figure 6. — The time-and-longitude averaged zonal wind $[\bar{u}]$ in July as estimated by Heastie and Stephenson (1958). Values are in m sec^{-1}

Figure 7. — The time-and-longitude averaged zonal wind $[\overline{u}]$ in northern winter (December-February) as estimated by Mintz (1954). Values are in m sec^{-1}



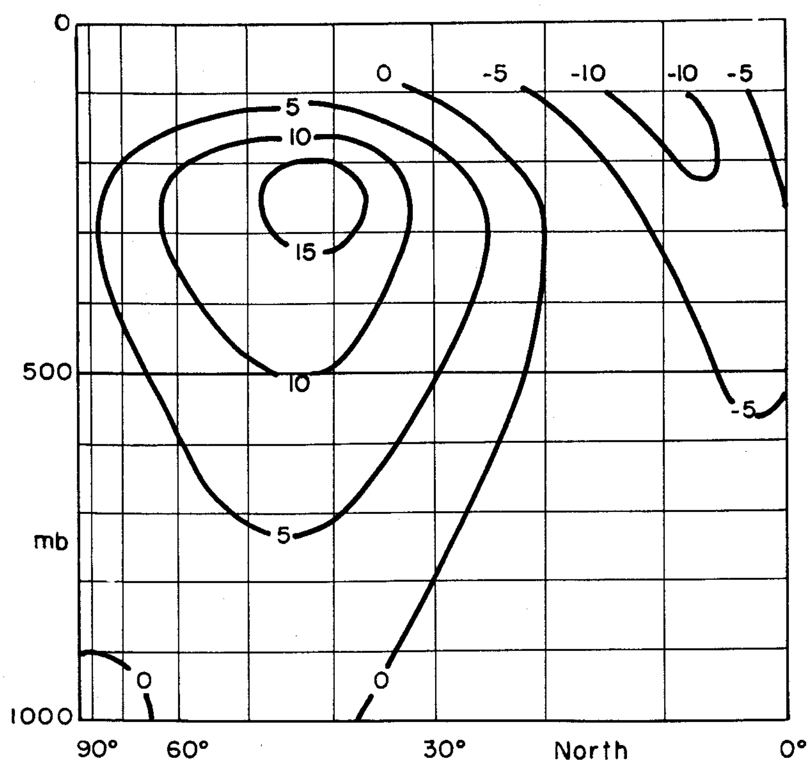


Figure 8. — The time-and-longitude averaged zonal wind $[\overline{u}]$ in northern summer (June-August) as estimated by Mintz (1954). Values are in m sec^{-1}

$$f[\bar{u}_g] = f[\bar{u}] + \frac{\tan\varphi}{a}[\overline{u^2}] + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\cos\varphi\,[\overline{v^2}] + \frac{\partial}{\partial p}[\overline{v\omega}] - [\bar{F}_\varphi] \tag{81}$$

The final term in (81) is presumably small, and the preceding term is difficult to estimate, but since it disappears in the vertical average it cannot be of one sign everywhere. The other terms depend upon readily estimated statistics. Crutcher's charts include the distributions of $\bar{v}$ and the standard deviations of u and v, from which $[\overline{u^2}]$ and $[\overline{v^2}]$ may be evaluated. Holopainen (1966) has evaluated the annual average nongeostrophic zonal wind $[\bar{u}]-[\bar{u}_g]$; it is shown in Figure 9. For the winter alone, $[\bar{u}_g]$ exceeds $[u]$ by 2.3 m sec^{-1} at 200 mb and 30°N, and the disparity between Figure 3 and 5 is thus nearly accounted for.

Yet, lest we be too hasty in maintaining that the three influences just mentioned completely explain the discrepancies, let us note that they are equally present in the southern hemisphere. Obasi's winter maximum of $[\bar{u}]$, based on observed winds, ought therefore to fall far short of the maximum given by Heastie and Stephenson. Yet reference to Figures 2 and 6 reveals essentially no difference. It appears that some of the differences between the estimates must result simply from the finite sizes of the samples. Different samples inevitably possess different mean values.

Despite the quantitative differences just noted, the qualitative features of $[\bar{u}]$ seem to be fairly well defined. At the surface there are the familiar trade winds and prevailing westerlies, with weak easterlies again in the polar regions. The eastward wind component increases upward from the surface to about 200 mb everywhere, except in low latitudes in northern summer. From 200 mb to considerably above 100 mb it decreases everywhere, except in high latitudes in winter, where it continues to increase. In the southern hemisphere there is a double maximum in winter. The only significant disagreement is in the equatorial regions, where Mintz's data show easterlies at all levels, while the remaining studies show westerlies near 200 mb in winter.

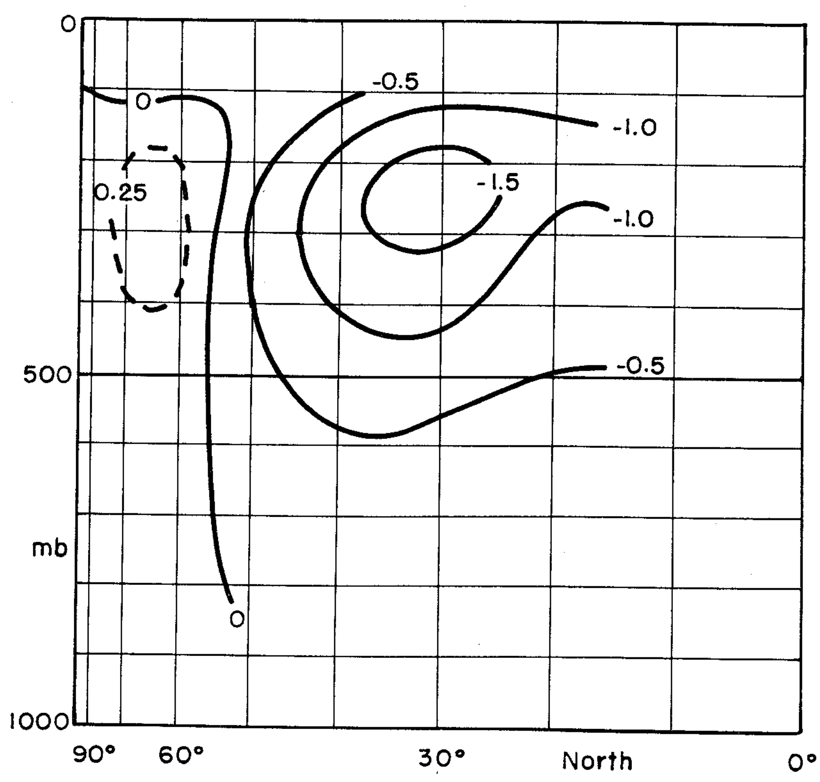


Figure 9. — The time-and-longitude averaged non-geostrophic zonal wind $[\bar{u}-\bar{u}_g]$ as estimated by Holopainen (1966). Values are in m sec^{-1}

Since the geostrophic wind affords a rather good estimate of $[\bar{u}]$ despite its tendency to overestimate the westerlies, many of the principal features just cited should have their counterparts in the field of $[\overline{T}]$. This proves to be the case. For uniformity, we present first the cross-sections of Peixoto (1960), supplemented by values at 50 mb and 30 mb obtained by Peng (1963, 1965), in Figures 10 and 11; these have been computed by a procedure similar to Buch's. Peixoto's data are again for the year 1950, and the network of stations is essentially the same; Peng's data are for 1958.

The cross-sections using the most complete compilation of data appear to be those of Palmén and Newton (1967). They are shown in Figures 12 and 13. They have been based largely upon the detailed maps of Goldie *et al.* (1958), which were constructed mainly from observations made during 1941-1952. Some aircraft measurements were used in regions where radiosonde data were scarce.

Unlike the estimates of $[\bar{u}]$, the estimates of $[\overline{T}]$ are in good quantitative agreement. The somewhat higher temperatures during the winter and lower temperatures during the summer obtained by Peixoto presumably occur because Peixoto's winter and summer were actually the six-month periods October-March and April-September, while Palmén and Newton presented data for January and July. We might add that no investigator has seen fit to estimate the temperature field geostrophically from wind observations.

The one feature of the temperature field which possesses no geostrophic counterpart in the wind field is its vertical variation. Here the principal feature is the separation of the atmosphere into the troposphere, where — except at low elevations in the polar regions in winter — the temperature decreases with elevation, and above it the stratosphere, where — again except in the polar regions in winter — the temperature no longer decreases. The tropopause separating these regions is not sharply defined in the averaged temperature fields, and will be mentioned later.

The increase of $[\bar{u}]$ with height in the troposphere is the geostrophic equivalent of the poleward decrease of $[\overline{T}]$. The decrease of $[\bar{u}]$ in the stratosphere is the equivalent of the poleward increase of $[\overline{T}]$ there. In the polar regions in midwinter the temperature decreases poleward even in the stratosphere, and the westerly wind speed increases with elevation. Near the Equator, where the geostrophic equation is less dependable, no important horizontal variations of $[\overline{T}]$ are revealed.

Peixoto and Crisi (1965) have also made estimates of the zonally averaged specific humidity $[\bar{q}]$, using northern hemisphere data for 1958. Data became much more plentiful between 1950 and 1958, and 345 stations were available. Their computational procedure was again similar to the one used by Buch. Their cross-sections are shown in Figures 14 and 15.

At the surface in tropical latitudes $[\bar{q}]$ is very high; in fact it is higher than the values which would prevail at temperatures a few degrees lower under saturated conditions. Thus $[\bar{q}]$ falls off with increasing latitude and elevation, and to a first approximation is determined by the field of $[\overline{T}]$.

Indeed, the variations of q follow those of the saturation specific humidity q_s so closely that some of the interesting aspects of the field of moisture are more readily seen in the field of zonally averaged relative humidity $[\overline{q/q_s}]$. Figures 16 and 17 present tropospheric estimates by London (1957). The outstanding feature is the dry region in the subtropics at middle levels in both winter and summer.

Other estimates (see the discussion by Manabe *et al.* 1965, p. 776) have indicated much lower relative humidities in the upper troposphere. Above 500 mb all estimates seem to be based upon rather limited data.

We finally consider the long-term meridional circulation $[\bar{v}]$ and the field of $[\bar{\omega}]$ related to it through continuity. Unlike $[\bar{u}]$, which often affords a moderately good approximation to instantaneous values of u,

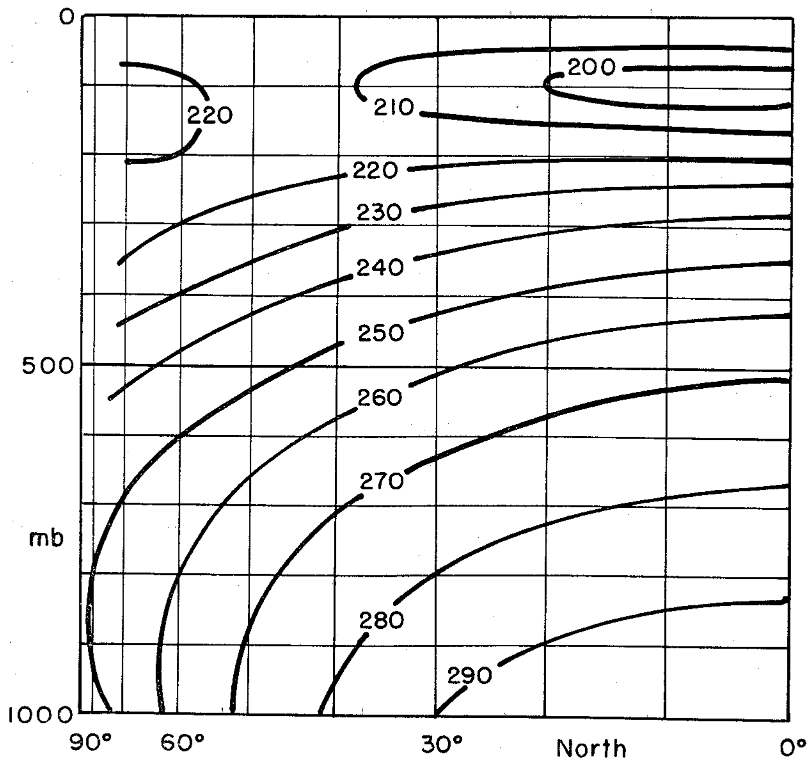


Figure 10. — The time-and-longitude averaged temperature $[\overline{T}]$ in northern winter (October-March) as estimated by Peixoto (1960) (1000 mb-100 mb) and Peng (1963, 1965) (100 mb -30mb). Values are in degrees K

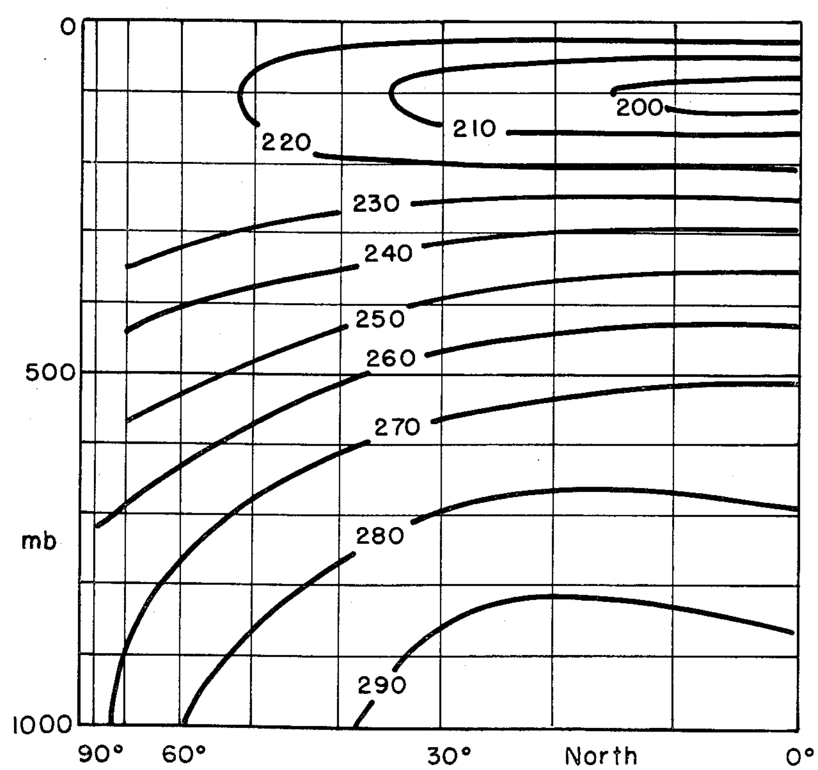


Figure 11. — The time-and-longitude averaged temperature $[\overline{T}]$ in northern summer (April-September) as estimated by Peixoto (1960) 1000 mb-100 mb) and Peng (1963, 1965) (100 mb-30 mb). Values are in degrees K

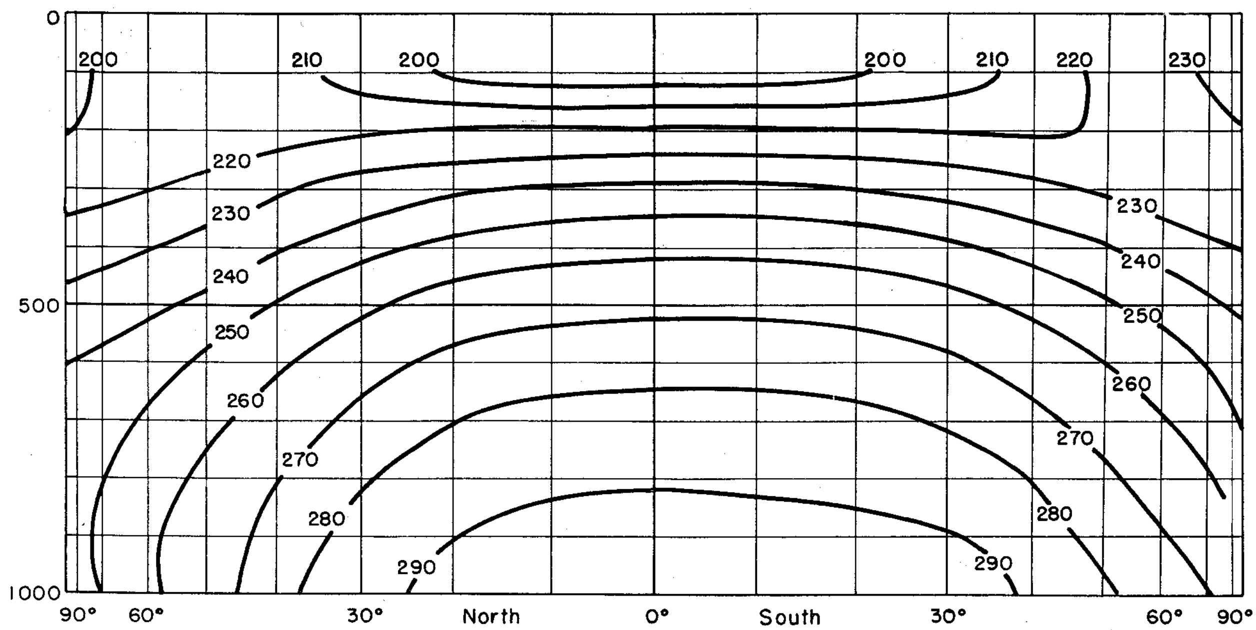


Figure 12. — The time-and-longitude averaged temperature $[\overline{T}]$ in January as estimated by Palmén and Newton (1967). Values are in degrees K

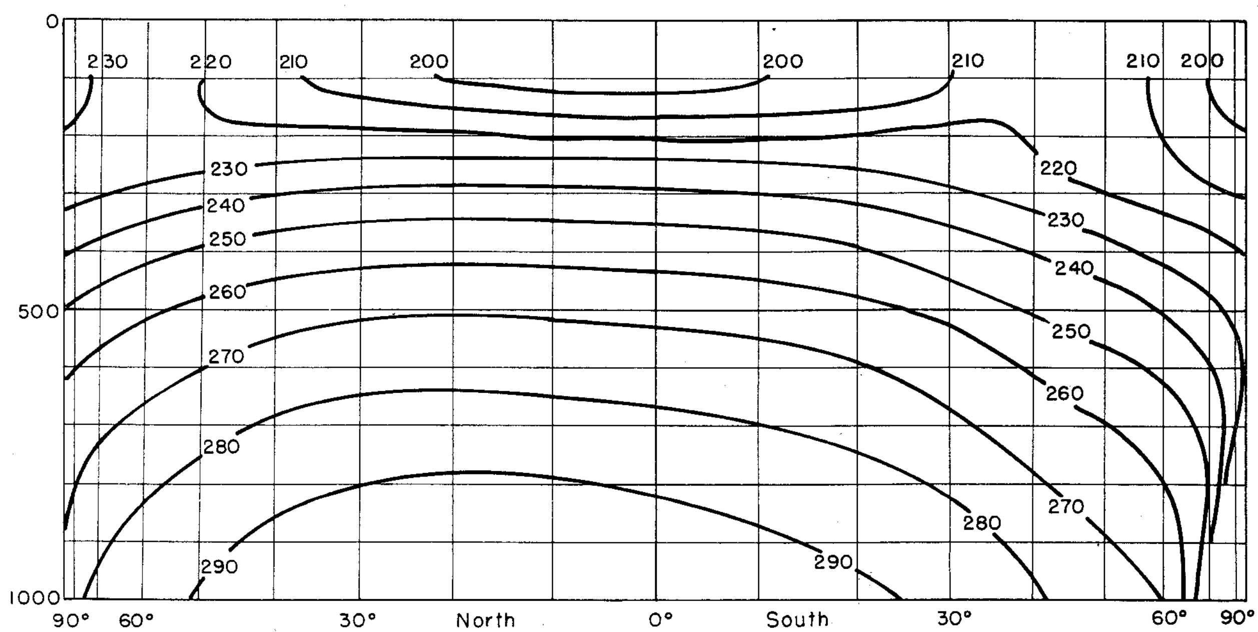


Figure 13. — The time-and-longitude averaged temperature $[\overline{T}]$ in July as estimated by Palmén and Newton (1967). Values are in degrees K

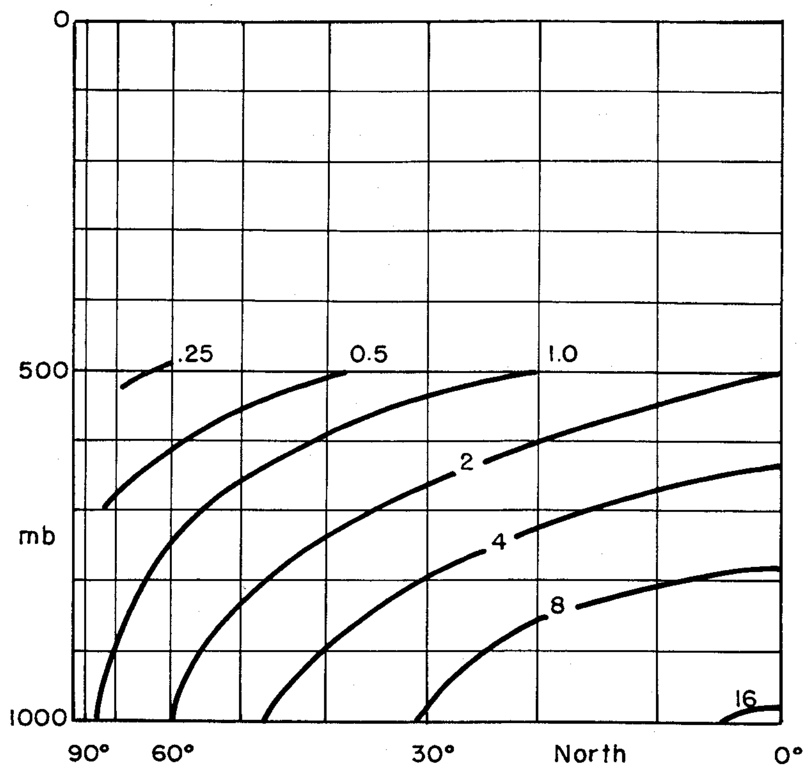


Figure 14. — The time-and-longitude averaged specific humidity $[\bar{q}]$ in northern winter (October-March) as estimated by Peixoto and Crisi (1965). Values are in thousandths

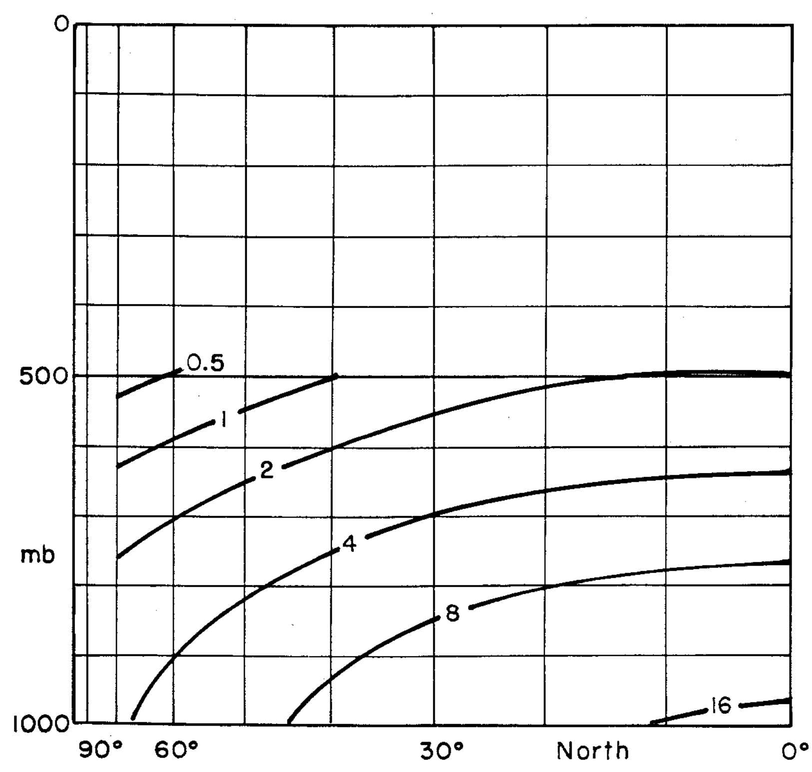
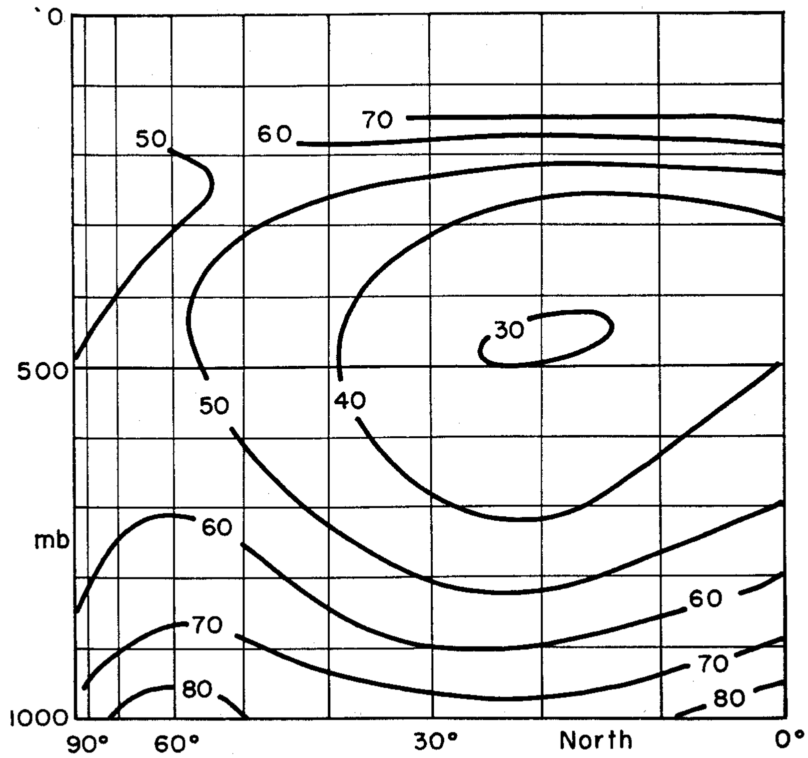


Figure 15. — The time-and-longitude averaged specific humidity $[\bar{q}]$ in northern summer (April-September) as estimated by Peixoto and Crisi (1965). Values are in thousandths

Figure 16. — The time-and-longitude averaged relative humidity $[\overline{q/q_s}]$ in northern winter as estimated by London (1957). Values are in per cent



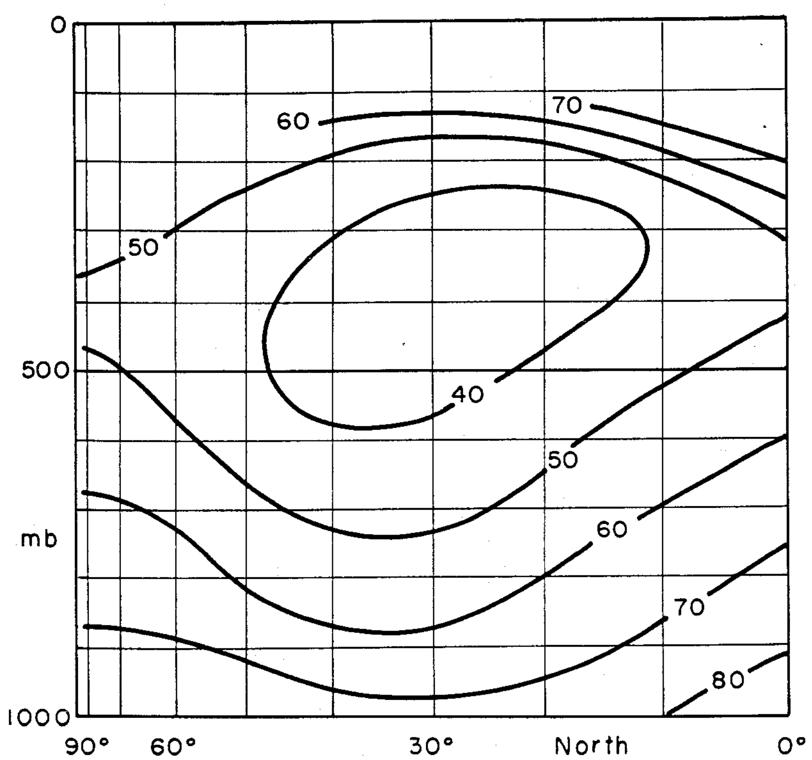


Figure 17. — The time-and-longitude averaged relative humidity $[\overline{q/q_s}]$ in northern summer as estimated by London (1957). Values are in per cent

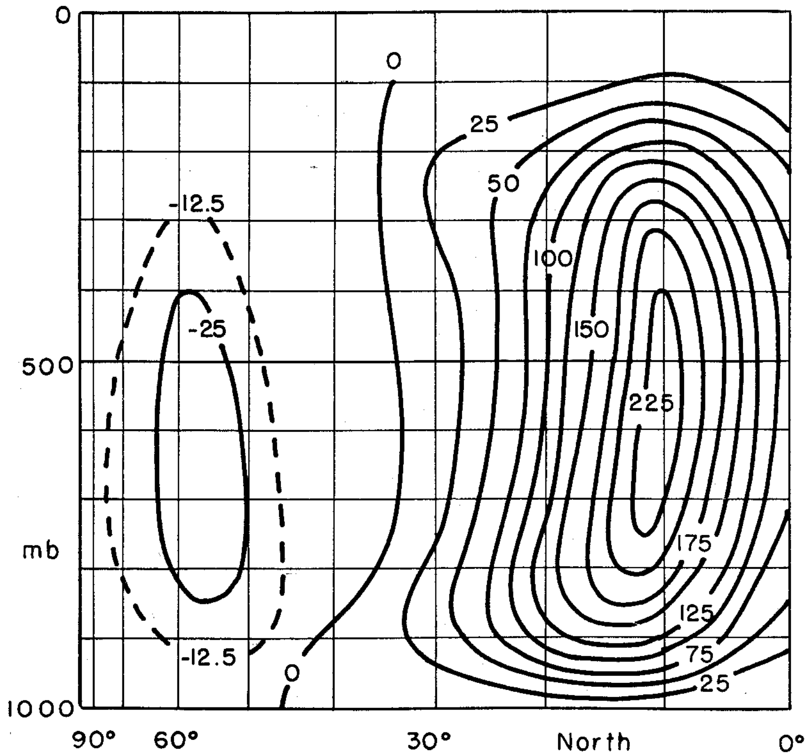


Figure 18. — The time-averaged meridional circulation in northern winter as estimated by Palmén and Vuorela (1963). The unit for stream function $\overline{\Psi}$ is 10^{12} g sec^{-1}

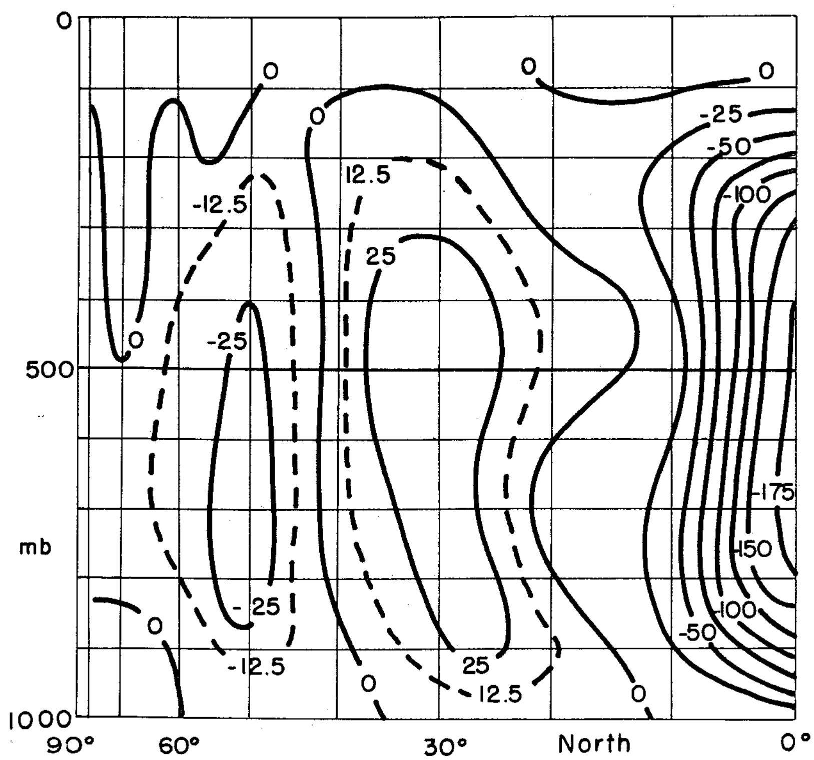


Figure 19. — The time-averaged meridional circulation in northern summer as estimated by Vuorela and Tuominen (1964). The unit for stream function $\overline{\Psi}$ is 10^{12} g sec^{-1}

$[\bar{v}]$ is essentially a statistical residual which results from averaging large values of v having opposite signs. An exception occurs within the trade winds, where the equatorward component is fairly persistent. Hence far less confidence can in general be placed in estimates of $[\bar{v}]$ than in those of $[\bar{u}]$, which themselves are somewhat uncertain. In Figures 18 and 19 we present recent estimates of the time-averaged meridional circulation by Palmén and Vuorela (1963) for winter, and by Vuorela and Tuominen (1964) for summer.

The lines shown are streamlines of mass flow. On the basis of equations (42) and (48) it is possible to introduce a "stream function" Ψ for total mass flow such that

$$2\pi\ a \cos\varphi\,[v] = g\,\partial\,\Psi\,/\,\partial p \tag{82}$$

$$2\pi\ a^2 \cos\varphi\,[w] = -\,g\,\partial\,\Psi\,/\,\partial\varphi\,. \tag{83}$$

With $\Psi = 0$ along the Earth's surface, the values of $\bar{\Psi}$ aloft can be determined from observed values of $[\bar{v}]$.

A meridional circulation ordinarily consists of one or more meridional cells, each cell being a circulation about an extreme value of Ψ. A cell in which the warmer air rises and the colder air sinks is called thermally direct; if the warmer air sinks and the colder air rises, it is thermally indirect. The circulation envisioned by Hadley contains a single direct cell in each hemisphere. As we shall show in Chapter V, in a thermally forced circulation a direct cell converts the energy introduced by the heating into kinetic energy; an indirect cell has the opposite effect, and must depend upon the remainder of the circulation for its kinetic energy.

Figure 18 shows a strong direct cell in low latitudes. Such a cell is now called a "Hadley cell". In higher latitudes there is a weaker indirect cell. In northern summer the southern-hemisphere Hadley cell extends well into the northern hemisphere. The equatorward flow in the Hadley cell is confined mainly to the surface friction layer, while the return flow, rather than being uniformly spread through the remainder of the atmosphere, is concentrated near the tropopause.

Needless to say, various estimates of $[\bar{v}]$ and $[\bar{\omega}]$ differ considerably. Tucker (1959), for example, finds a less intense Hadley cell (in winter), a well developed middle-latitude indirect cell, and good indications of another direct cell in the polar regions. The long-term meridional circulation can also be estimated by various indirect methods. We do not regard these estimates as part of the "observed circulation", and we shall delay their description until the following chapter.

The eddies and the transient motions

As noted, a reasonably complete presentation of the features in the remaining three categories previously mentioned would require an inconveniently large collection of charts. We shall merely describe some of the main qualitative features. The quantitative influence of these features upon the fields of $[\bar{\mathbf{U}}]$, $[\bar{T}]$, and $[\bar{q}]$ will be considered in the following chapter.

Features of the second category serve to distinguish the climate of one locality from that of another in the same latitude, and have received much attention from climatologists. They would not be present in an idealized atmosphere without geographical features to distinguish one longitude from another. As a result, they have been disregarded altogether in many theoretical studies of a fluid dynamical nature, while in other studies simple idealized oceans and continents have been introduced.

In general there is a tendency toward high sea-level pressure with anticyclonic circulation over the continents and low sea-level pressure with cyclonic circulation over the oceans in winter, except at very low latitudes. The opposite situation tends to prevail in summer, except at rather high latitudes. This tendency is most clearly revealed by the Asiatic winter and summer monsoons — the intense winter anticyclone centred over northern Asia and the equally great summer cyclone centred over southern Asia. The great Icelandic and Aleutian cyclones in winter also fit this pattern. In the southern hemisphere, where the continents are smaller, the tendency is present but less pronounced.

The temperature tends to be low over the continents and high over the oceans in winter, while the opposite tendency prevails in summer. Thus, in agreement with the thermal wind relation, the monsoons decrease in intensity with elevation. The temperature field is also influenced by the circulation itself, and the Icelandic and Aleutian cyclones are colder on their western sides, whence, again in agreement with the thermal wind equation, they are displaced westward with elevation. In the middle and upper troposphere there are two principal troughs in the westerly flow, located near the east coasts of North America and Asia. In the southern hemisphere the time-averaged flow is more nearly parallel to the latitude circles.

It might be supposed that features of the third category, like those of the second, would be absent in an idealized atmosphere, but numerical solutions reveal that pronounced time variations of $[u]$ and $[T]$ may occur, even though longitudinal variation of $\bar{u}$ and $\overline{T}$ do not occur. The distinction is that the time interval over which the averaging is performed is infinite, while the space interval is limited by the circumference of the Earth. In an idealized atmosphere circulating above a plane of infinite west-east extent, $[u]$ and $[T]$ would presumably not vary with time, while $\bar{u}$ and $\overline{T}$ would vary with longitude if the averages were taken over finite time intervals.

In the real atmosphere the best-known feature in the third category is probably the continual irregular oscillation between high-index and low-index patterns. The index of the zonal westerlies, or simply the zonal index, was originally defined by Rossby (1939) as the average geostrophic eastward wind component at sea-level between 35°N and 55°N. A typical low-index situation is however regarded as one where the westerly winds are not only weaker than average but are also displaced farther toward the Equator, while the northward and southward motions are unusually well developed. The opposite conditions characterize a typical high-index situation. Successive maxima or minima of the zonal index tend to occur at intervals of from two weeks to two months, but the intervals are not uniform enough for the index to show any periodicity. The zonal index serves as a fundamental quantity in some systems of extended-range forecasting.

Irregular fluctuations also characterize the easterlies. The trade winds are virtually always present, and vary only in intensity, but the polar easterlies are often replaced by westerlies. Indeed, there is usually neither an anticyclone nor a cyclone centred at the Pole, and a negative or positive value of $[u]$ near the pole is simply a measure of whether the vorticity is anticyclonic or cyclonic. As Mintz (1954) points out in his discussion of the long-term zonal circulation, the polar easterlies in the northern hemisphere have even been absent in averages over entire seasons; in essence they are a statistical residual.

In marked contrast to the non-periodic fluctuations in middle latitudes are the variations in the middle stratosphere in equatorial latitudes. During the past ten years the zonal winds have been observed to alternate between strong easterlies and strong westerlies, with a period of slightly more than two years. Moreover, superposed variations of shorter period are rather minor. The variations of u at different longitudes appear to be in phase, so that $[u]$ exhibits similar variations. The oscillation is most intense at the 30-mb level, and the various phases of the oscillation occur later at lower elevations. This so-called 26-month or quasi-biennial oscillation is generally regarded as a separate phenomenon from the irregular fluctuations which predominate in most of the atmosphere, and it has been the subject of numerous studies (see Reed 1965). Five or six complete cycles, which are certainly not perfect duplications of one another, do not prove that the equatorial stratospheric zonal wind will continue to oscillate in this manner. Assuming, however, that it continues to oscillate as it has ever since it was first regularly observed, it can be predicted two years or more in advance with considerable accuracy. Certainly the zonal index cannot be predicted a whole cycle in advance, much less two years, on the basis of its own past behaviour alone (see Namias 1950).

At the end of the spectrum are the fluctuations of the circulation which presumably accompanied the changes between glacial and interglacial epochs. Despite the evidence for major temperature variations, it is difficult to reconstruct the accompanying variations in the field of motion. It has nevertheless been surmised that the changes are very much like those characterizing the changes of the zonal index, with a low-index type of circulation prevailing during periods of extensive glaciation (see Willett 1949).

Features in the fourth category include many of those entities most familiar to the synoptic meteorologist. The most obvious ones are the migratory anticyclones and cyclones, including tropical cyclones. At upper levels large troughs and ridges, which give a wave-like appearance to the westerly wind current, frequently occur in preference to closed cyclones and anticyclones. While individual cyclones and anticyclones and individual troughs and ridges are usually regarded as secondary circulation systems, their frequency of occurrence, average intensity, and average daily displacement as functions of geographical location are properly regarded as additional characteristics of the general circulation. Likewise the wave number, or the number of principal troughs or ridges intersecting a given latitude circle, is a general-circulation characteristic.

Fronts and frontal surfaces form another feature in the fourth category. While frontal passages at individual locations are usually regarded as rather small-scale phenomena, the entire polar front — the principal discontinuity between air masses of tropical and polar origin — is often looked upon as a feature of the general circulation, and it has played a prominent role in some theories. Because its position is continually oscillating, no corresponding discontinuities appear in the time-averaged or longitude-averaged fields of motion and temperature.

Similarly the intertropical convergence zone — the principal zone of convergence between low-level currents originating in the southern and northern hemispheres, generally extending most of the way around the globe near the Equator — is logically regarded as a general-circulation feature. Since its position also oscillates, it appears as a fairly broad zone of transition rather than a narrow zone in averaged fields of motion.

A further feature which is not clearly evident in averaged temperature fields because its position is continually oscillating is the tropopause — the surface separating the stratosphere from the troposphere. The upward temperature decrease which prevails in the troposphere ordinarily ceases rather abruptly at the tropopause, but in the fields of $\overline{T}$ or $[T]$ it ceases more gradually. To obtain an abrupt tropopause in the field of $[\overline{T}]$ one would have to perform the averaging in a new co-ordinate system in which elevation above the tropopause replaced absolute elevation as a vertical co-ordinate.

The presence of a narrow meandering jet stream in either hemisphere, or often several jet streams, is not clearly represented by any single term in equation (80). The zonal westerly wind maximum which is such a prominent feature in the field of $[\bar{u}]$ is sometimes identified with the jet, but it does not possess the full average strength of the jet, nor does it occupy the proper average latitude. The fields of $\bar{u}$ at different longitudes possess maxima at different latitudes and elevations; if these maxima are averaged with respect to longitude, the result according to Crutcher's winter data is 38 m sec^{-1}, as opposed to a maximum of 34 m sec^{-1} for $[\bar{u}]$, while at longitude 140°E, just south of Japan, $\bar{u}$ reaches 65 m sec^{-1}. Likewise the fields of $[u]$ at different times possess maxima at different latitudes and elevations. But the full strength of the jet, often exceeding 100 m sec^{-1} at individual points, and the full amplitude of its meanders, are evident only in observations which have not been averaged.

CHAPTER IV

THE PROCESSES WHICH MAINTAIN THE CIRCULATION

In Chapter II we introduced the physical laws which govern the circulation of the atmosphere, and the dynamic equations which express these laws in mathematical form. In Chapter III we described some of the features of the circulation as revealed by observations, giving particular attention to the average fields of wind velocity, temperature, and water-vapour content. The application of the physical laws to the explanation of the observed circulation is the task which now confronts us.

The most direct way to accomplish this task would be to solve the dynamic equations. At present we lack a suitable means of solution. We must therefore proceed more indirectly.

In their usual form the dynamic equations enumerate the physical processes which directly affect any quantity. For example, the thermodynamic equation states that the temperature may be altered by advection, adiabatic compression or expansion, and net heating. It is sometimes possible to evaluate the long-term influence of each process affecting some feature of the circulation by recourse to the observational data. A knowledge of the magnitude of each process will not by itself constitute an explanation of the circulation, since it will not reveal why each process assumes the value which it does. Nevertheless, an understanding of the relative importance of the separate processes can be of considerable aid in formulating a qualitative explanation or in assessing the worth of any explanation which may have been offered.

The balance requirements

Consider the total mass of water contained in the region of the atmosphere north of a given latitude. This quantity may be temporarily increased by evaporation from the underlying Earth or decreased by precipitation falling to the Earth. It may also be increased or decreased by an inflow or outflow of moist air across the southern boundary. There is no necessity for the amount of evaporation taking place north of a given latitude to balance the amount of precipitation falling there, but, if the long-term average rates of evaporation and precipitation fail to balance, enough water must be transported within the atmosphere into or out of the region to balance the deficit or excess of evaporation. The need for this condition to be fulfilled constitutes the balance requirement for the transport of water in the atmosphere. It is a matter of convention that the evaporation and precipitation are regarded as determining a balance requirement for the transport rather than vice versa; nothing implies that one process is the cause while another is the effect.

It has been known for some time that evaporation exceeds precipitation in subtropical latitudes, while precipitation exceeds evaporation in middle and higher latitudes and also near the Equator. It follows that the motions of the atmosphere must be such as to transport water from the subtropics to lower and also to higher latitudes. A compensating return transport results from the combined effect of oceans, rivers, and flow within the ground.

A number of estimates of the average annual evaporation and precipitation at various latitudes are available. Figure 20 illustrates a recent Pole-to-Pole compilation of estimates by Sellers (1966). From the excess of precipitation over evaporation the required northward transport of water by the atmosphere has been derived, and is shown in Figure 21. The peak values in the tropics and in middle latitudes are the most prominent features. An interesting detail is the northward transport across the Equator needed to feed the intertropical convergence zone, whose mean position is somewhat to the north.

The precipitation curve in Figure 20 is based mainly upon direct measurements. These, however, are almost non-existent over the open ocean, while the amounts measured at island stations may differ considerably from the amounts falling over the oceans nearby. The evaporation curve is derived largely from the detailed work of Budyko (1956, 1963) and his collaborators. Over the oceans the evaporation has been computed from an empirical formula involving temperature, humidity, and wind speed, while over land areas it has been computed from precipitation and run-off. Much care has gone into the estimates of evaporation and precipitation, but in view of the many uncertainties we cannot yet regard them as the final word.

For computational purposes it is desirable to express the water balance in analytic form. If we integrate the general formula (10) over the volume of the region north of latitude φ_1, using (11), and then average over time, we find that

$$\int_0^\infty 2\pi r \cos\varphi_1 \, [\overline{\rho X v}] \, dz = -\int_0^\infty \int_{\varphi_1}^{\pi/2} 2\pi r^2 \cos\varphi \, [\overline{\rho dX/dt}] \, d\varphi \, dz \, , \tag{84}$$

where X may be any scalar quantity, $[\overline{\rho X v}]$ is measured at latitude φ, and it is assumed that there is no transport across the lower boundary by the motion of the atmosphere. An approximation to (84) obtained by integrating (47) over the mass of the region, using (48), is

$$\int_0^{p_0} 2\pi a \cos\varphi_1 \, [\overline{Xv}] \, g^{-1} \, dp = -\int_0^{p_0} \int_{\varphi_1}^{\pi/2} 2\pi a^2 \cos\varphi \, [\overline{dX/dt}] \, g^{-1} \, d\varphi \, dp \, . \tag{85}$$

In (84) the longitude and time averages denoted by brackets and bars are for fixed φ and z; in (85) they are for fixed φ and p. The latter form is more convenient for computation when the available data are at standard pressure-levels.

If q represents the mass of water per unit mass of air,

$$\int_0^{p_0} (dq/dt) \, g^{-1} \, d\, p = E_0 - P_0 \, , \tag{86}$$

where E_0 and P_0 denote the rates of evaporation from the Earth and precipitation upon the Earth. Equating X to q in (85), we obtain the water-balance equation

$$\int_0^{p_0} 2\pi a \cos\varphi_1 \, [\overline{qv}] \, g^{-1} \, dp = -\int_{\varphi_1}^{\pi} 2\pi a^2 \cos\varphi \, [\overline{E_0 - P_0}] \, d\varphi \, . \tag{87}$$

The left-hand side of (87) represents the total transport of water across latitude φ_1. The right-hand side has been used in constructing the transport curve in Figure 21 from the values of E_0 and P_0 in Figure 20. In most applications the horizontal transport of liquid and solid water is disregarded, and q is assumed to represent specific humidity.

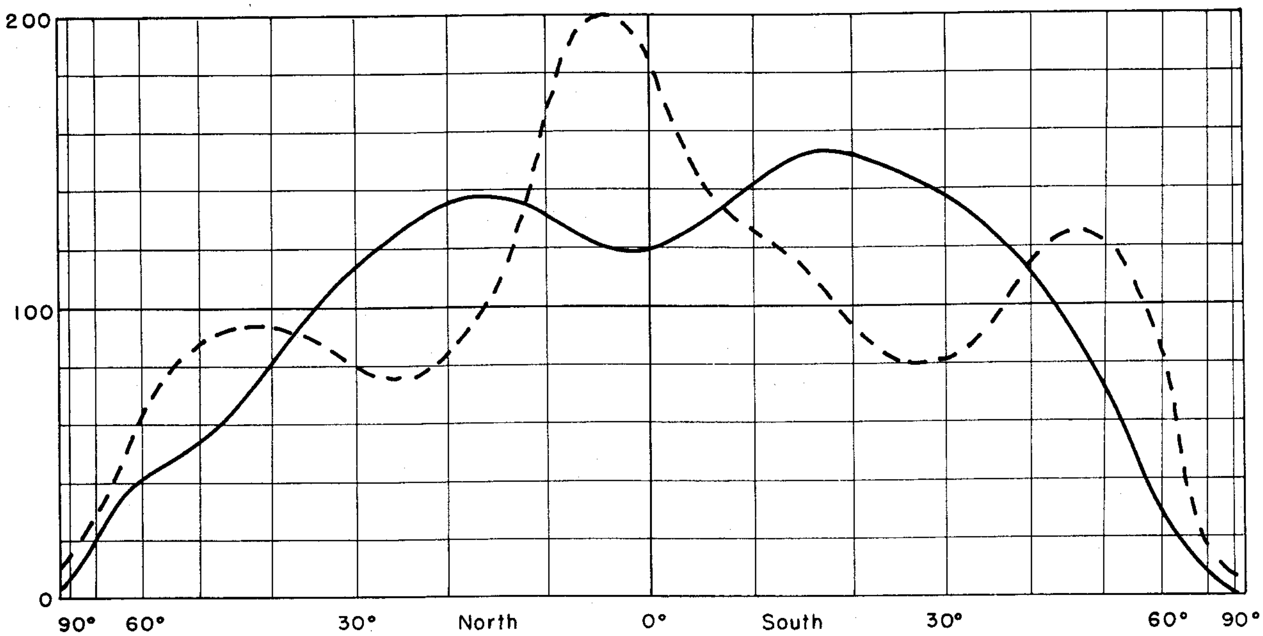


Figure 20. — Average annual evaporation (solid curve) and precipitation (dashed curve) per unit area as given by Sellers (1966). Values are in centimetres of water per year, or g cm^{-2} year $^{-1}$ (scale on left)

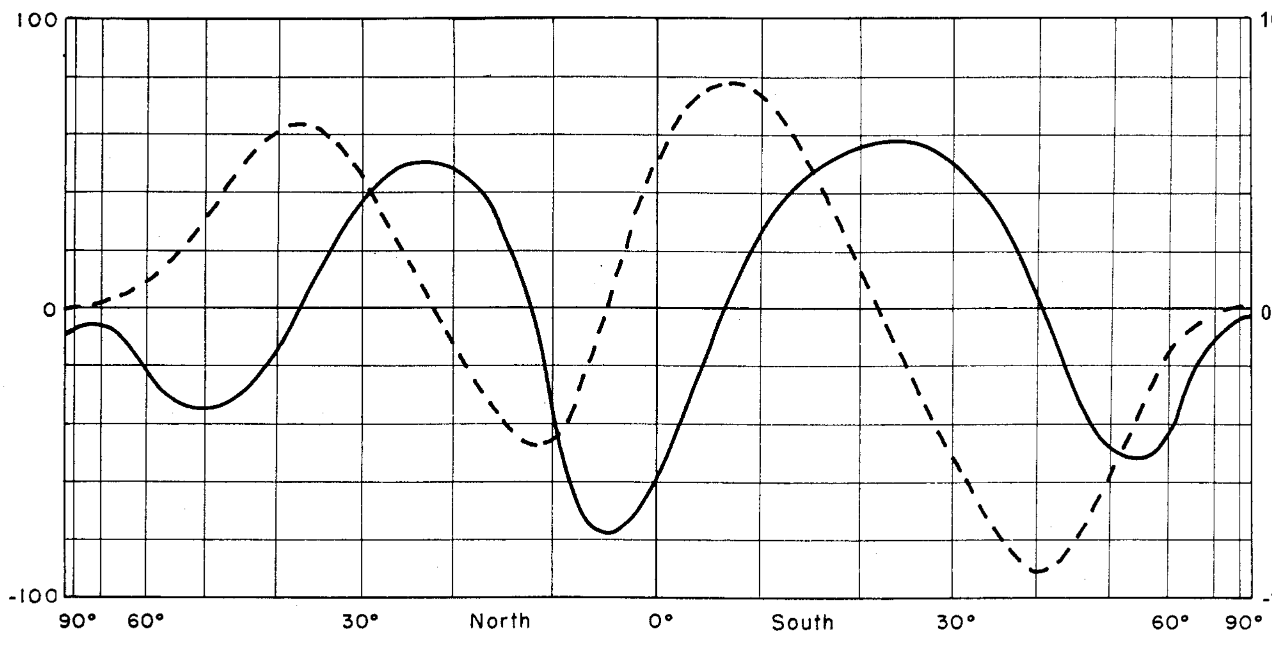


Figure 21. — Excess of evaporation over precipitation (solid curve) as given by Sellers (1966), in g cm^{-2} year $^{-1}$ (scale on left), and northward transport of water in the atmosphere required for balance (dashed curve) in units of 10^{11} g sec^{-1} (scale on right)

It is to be stressed that the evaporation and precipitation require a transport of water only in the sense that if they exist, the transport must exist. They cannot be regarded as the cause of the transport. If the circulation were unable to carry water into the equatorial zone, for example, there would simply be less precipitation or more evaporation there.

The atmosphere does not exchange dry air — the fraction of the air other than water in its various phases — with its environment in significant amounts. The balance requirement for dry air is therefore very simple; there can be no net long-term flow of dry air across any latitude. It follows that there is a net flow of air across each latitude, equal in mass to the net flow of atmospheric water. As much mass flows across latitude 40°S, for example, as would if there were a uniform north wind of 0.3 cm sec^{-1} at all levels. The fact that there is no net flow of dry air rather than no net total flow of air across each latitude is inconsequential for many purposes, but it must be recognized for a proper appreciation of the energy balance.

The balance of absolute angular momentum is analogous to that of water. Angular momentum may be exchanged between the atmosphere and the underlying surface, and it may be transported horizontally by the motion of the atmosphere. Although individual masses of air do not conserve their angular momentum even approximately, still the pressure torque within the atmosphere transfers angular momentum only from one longitude to another, while the frictional torque transfers it almost entirely from one elevation to another. Any net exchange of angular momentum with the underlying Earth by the region of the atmosphere north of a given latitude must therefore be balanced by a transport of angular momentum across that latitude.

As Hadley noted long ago, the atmosphere exerts a westward frictional drag upon the Earth in the latitudes of the trade winds, whence angular momentum is transferred to the atmosphere from the Earth. In middle latitudes where the westerlies prevail, angular momentum is returned to the Earth. There is an additional weak transfer to the atmosphere in the polar caps.

The frictional drag is often spoken of as if it were the only means for exchanging angular momentum between the atmosphere and the Earth, but another mechanism can operate wherever there are mountains, hills, or smaller irregularities. If there is a horizontal pressure difference across a mountain range, the air will effectively push the mountain, and the rest of the Earth with it, toward lower pressure; the mountain therefore pushes the air toward higher pressure. Although it seems natural to picture the air as piling up on the windward sides of mountains, and thereby augmenting the frictional torque, it is not obvious that this should be the case, since many mountain masses are so large that the pressure difference across them depends mainly upon the positions of migratory cyclones and anticyclones.

If X represents the angular momentum M in (84), we find from (17) and (25) that

$$\int_0^\infty 2\pi r^2 \cos\varphi_1 \, [\overline{\rho u v}]\, dz + \int_0^\infty 2\pi r^3 \,\Omega \cos^3 \varphi_1 \,[\overline{\rho v}]\, dz$$

$$= \int_{\varphi_1}^{\pi/2} 2\pi r^3 \cos^2\varphi \,[\overline{T_{0\lambda}}]\, d\varphi - \int_0^\infty \int_{\varphi_1}^{\pi/2} r^2 \cos\varphi \,(\Sigma \overline{p}_{\mathrm{E}} - \Sigma \overline{p}_{\mathrm{W}})\, d\varphi\, dz, \tag{88}$$

where Σp_{E} and Σp_{W} are the sums of the pressures on the east and west sides of the mountains or other sloping terrain intersecting the latitude circle in question, and $T_{0\lambda}$ is the eastward component of the frictional stress T_0 at the Earth's surface. The terms on the left represent the transports of relative angular momentum and Ω-momentum.

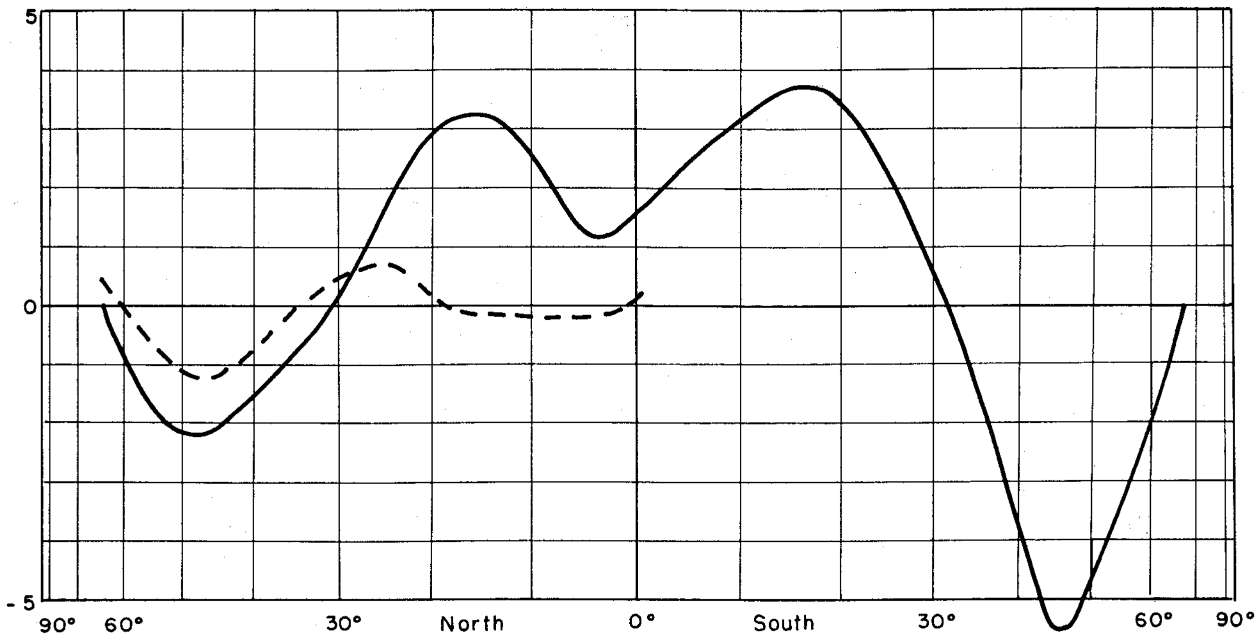


Figure 22. — Average eastward torque per unit area exerted upon the atmosphere by surface friction (solid curve) as estimated by Priestley (1951), and by mountains (dashed curve) as estimated by White (1949) north of 25°N, and Yeh and Chu (1958) south of 25°N. Unit is 10^8 g sec^{-2} (scale on left)

Figure 22 shows the frictional torque as determined by Priestley (1951), together with the mountain torque north of 25°N as determined by White (1949), and south of 25°N as given by Holopainen (1966) on the basis of a study by Yeh and Chu (1958). Priestley used the familiar empirical formula

$$T_{0\lambda} = C_D\, \rho\, [u^2 + v^2]^{1/2}\, u \tag{89}$$

to compute the frictional drag over the oceans, choosing the value 0.0013 for the dimensionless drag coefficient C_D. He then determined the torques which would follow if the ocean stresses were representative of entire latitude belts. He omitted the polar caps, where the torques should be small in any case. White based his computations upon normal pressure charts and simplified profiles of the principal mountain ranges.

Evidently the mountain torque is by no means negligible; in temperature and subtropical latitudes it tends to have the same sign and order of magnitude as the frictional torque. If this result holds also in the southern hemisphere, the effect of the mountain torque might be fairly well incorporated by increasing the drag coefficient C_D, whose proper value is not very well known in any case. Hutchings and Thompson (1962) have found that the relatively small New Zealand Alps add nearly 10 per cent to the total torque in their latitude band; perhaps the much higher Andes could nearly double the torque. Possibly there is a further contribution of the same sign from hills and other irregularities of intermediate size, whose effects are presumably not accounted for by Priestley's value of C_D.

It is well to note the implications of the apparent agreement between the mountain torque and the frictional torque. If the torques were of equal strength, and if the mountain torque did not agree in sign with the frictional torque, the total torque could not be inferred from the surface wind field. The theory of the general circulation on an idealized uniform Earth would then not be applicable to the real atmosphere.

Priestley observed that his middle-latitude torques were insufficient to balance his low-latitude torques, and noted that the situation could not be remedied simply by assuming a larger drag coefficient over land. He thereupon multiplied the middle-latitude torques by the factor 1.4 needed to achieve a balance. His amended torque, and the transport of angular momentum demanded by the balance requirements, are shown in Figure 23. The mountain torque is not explicitly included. The outstanding feature is the poleward transport across the subtropics in either hemisphere.

This transport must be almost entirely a transport of relative angular momentum. Although Ω-momentum is typically much greater than relative momentum, it is nearly constant at a fixed latitude, and in the absence of any appreciable net mass transport there can be no large transport of Ω-momentum. Actually, since Ω-momentum increases with elevation, a direct meridional cell will bring about some poleward transport, while the net flow of mass across certain latitudes requires an additional flow of Ω-momentum; these amounts may be computed from the second term on the left of equation (88). The intense Hadley cell determined by Palmén and Vuorela (1963), shown in Figure 18, yields an extreme poleward Ω-momentum transport of 2.2×10^{25} g cm^2 sec^{-2} across latitude 12°N in winter, but there is little transport across this latitude in summer. Moreover, the contribution by the Hadley cell is virtually cancelled by an equatorward flow of 1.4×10^{25} g cm^2 sec^{-2} across 12°N brought about by the net mass flow. The extreme poleward transports by the mass flow of 1.3×10^{25} and 1.7×10^{25} g cm^2 sec^{-2} occur near 35°N and 40°S, where the meridional cells are too weak to contribute appreciably. Comparison with Figure 23, where the required transports reach 20×10^{25} and 40×10^{25} g cm^2 sec^{-2} in the two hemispheres, suggests that most of the total transport is a transport of relative angular momentum.

There can be little doubt that the estimates in Figures 22 and 23 have the proper sign and order of magnitude, but by comparison with Figures 20 and 21 the actual values are poorly known. The mountain torque has received insufficient attention, but the uncertainty of the frictional torque is due largely to our inadequate knowledge of friction, particularly over irregular land areas.

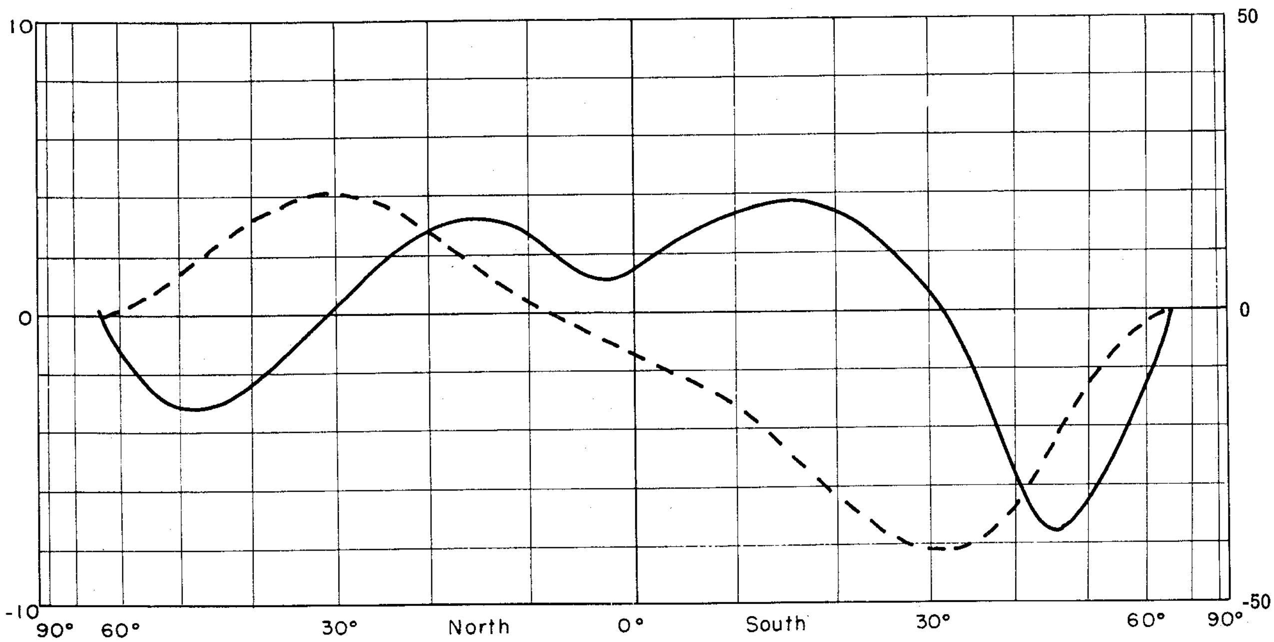


Figure 23. — Modified estimate by Priestley (1951) of average eastward torque exerted upon the atmosphere by surface friction (solid curve) in units of 10^8 g sec^{-2} (scale on left), and the northward transport of absolute angular momentum required for balance (dashed curve) in units of 10^{25} g cm^2 sec^{-2} (scale on right)

As in the case of the water balance, the torques are not a causal requirement for the transport of angular momentum. If the circulation were unable to transport angular momentum to middle latitudes, the surface westerlies there simply would not occur.

The balance of total energy presents a more complicated problem. Not only does the atmosphere exchange energy with the underlying Earth, but both the atmosphere and the underlying Earth gain energy from the sun and lose it to outer space through radiation. On this account it is desirable to examine first the energy balance of the entire atmosphere-ocean-Earth system, and then the more complicated energy balance of the atmosphere alone.

The incoming solar energy, which is the ultimate driving force for the atmospheric and oceanic circulations, is more intense in low than in high latitudes. Some of this energy is reflected or scattered back to space and plays no further role in the energy balance. The remainder is absorbed by the atmosphere and the Earth's surface; this portion, like the total, is more intense in low latitudes.

The energy re-radiated to space by the atmosphere and the Earth's surface is also more intense in low latitudes, although not so much more intense as one might expect in view of the higher temperature. Much of the outgoing radiation takes place from the uppermost layers of water vapour in the atmosphere; these extend to great heights in low latitudes and are therefore about as cold as the uppermost water vapour in higher latitudes. The net result is therefore a considerable excess of heating in low latitudes. It follows that there must be a poleward transport or transfer of energy across virtually every latitude. This transport may occur within the atmosphere or the oceans.

In contrast to the scarcity of numerical estimates of the angular-momentum exchange, there are numerous estimates of the incoming and outgoing radiation. Figure 24 is again based upon the values compiled by Sellers (1966) from a number of sources. The upper curve shows the solar energy reaching

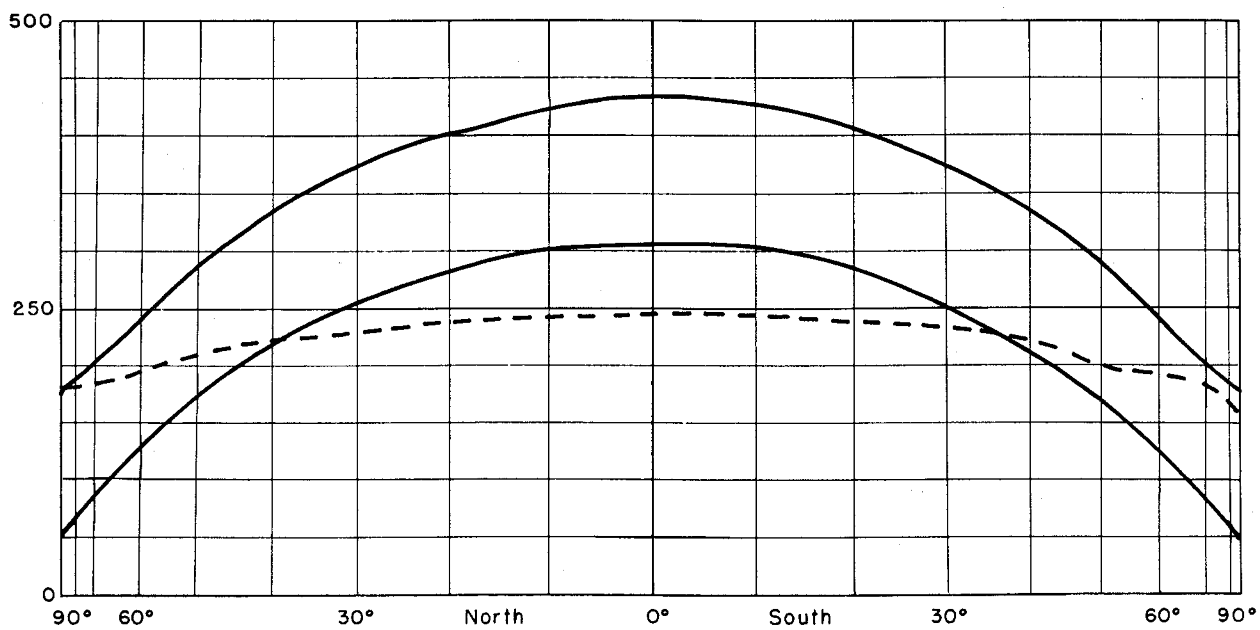


Figure 24. — Average solar energy reaching the extremity of the atmosphere (upper solid curve), average solar energy absorbed by the atmosphere-ocean-Earth system (lower solid curve), and average infra-red radiation leaving the atmosphere-ocean-Earth system (dashed curve), as given by Sellers (1968). Values are in watts m^{-2} (scale on left). (1 watt $m^{-2} = 1.435 \times 10^{-3}$ cal cm^{-2} min^{-1} = 0.754 kilolangleys $year^{-1}$.)

the extremity of the atmosphere; by comparing this amount with the amount absorbed, we see that the albedo, or the fraction reflected or scattered back to space, is about 30 per cent in low latitudes but exceeds 50 per cent in the polar regions. Figure 25 presents the net radiation and the required northward transport and transfer of energy. The principal features are again the peak values in middle latitudes.

The usual method of estimating the outgoing radiation requires rather involved computations of the emission in various wavelengths from the various atmospheric constituents, and individual estimates are far from being in complete agreement. Direct measurements have recently been made possible by the satellite. Figure 26 compares the values already shown in Figure 23 with the values obtained by Winston (1967) from one year of satellite observations. The calibration problems of the satellite-borne radiometers are not completely solved, and it would be premature to replace the conventional values by the new ones, but the general shape of the satellite curve is of considerable interest. It shows a pronounced relative minimum at 5° north, which actually appears near the Equator in every season. This is apparently due to the presence of the intertropical convergence zone, where the clouds and moisture ordinarily extend to great heights and hence radiate at a lower temperature than do the surrounding latitudes. Some of the pre-satellite estimates have indicated a similar although less pronounced minimum just north of the Equator.

Again the net heating cannot be regarded as the cause of the energy transport, except in the sense that heating is the ultimate cause of the entire atmospheric and oceanic circulations. If the circulations were unable to transport so much energy, the low latitudes would simply be warmer and the high latitudes would be colder, and the excess net heating in low latitudes would not be so great.

The energy balance of the atmosphere alone presents still further complications. Total energy need not include all forms of energy, but, having included one form, it must include all others which are directly or indirectly converted into this form in significant amounts. In the atmosphere and the underlying ocean and land the important forms are kinetic energy, potential energy, and internal energy, the last form including thermal internal energy and the latent energy of condensation and fusion of water. Moreover, in addition to being directly transported in each form, energy may be transferred horizontally by the pressure forces.

Before describing the atmospheric energy balance we must note that there is some ambiguity in defining the transport of energy by the atmosphere alone or the ocean alone; this ambiguity is the combined result of two circumstances. First, the zero marks on the scales for measuring internal and potential energy are somewhat arbitrary. Second, there is an exchange of mass between the atmosphere and the ocean, and hence a net flow of mass within the atmosphere and also within the ocean. The ambiguity may be removed by choosing zero marks, but different choices will lead to different pictures of the energy balance.

Ordinarily sea-level is chosen as the surface of zero potential energy, in which case the atmosphere will not exchange potential energy with the oceans, although the Earth will gain some potential energy when rain or snow falls at higher elevations. Likewise, absolute zero is generally chosen as the temperature of zero thermal internal energy, in which case the horizontal transfer of energy by internal pressure forces will be proportional to, and additional to, the transport of thermal internal energy, in the ratio R/c_v, or about 2/5. The sum of these quantities will then equal the transport of sensible heat, whose value per unit is $c_p T$.

Both the liquid and vapour phases of water have been chosen as the phases of zero latent energy. The former choice is the most frequent. In this event the oceans transport no latent energy, but the atmosphere transports large amounts poleward across middle latitudes. If the vapour phase is chosen

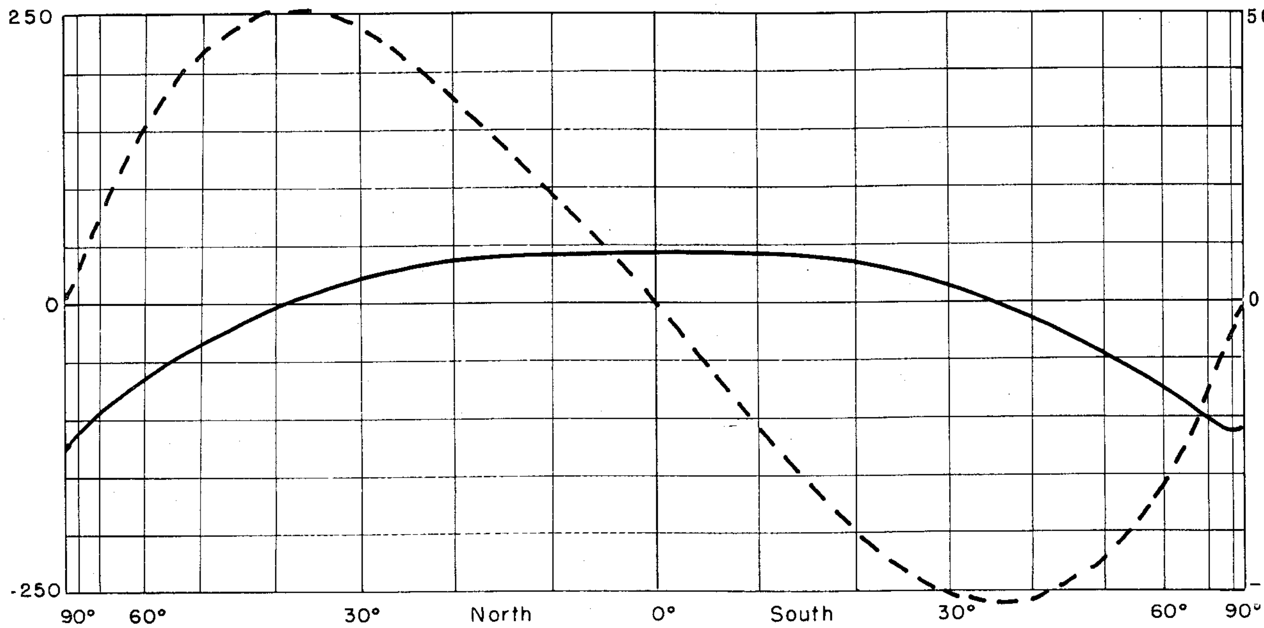


Figure 25. — Excess of absorbed solar radiation over outgoing infra-red radiation (solid curve), as given by Sellers (1966), in watts m^{-2} (scale on left); and northward transport of energy by the atmosphere and oceans required for balance (dashed curve), in units of 10^{14} watts (scale on right)

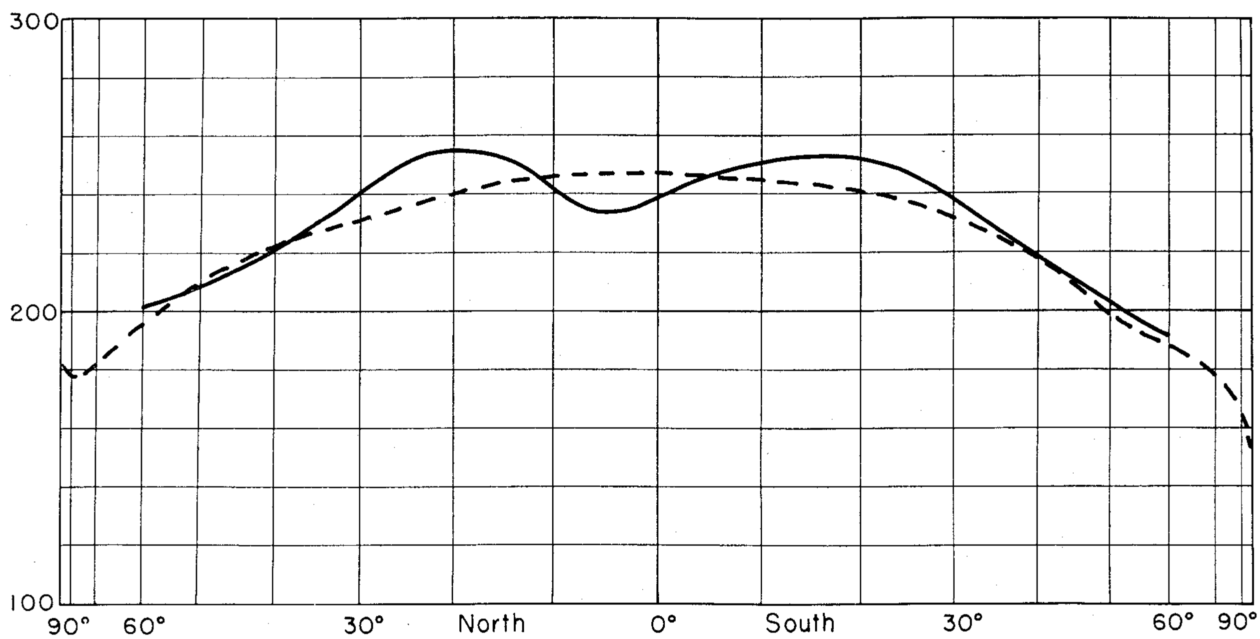


Figure 26. — Average infra-red radiation leaving the atmosphere-ocean-Earth system determined by conventional methods (dashed curve), as given by Sellers (1966); and determined from satellite measurements (solid curve), as given by Winston (1967). Values are in watts m^{-2} (scale on left)

instead, the oceans transport large amounts of negative latent energy equatorward across middle latitudes, i.e. they transport energy poleward, while the atmosphere transports only minor amounts.

Since the atmosphere must satisfy the balance requirement for total energy in either event, this requirement must also depend upon the choice of phase. If the liquid phase is chosen, the net radiation received or emitted by the atmosphere, the thermal internal energy transferred to the atmosphere from the Earth, and the latent energy supplied to the atmosphere by evaporation must together be balanced by transports of sensible heat, potential energy, kinetic energy, and latent energy within the atmosphere. If the gaseous phase is chosen, evaporation adds no energy to the atmosphere, but precipitation removes negative energy, i.e. adds energy.

If we set $X = K + \Phi + I$ in (84), we find from (24) after some arrangement of terms that

$$\int_0^\infty 2\pi r \cos\varphi_1 \, [\overline{\rho\,(K + \Phi + I)\, v + pv}] \; dz = -\int_0^\infty \int_{\varphi_1}^{\pi/2} 2\pi r^2 \cos\varphi \, [\overline{Q + \mathbf{V}\cdot\mathbf{F}}] \; d\varphi \, dz. \tag{90}$$

The term pv on the left represents the work done against a unit area of the southern boundary by the pressure forces. As already noted, for a dry atmosphere it would be proportional to, and additional to, the term $\rho I v$, which would represent the transport of internal energy. For the real atmosphere pv is still proportional to the transport of thermal internal energy.

In the strictest sense equation (90) is not applicable, because the mass transport across the Earth's surface was neglected in deriving it. However, it may be used if I includes both thermal and latent internal energy, and if the gain of latent energy resulting from evaporation from the surface is included in Q. Alternatively, I may include only thermal internal energy, and the release of latent heat by condensation may be included in Q.

Figure 27 shows the amounts of energy gained or lost by the atmosphere by various processes, including evaporation rather than precipitation. Again the values are those given by Sellers. The sum of these amounts and the required transport of energy in the atmosphere are shown in Figure 28. Again the transport presents a rather smooth curve, with peak values in middle latitudes. One may again question how accurately the various curves are really known.

The transport of latent energy in the atmosphere is for practical purposes proportional to the transport of water, which balances the excess of evaporation over precipitation. It follows by subtraction that the transport of the remaining forms of energy must balance the net radiation, the internal energy supplied from the Earth, and the latent energy released by condensation. This is the result which would have been deduced directly if the vapour phase of water had been chosen as reference.

The curves in Figure 29 are determined directly from the curves in Figures 21, 25, and 28. They show separately the required transports of thermal internal energy by the ocean currents, of sensible heat plus potential and kinetic energy by the atmospheric circulation, and of latent energy. The last quantity is regarded as being transported by the atmosphere under the usual convention, but would be considered to be transported mainly by the ocean under the less common convention that the vapour phase possesses zero latent energy.

The transport of internal energy by the ocean again conforms to the pattern of a single peak in each hemisphere, but a striking feature of the remaining transports is their relative irregularity as compared to the total transport. The two curves have steep slopes of opposite signs in the tropics. A simple explanation, which however requires verification, would be that the air flowing into the intertropical convergence

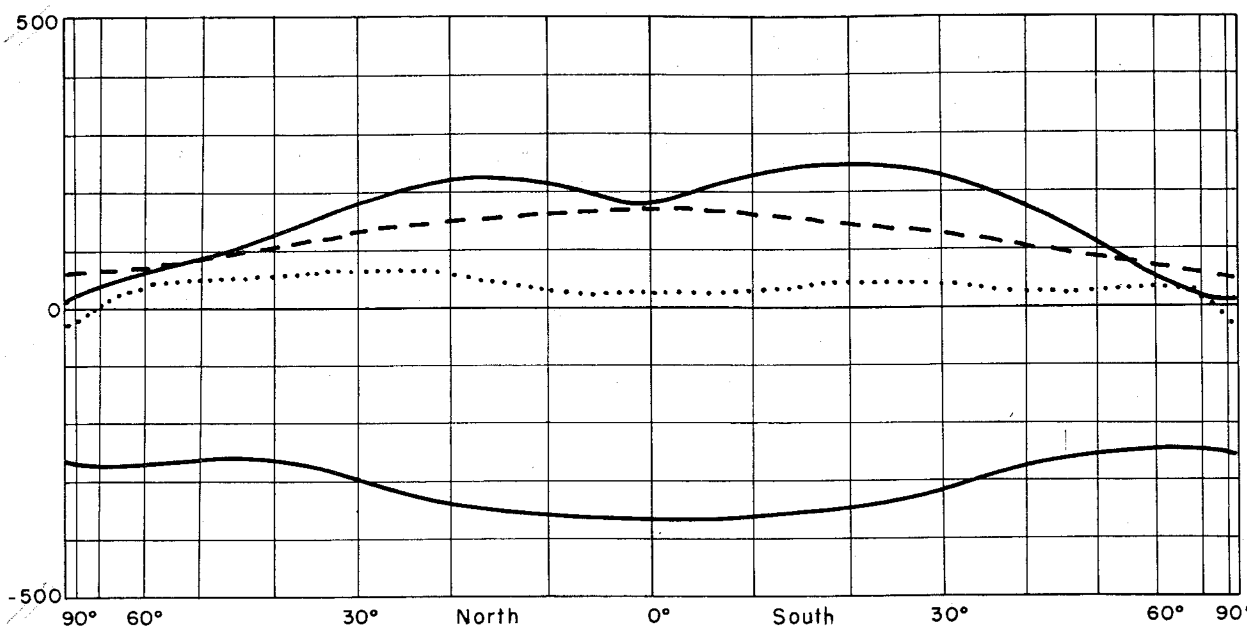


Figure 27. — Average short-wave radiation absorbed by the atmosphere (upper solid curve), average sensible heat received by the atmosphere from the underlying surface (dotted curve), average latent heat received by the atmosphere from the underlying surface (dashed curve), and average infra-red radiation leaving the atmosphere (lower solid curve), as given by Sellers (1966). Values are in watts m^{-2} (scale on left)

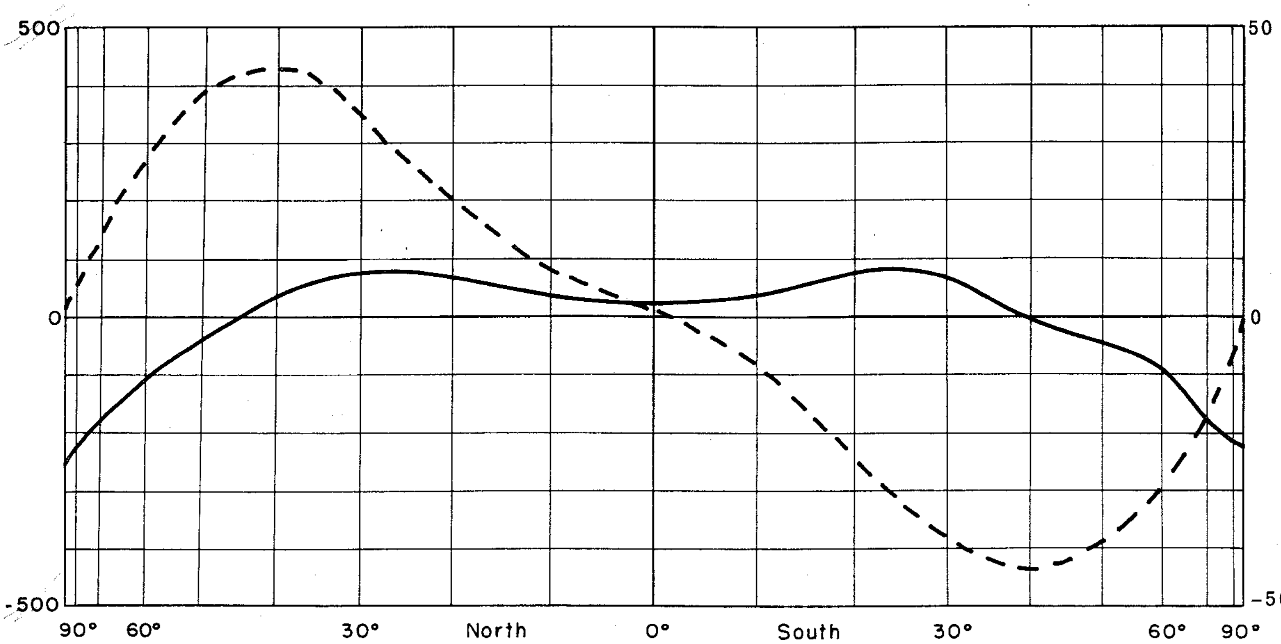


Figure 28. — The average net gain of energy from environment by the atmosphere (solid curve), as given by Sellers (1966) in watts m^{-2} (scale on left); and the northward transport of energy by the atmosphere required for balance (dashed curve), in units of 10^{14} watts (scale on right)

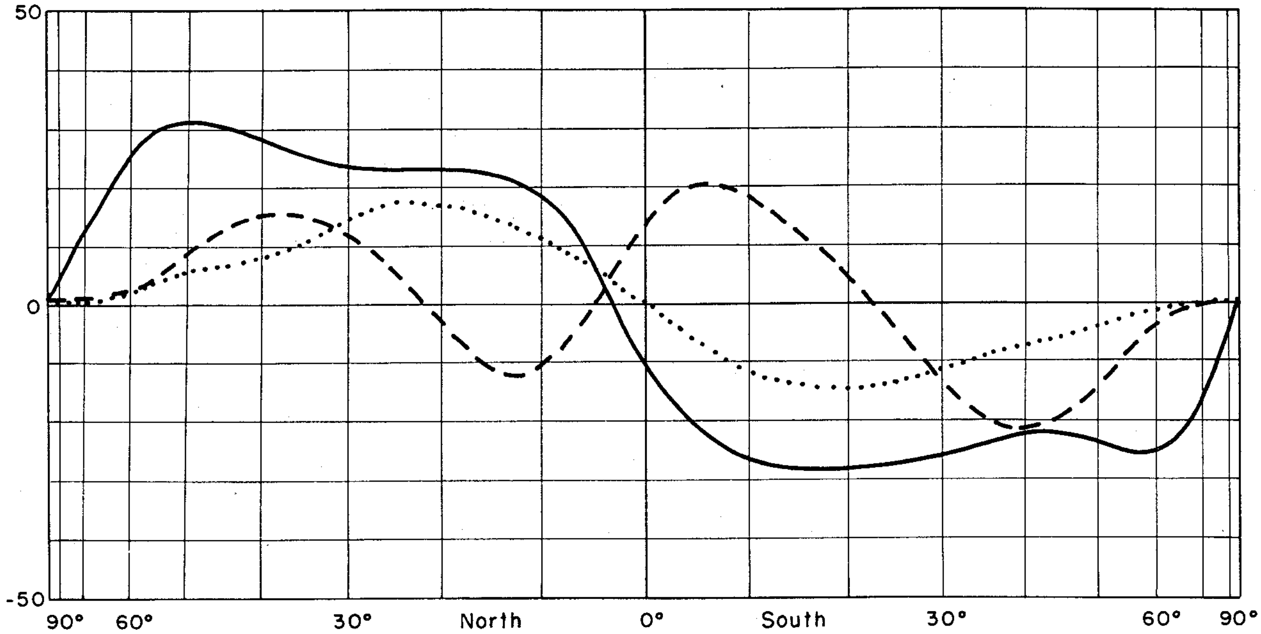


Figure 29. — The northward transport of sensible heat plus potential energy plus kinetic energy by the atmosphere (solid curve), of latent heat by the atmosphere (dashed curve), and of sensible heat by the oceans (dotted curve), required for balance, in units of 10^{14} watts (scale on left)

zone is loaded with latent energy; as the air rises and precipitation is released this latent energy is converted into thermal internal energy and potential energy, which is then transported out of the convergence zone.

It is to be stressed that the preceding curves show only the transports demanded by the observed exchanges of water, angular momentum, and energy between the atmosphere and its environment or between the system and its environment. Whereas the general features of these curves have been known for many years, it is only recently that observations have enabled us to evaluate the transports directly.

Former theories of the general circulation

Before considering the manner in which the balance requirements now appear to be fulfilled, we shall examine some of the views which prevailed in the past. So-called theories of the general circulation, whether they were real attempts to account for the circulation by dynamical theory, or merely descriptive schemes unaccompanied by explanations, appeared in abundance during the nineteenth and early twentieth centuries. Bergeron (1928) even remarked that there were as many theories as authors. We cannot discuss or even mention the great majority of these, but we shall attempt to identify those ideas which most greatly influenced the subsequent development of the subject, and which have led us to our present state of knowledge.

It is a relatively easy matter today to determine whether any newly proposed scheme of the general circulation agrees in its main features with observations, and to discard the scheme if it does not. In judging the worth of an older theory, we must therefore recall that much of what we now look upon as the observed circulation was unobserved as recently as World War II, and that at the close of the nineteenth century even such familiar entities as the stratosphere had not been discovered. Thus the main features of some of the former schemes were their speculations as to the circulation in the regions where observations were not available.

If the circulation were uniquely determined by the governing laws, any proposed scheme later found not to agree with the newer observations would necessarily violate some dynamical principle. We must therefore note that there may be many different circulations which satisfy the dynamic equations. Moreover, even if the external conditions should determine the circulation uniquely, considerably different circulations might be properly deduced from slightly different assumptions concerning the external conditions; this possibility has been cited by Bergeron (1928) as a contributing factor to the abundance of theories.

We should therefore regard a theory as a worthy contribution to the knowledge of its time if it contains no flaw in its dynamical reasoning, and if it is consistent with the observations available when it was formulated. A necessary condition for a theory to be dynamically sound and also compatible with observations is that it satisfy the balance requirements demanded by these observations. The condition is not sufficient; a proposed circulation may transport the proper amounts of angular momentum, water, and energy across each latitude and still be deficient in other respects. Noting this possibility, we may yet judge the worth of a proposed scheme partly by its ability to satisfy the balance requirements.

The circulation pictured by Hadley (1735), discussed in detail in the first chapter, satisfies the balance requirements demanded by observations which were then available, although not all of the requirements which more recent observations demand. The upper current carries as much mass poleward as the lower current carries equatorward. It carries more angular momentum, since the westerlies aloft are stronger. It also carries more sensible heat and internal energy, if the stratification is stable.

Hadley's scheme does not contain the weaker equatorward flow of angular momentum at high latitudes, but neither does it contain the polar easterlies which would demand it. Hadley did not consider water, but presumably his circulation would carry more water equatorward than poleward across every latitude, yielding the equatorial excess of precipitation and perhaps the deficit in the subtropics, but also giving a deficit in the polar regions, in contrast to what is observed.

Figure 30 illustrates Hadley's circulation schematically. Hadley himself presented no figure; we have introduced Figure 30 for comparison with the figures which have accompanied the numerous subsequent works.

In Hadley's picture the horizontal transports needed to satisfy the balance requirements are accomplished by the simplest possible mechanism — a meridional circulation where the uniform poleward current at one elevation carries a different amount of each transported property from the uniform equatorward current at another elevation. Since the atmosphere is not constricted to behave independently of longitude and time, other mechanisms are available. Whenever large-scale eddies such as cyclones are present, poleward flow at one longitude is accompanied by equatorward flow at the same time and elevation at another longitude. The oppositely directed currents may carry different amounts of any property. Theories of the general circulation therefore conform to one of two general schemes — those in which eddy motions are either absent or irrelevant, and those in which the eddies influence the zonally averaged motion by transporting some property from one latitude to another.

Following the publication of Hadley's paper there was a period of more than half a century during which the scientific community was generally unaware of its existence. Similar explanations were even rediscovered by such savants as Immanuel Kant (1756) and John Dalton (1793). Later Dalton encountered Hadley's paper, and, in acknowledging that his own contribution had been completely anticipated, noted

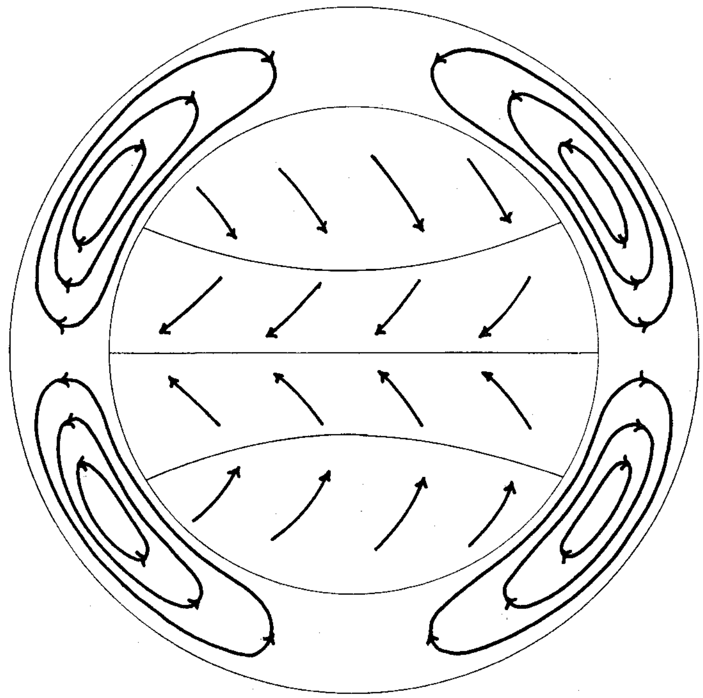


Figure 30. — A schematic representation of the general circulation of the atmosphere as envisioned by Hadley (1735)

the tendency for current works to continue to quote the older inadequate theories, while continuing to ignore the more recent and more acceptable ones. His remarks remain true today.

In due time, however, Hadley's explanation became the one which was quoted in the standard works. Early nineteenth century observations being what they were, there was little reason to doubt the explanation in its essential features.

As the nineteenth century progressed, observations began to cast some doubt upon Hadley's scheme. In particular, there was growing evidence that the prevailing westerlies in the northern hemisphere tended to blow from somewhat south of west, instead of from somewhat north of west as Hadley's explanation would have demanded. Undoubtedly the available data did not really justify such a conclusion, as they were confined largely to oceanic regions. Nevertheless the conclusion was evidently correct, as indicated by today's vastly more complete observations.

With the realization that the surface separating the trade winds from the south-westerlies above them sloped downward toward the north, and reached the Earth's surface in the horse latitudes, the notion became established that the middle-latitude south-westerly current at the surface was simply an extension of the current above the trades. The question then arose as to how the air returned from higher latitudes.

An answer was provided by the eminent meteorologist and climatologist Heinrich Wilhelm Dove (1837). Earlier (1835) Dove had been one of those to rediscover Hadley's explanation of the trades, again under the assumption that absolute velocity rather than absolute angular momentum would be conserved in the absence of east-west forces. He now accepted Hadley's ideas completely as far as low latitudes were concerned, but he favoured the prevailing idea that the south-westerly winds in middle latitudes were a continuation of the south-westerlies above the trades, since he felt that their warmth and humidity demanded an equatorial origin. At that time it was not generally realized that air rising to high levels and sinking again would have to lose most of its moisture.

It followed naturally that the trades themselves should be a continuation of a return current from higher latitudes. Dove rejected the possibility that this current could occur at upper levels, since it appeared impossible for oppositely directed currents to cross in the horse latitudes without altering one another. He was thereby led to a scheme where south-westerly and north-easterly winds in middle latitudes flowed side by side at different longitudes at the same level, rather than one above the other. His warm moist equatorial current was fed by the south-westerlies above the trades while his cold dry polar current fed the trade winds. The equatorial current preferred the oceans and the west coasts of continents, while the polar current preferred the interiors and east coasts, but the longitudes of the currents were not fixed, and familiar local weather changes accompanied the replacement of one current by the other. He could explain the net northward flow, volumewise at least, by the greater specific volume of the equatorial current, but he also felt that it might be largely fictitious, since the observed northward flow could be compensated by southward flow over the interiors of continents, where observations were less abundant.

He even maintained that there were only two winds in middle latitudes — the north and the south — other directions being simply variants. Thus his polar and equatorial currents seem to be none other than what we now call polar and tropical air masses, which he chose to identify by their preferred motion rather than their quasi-conservative thermodynamic properties. He furthermore regarded the middle-latitude storms as resulting from the conflict of the two currents. His circulation is shown schematically in Figure 31.

Dove's scheme can certainly satisfy the energy balance requirements, since the equatorial current is warmer than the polar current. It can satisfy the momentum balance requirements, since the south-westerly winds carry more momentum than the north-easterlies. Under the assumption that the equatorial current cannot carry its water aloft at low latitudes, as Dove had supposed, but must lose it and then reacquire it from the ocean after descending, the scheme can satisfy the balance requirements for water.

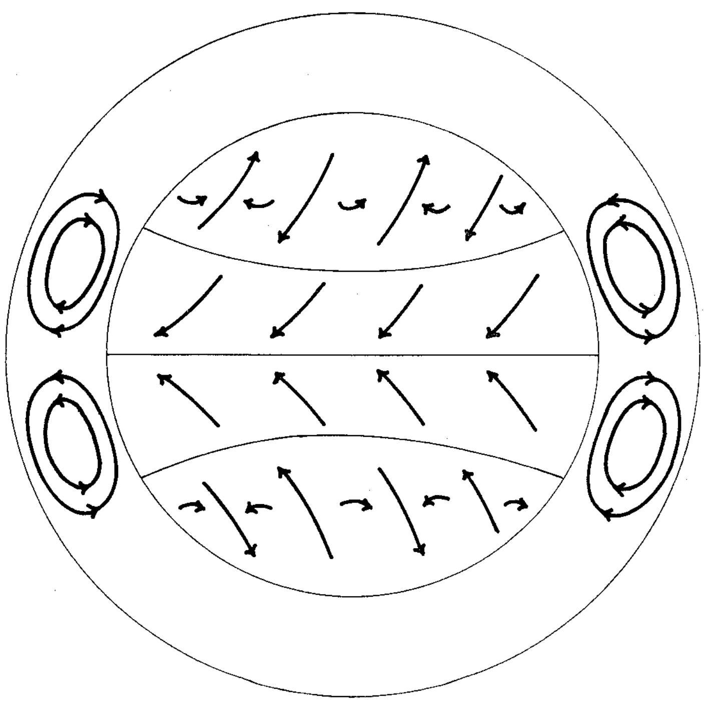


Figure 31. — A schematic representation of the general circulation of the atmosphere as envisioneed by Dove (1837)

Yet it does not have the appeal of Hadley's scheme. As a keen observer rather than a theoretician, Dove offered no tidy explanation for the circulation which he so carefully recorded. His arguments involving the Earth's rotation are more applicable to Hadley's circulation than to his own.

Among the most diligent compilers of weather observations was the naval officer Matthew Maury, whose charts of the winds over the oceans had led to considerable reductions in the normal sailing times between distant points. In 1855 he came forth with his own scheme of the general circulation, which departed considerably from Hadley's, and contained precisely the features which Dove had rejected some years before. It is shown in Figure 32. Instead of the single meridional cell in either hemisphere, or opposing currents side by side, there are two cells — a direct cell like Hadley's within the tropics, and an indirect cell in higher latitudes. The flow above the north-east trades is from the south-west, and the upper-level flow at higher latitudes is apparently supposed to be from the north-east.

Like Dove, Maury used no mathematical formulae, but he was extremely conscious of the balance requirements for water. A distinctive feature of his scheme was the crossing of the meridional currents as they sank in the horse latitudes and also as they rose in the doldrums, and he devoted great efforts to justifying these crossings. He was a great believer in the Grand Design, and he rejected the possibility that the converging currents would mingle and then depart, now in the direction from which they came, and now in the opposite direction, on the grounds that the circulation could not be left so completely to chance. He could see no reason why the currents must cross instead of returning, but he insisted that the lack of balance between precipitation and evaporation in low latitudes and also in high latitudes indicated that they did cross. Like Dove, he was unaware that a high-level current cannot retain its moisture. He believed that crossing without mixing could occur by having vertical columns of air pass one another; his envisioned columns seem to have the dimensions of cumulonimbus towers.

Maury was unable however to offer an explanation for what seems now to be the principal feature of his scheme — the indirect cells in higher latitudes. He accepted Hadley's explanation for the trade-wind cells, and simply said that the cause of the indirect cells had not been explained by philosophers.

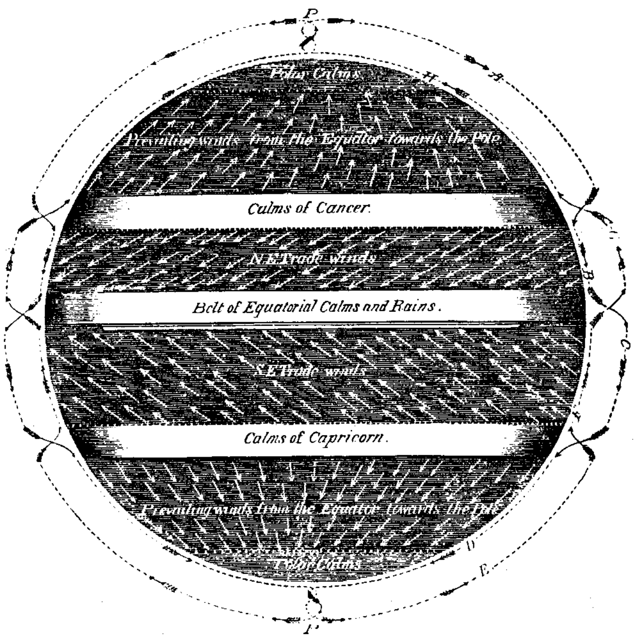


Figure 32. — The general circulation of the atmosphere according to Maury (1855)

Maury's scheme seems to satisfy the balance requirements for angular momentum, in view of the assumed crossing currents in the horse latitudes and the upper-level easterlies in higher latitudes, which, however, are inconsistent with modern observations. It certainly cannot satisfy the energy balance requirements, except for an atmosphere which is heated at the Equator and the Poles and cooled in between. Nevertheless, Maury's book is extremely readable. It became rather popular in his day, and it was instrumental in initiating some of the more rational theories which were to follow.

Among those who read Maury's book was the school teacher William Ferrel, whose interest in the subject was thereby aroused. Here he first learned that the normal pressure was not uniform over the Earth's surface, but highest in the horse latitudes, and lower in the doldrums and especially in the polar regions. He found that he disagreed with some of Maury's ideas, particularly the crossings of the meridional currents, which he felt ought to mix rather than cross. In the following year (1856) he came forth with a theory of his own.

The circulation which he envisioned is shown in Figure 33. It is somewhat like Maury's, except that there are now three cells in either hemisphere, which he felt were demanded by the observations. Unlike Maury, however, Ferrel believed that he could present a complete explanation.

Ferrel's great contribution in this paper was the introduction of a "new" force, the north-south component of the Coriolis force, which he incidentally identified with one of the terms in Laplace's tidal equations, formulated long before Coriolis, and which he believed had not been previously recognized in any meteorological work. Actually he had been anticipated in an unnoticed paper by Tracy (1843), who with inadequate arguments nevertheless deduced the proper direction of the deflection. Ferrel believed that proper consideration of the new force would account for the previously unexplained features not only of the general circulation but of cyclones and smaller disturbances as well.

Ferrel agreed with Hadley that the prime moving force of the atmospheric circulation was the Pole-to-Equator density gradient brought about by solar heating, which he believed should lead to meridional

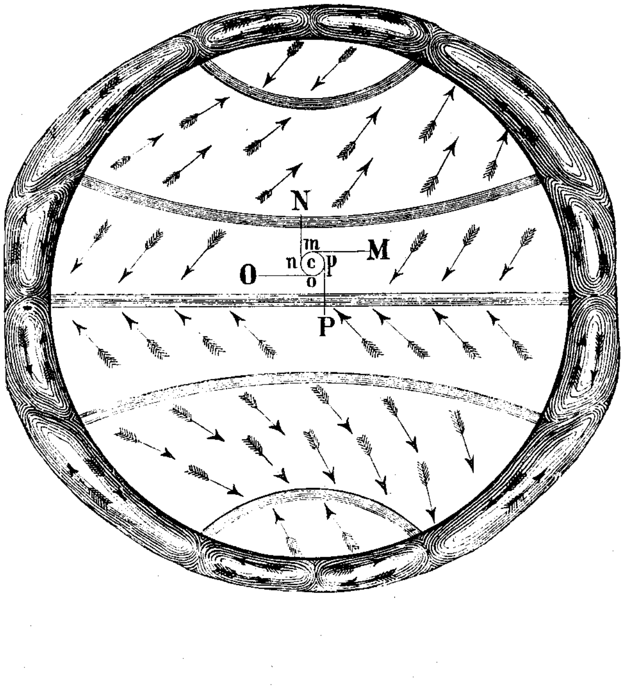


Figure 33. — The general circulation of the atmosphere according to Ferrel (1856)

motions, and hence, through the action of the east-west Coriolis force, to easterly and westerly winds distributed much as Hadley had supposed; but he did not find in Hadley's theory any explanation for the distribution of pressure. He then noted that through the action of the new force the easterlies in low latitudes and the westerlies in higher latitudes should be deflected away from the Equator and the Poles toward the subtropics, thereby creating the observed deficit of pressure at the Equator and the Poles and the excess in the subtropics. To explain the poleward drift in the surface westerlies, he observed that because of surface friction, the winds near the ground would be considerably weaker than the winds somewhat higher up, while the horizontal pressure gradient would be reasonably uniform. The southward Coriolis force near the ground would therefore be insufficient to balance the pressure gradient, and the westerlies would turn poleward, later to rise and return equatorward.

Ferrel also noted that for hydrostatic reasons the latitude of highest pressure must be displaced toward the Equator with elevation. He apparently felt that the opposing currents aloft must meet at this latitude in order to maintain the high pressure, whence he showed inclined boundaries between the low-latitude and middle-latitude cells.

There are certain obvious deficiences in Ferrel's scheme, as well as in his explanation of it. The indirect cells in middle latitudes must transport angular momentum and energy toward the Equator, and neither balance requirement can be satisfied. The middle-latitude westerlies aloft were originally supposed to be maintained by the action of the Coriolis force upon the poleward currents, but now these currents have been replaced by equatorward currents while the westerlies are allowed to remain.

Nevertheless it would be difficult to overestimate the importance of Ferrel's paper. Here he first presented to the meteorological world a correct account of the Coriolis force, a quantitative description of the geostrophic wind, and a partial explanation for its occurrence. His demonstration that the pressure field could adjust itself to fit the wind field, rather than forcing the wind to do all of the adjusting, has often been overlooked by succeeding generations of meteorologists.

Another scientist who read Maury's book was the physicist and inventor James Thomson, who found himself in considerable disagreement with some of Maury's ideas. Thomson was understandably unaware of Ferrel's work, which had been published in a local medical journal, but he had attended a lecture delivered by Murphy (1856), who had also read Maury's book and had suggested that the low pressure at the Poles resulted from the centrifugal force of the westerly currents, which could be treated as large circumpolar vortices. Thomson soon produced his own scheme (1857), which is shown in Figure 34.

After noting Hadley's error concerning the conservation of angular velocity, he otherwise accepted Hadley's arguments with regard to the bulk of the atmosphere, but maintained that the westerly winds near the ground, being slowed by friction, would possess a deficit of centrifugal force relative to the stronger westerlies immediately above, and would therefore drift poleward. In this respect his argument is the same as Ferrel's, differently worded. The southward or northward component of the Coriolis force, as Ferrel pointed out, is simply the excess or deficit of centrifugal force as compared to the centrifugal force of a particle rotating with the Earth. The excess Coriolis force of a rapid west wind over that of a slow west wind is therefore the same thing as the excess centrifugal force of the rapid wind over the slow wind.

Just as there was little observational evidence in Hadley's day to contradict his scheme, so there was little evidence in Thomson's day to contradict his. Thomson's scheme is admirable for its simplicity, and it also satisfies the balance requirements. The indirect cell is confined to such a shallow layer that it transports very little angular momentum or energy. At the same time it can produce the needed poleward transport of water, since the water-vapour content decreases so rapidly with elevation.

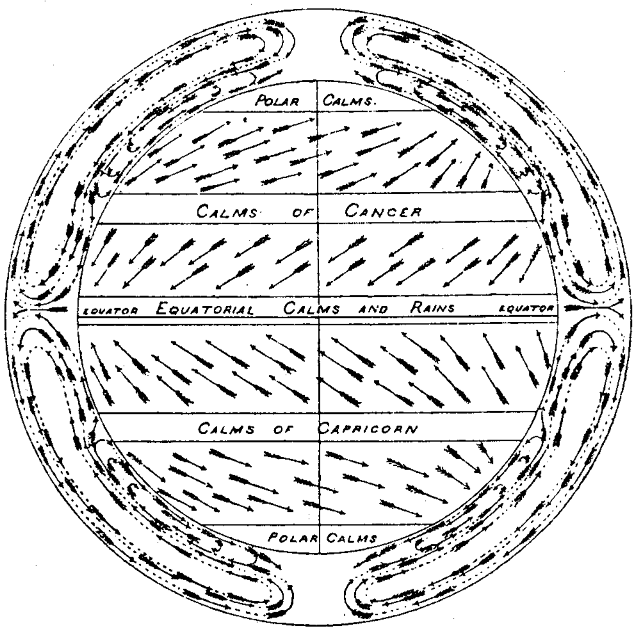


Figure 34. — The general circulation of the atmosphere according to Thomson (1857, 1892)

Thomson's published contribution was limited to an abstract, but years later (1892) in his Bakerian Lecture, delivered two months before his death, he returned to the problem of the atmosphere, and reiterated his belief in his former scheme. He also discussed critically the other schemes which had been offered, and noted the difficulties entailed by Maury's and Ferrel's converging currents aloft.

Ferrel was, however, a relative newcomer to the field, and his ideas were anything but static. He set about formulating his work in mathematical terms, and as result he came up with a revised scheme (1859). It is shown in Figure 35. It is very much like Thomson's, except in the polar regions where definitive observations were lacking in any case. It is therefore equally effective in fulfilling the balance requirements. His paper contains the complete equations of motion for the atmosphere, and an account of the thermal wind relation.

In justifying his scheme, he maintained that if surface friction were absent, while internal friction still existed, the atmosphere would assume a condition of uniform absolute angular momentum. It is often pointed out that such a condition would lead to unrealistically violent winds at high latitudes, but Ferrel went a step beyond his successors and noted that the accompanying pressure gradients required by geostrophic balance would leave the polar regions completely devoid of air. In his computations he had treated the atmosphere as a liquid; as a gas the only real singularities would be at the Poles, but even between 30° and 60° latitude the pressure would drop by a factor of three.

Ferrel maintained that with surface friction the atmosphere would tend toward the same distribution, but to a much lesser extent, the latitudes separating the easterly and westerly surface winds being ultimately determined by the requirement of no net surface torque. He thus purported to account for the observed distribution of zonal wind. He explained the poleward drift of the surface westerlies as in his earlier paper, observing that this drift required a return current somewhere aloft. He noted, however, that in view of the thermal wind relation, the upper-level westerlies must be stronger than the

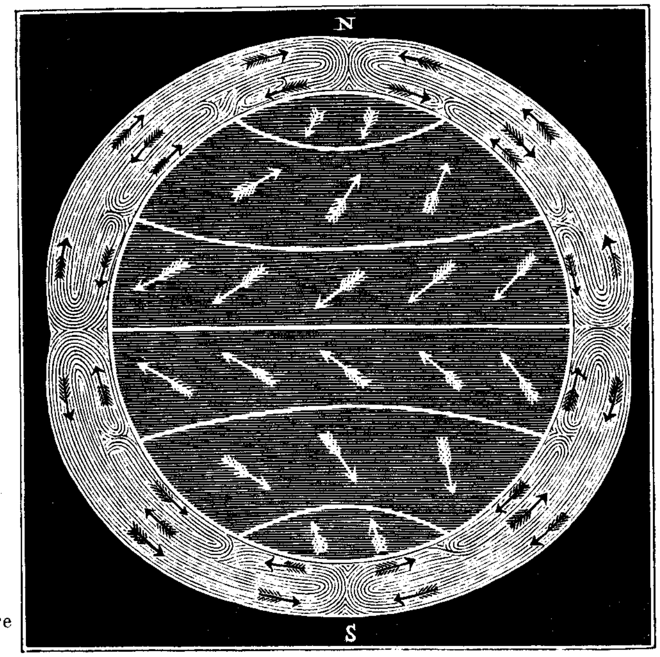


Figure 35. — The general circulation of the atmosphere according to Ferrel (1859)

surface westerlies and must be maintained by the action of the Coriolis force upon a poleward current. He therefore placed the equatorward current at an intermediate level, noting that according to observations it should lie above the fair-weather clouds.

We cannot agree with Ferrel's premise that with internal viscosity but without surface friction the atmosphere would tend to acquire a state of uniform absolute angular momentum. Such a circulation would possess strong internal stresses. Neither does it seem very likely that the ultimate circulation with surface friction would be an attenuated form of the circulation without friction, despite Ferrel's observation that there can be no resistance to motion until there is motion. Whereas Thomson by-passed an explanation of the surface easterlies and westerlies by simply agreeing with Hadley, Ferrel's attempt in this respect yielded no improvement. Beyond this point, Ferrel presented some penetrating arguments, and he used the thermal wind relation to good advantage.

Ferrel's subsequent work led to further modifications, his final scheme (1889) differing slightly from his second one. He was intrigued by the possibility of deriving mathematical expressions for the circulation, but felt that this could not be done because the frictional forces could not be properly formulated. He continually maintained that the circulation must be derived from a knowledge of the temperature field, rather than the field of solar heating, and his system of equations does not contain the thermodynamic equation. It was a great loss to nineteenth-century meteorology that the man who introduced the equations of motion never saw fit to seek a complete solution of them.

The task which Ferrel had regarded as unfeasible was finally attempted in a pair of papers by Oberbeck (1888), who represented the effects of friction by a simple coefficient of viscosity. Like Ferrel, Oberbeck sought to derive the motion from the temperature field, and he did not use the thermodynamic equation. He represented the temperature by a simple analytic function of latitude and elevation.

Oberbeck sought first the circulation which would prevail in the absence of rotation and advection, and the set of equations which he first solved expressed a balance between the effects of friction and the pressure forces. The circulation which he obtained was necessarily entirely meridional, and consisted of a single direct cell. To obtain the next approximation he balanced the east-west Coriolis force, as determined by his first approximation, against friction. The added circulation was entirely zonal and proportional to the Earth's angular velocity Ω, and consisted of low-latitude easterlies and high-latitude westerlies at low levels, and westerlies at all latitudes at high levels.

On the whole his circulation bears considerable resemblance to Hadley's. We feel, however, that this resemblance is fortuitous. In a steady symmetric circulation, the Coriolis force resulting from the net north-south motion in any vertical column is zero. Hence Oberbeck was balancing the frictional drag at the base of each column against the net Coriolis force resulting from the weak vertical currents, and he thus obtained easterlies and westerlies just above the surface, in their proper latitude. In a mathematical description of Hadley's circulation, the frictional drag is balanced by non-linear terms, which Oberbeck had not used at this point.

In his second paper Oberbeck sought the final corrections needed for an exact solution, but since the system of equations governing these corrections was as complicated as the original system, containing all the non-linear terms, he found it necessary to make further approximations. The added circulation was again entirely meridional, and proportional to Ω^2, and consisted of a direct cell in low latitudes and an indirect cell in high latitudes. In essence he had found the first three terms of a power series in Ω. For the value of Ω appropriate to the Earth, the added cell in high latitudes was insufficient to reverse the direction of the original cell, and simply weakened it there while intensifying it in low latitudes.

It is no discredit to Oberbeck that he was forced to stop with the quadratic terms in Ω, yet it must be conceded that on this account his solution is not a particularly good approximation to the exact solution which he sought. His increase of westerly wind speed with height is proportional to Ω, whereas, according to the thermal wind relation, it should be inversely proportional to Ω. It is not at all certain that Oberbeck could have improved his results by computing more terms, since, as noted by Brillouin (1900), there is no assurance that the series would converge. A power-series expansion does not reveal that $\Omega/(1 + \Omega^2)$, for example, becomes small as Ω becomes large.

Oberbeck's work marks the beginning of a new field of endeavour — representing the global circulation by solutions of the dynamic equations, as opposed to using the equations simply to deduce general properties. A mathematical solution is simply one type of description of the circulation, but its advantages are obvious. If the equations have been correctly formulated and correctly solved, with no crippling approximations, the description is assured of being internally consistent in every way, and in particular it will satisfy its own balance requirements.

If more theoretical meteorologists had followed Oberbeck, and had sought actual solutions of the dynamic equations in preference to circulations which could merely be rendered plausible by qualitative arguments based upon the dynamic equations, many of the impossible schemes which were subsequently offered might never have appeared. Still, the idea that manipulation of mathematical symbols ought to replace qualitative reasoning could scarcely have appealed to the many competent meteorologists who were nevertheless not mathematically inclined. When further attempts to solve the equations yielded circulations which were no more realistic than Oberbeck's, this was cited as evidence that the whole procedure was meaningless. The fact that the equations had not really been solved was disregarded.

It must be admitted that even very recent analytic solutions of the equations have had a certain unrealistic flavour. It is only with the advent of numerical solutions by digital computers that the equations have begun to acquire the status which they deserve.

In Oberbeck's work, as in that of most of his predecessors with the notable exception of Dove, the general circulation was treated as being completely symmetric with respect to the Earth's axis. It must not be supposed on this account that the various authors were unaware of the presence of cyclones and other disturbances. Both Ferrel and Oberbeck were deeply concerned with the cyclone problem, and Ferrel often dealt with the general circulation and cyclones in separate sections of the same papers, regarding the cyclone circulations as being much like the general circulation on a smaller scale. Yet nowhere in Ferrel's work is there any suggestion that cyclones owe their origin or subsequent behaviour to the general circulation, or that the general circulation in turn is affected by the presence of cyclones.

The idea that storms were dependent upon the general circulation had been proposed long before Ferrel's time, and it formed an essential part of Dove's work. In modern studies where the field of motion has been analysed into "zonal" and "eddy" components, there is often a tendency to regard all departures from zonal symmetry as having a similar nature, and to refer to them loosely as storms. As an observer of weather phenomena rather than a formalist, Dove distinguished between cyclonic storms on the one hand and the equatorial and polar currents on the other, regarding the storms as originating from disturbances of the opposing currents. It would have meant nothing to him to inquire whether these currents influenced the general circulation; to him they were the general circulation. Indeed, such a question is meaningful only if the general circulation has been defined. A less ambiguous question would ask whether the zonally averaged motions are different from what they would be if departures from zonal symmetry were absent. Dove might have answered this question in the affirmative.

It is remarkable that Dove's rather advanced description of the circulation, which would have been dynamically possible in a dry atmosphere, went completely unmentioned by many subsequent writers (for example Brillouin, 1900) who included thorough treatments of the works of Maury, Ferrel, Thomson, Oberbeck and others in their historical discussions of general-circulation theories. Among those who did mention Dove, Waldo (1893), while presenting extensive reviews of the other works, merely states that Dove made certain modifications of Hadley's theory; he does not even say what these modifications were.

Perhaps the neglect of Dove's work may be traced to his refusal in his later years to accept any of the newer ideas, with the result that all of his work tended to become discredited. Perhaps his work was ignored because he offered only descriptions rather than explanations, although the same criticism could be made of Maury. It seems very likely, however, that most of the writers of the later nineteenth century simply did not consider that the motion of which Dove spoke was the general circulation. The notion that the general circulation meant the time-averaged or the time-and-longitude-averaged circulation had become rather well established, and Dove's currents varied with time and longitude. It is noteworthy that Hann (1901), in what is still one of the most comprehensive meteorological treatises yet produced, makes no mention of Dove in his chapter on the general circulation, but reviews Dove's ideas in detail in the following chapter on storms.

Yet despite the apparent rejection of Dove's scheme, the symmetric theories of the general circulation could not endure forever. Even as Ferrel and Thomson were making their final contributions, specific objections to them were being raised.

One of these was based mainly upon theoretical considerations. Except near the Earth's surface, friction was generally considered to be negligible. It was often pointed out that in the schemes of Ferrel and Thomson the poleward-moving air aloft would acquire unheard-of velocities in middle latitudes, in conserving its absolute angular momentum. Some writers took the attitude that such high velocities, never being observed, could not possibly exist, and that the schemes were dynamically impossible.

Others took the more moderate attitude that the high velocities simply did not exist, and that the schemes, while perhaps possible, were incorrect. In any event the thermal wind relation, adhered to by Ferrel, would not allow such high velocities to occur aloft unless excessively high velocities occurred at the Earth's surface also. The simple scheme of having the poleward current aloft so weak that there would be ample time for even weak friction to reduce the westerlies does not seem to have found favour.

In a remarkable paper the renowned physicist Hermann von Helmholtz (1888) attacked this problem. He had previously (1868) been the first to emphasize that the motion in a fluid need not be everywhere continuous. He began the present work by noting that friction was extremely ineffective in the atmosphere, except at the Earth's surface and at internal surfaces of discontinuity. Clearly, he was here referring to molecular friction. He then noted the large velocities which would be required by a circulation between the Equator and 30° latitude, and maintained that while such a circulation did occur, the large velocities were not found. He thereupon sought the means by which the winds were prevented from attaining them.

In seeking a solution Helmholtz virtually developed a theory of the general circulation. By reasoning somewhat like Ferrel's, he deduced that in the absence of friction there would be easterly winds in low latitudes and westerlies in higher latitudes. He then proceeded to determine how this circulation would be modified by heating and friction. Like Ferrel and Thomson, he found that surface friction would produce a poleward drift in the surface westerlies, and intensify the equatorward drift in the surface easterlies.

In his next step he proceeded beyond the earlier authors. He maintained that the returning air above the trades must come into immediate contact with the cooler and more slowly moving air below, with the formation of a surface of discontinuity. At such a surface the equilibrium would be unstable, so that vortices would form, and ultimately bring about vertical mixing.

In the polar regions he felt that the effect of cooling would outweigh the effect of surface friction, and lead to additional equatorward spreading. Easterly winds would thereby develop, and the resulting friction would cause further spreading. Again a surface of discontinuity would form between this air and the returning air aloft, and vertical mixing would again occur.

He concluded with the opinion that the principal deterrent to stronger winds aloft was not surface friction, but the mixing of layers of different velocities by means of vortices forming on surfaces of discontinuity.

Helmholtz's paper is often regarded as the original statement that cyclones must form upon surfaces of discontinuity, and that these cyclones will in turn alter the general circulation. Admittedly some statements are subject to more than one interpretation, but we do not feel that this is what Helmholtz was saying. The vortices which he visualized seem to have horizontal scales of hundreds of metres, or perhaps a few kilometres, but not thousands of kilometres, and he frequently mentioned billow clouds. He referred to cyclones in only two connexions, in neither case as disturbances on unstable surfaces of discontinuity. He first suggested that they should form in middle latitudes under the masses of ascending air. Later he clearly maintained that the permanent circumpolar anticyclonic motion at the Earth's surface and the cyclonic motion above it should break up into smaller cyclones and anticyclones as a result of surface irregularities, such as mountains. In neither case did he specifically say that these cyclones would affect the general circulation; it is the vortices which develop on the surfaces of discontinuity which were assigned this role. In a second paper (1889) he mentioned that the numerous disturbances should cause the principal surface of discontinuity to break "into separate pieces which must appear as cyclones," but he did not elaborate further.

If Helmholtz did not give the meteorological world the wave theory of cyclones, he gave it the concept of turbulent viscosity, whose presence is an essential feature of every modern theory of the general circulation. It is worth noting in this connexion that if the coefficient of turbulent viscosity is assumed to have a value throughout the atmosphere of from 10 to 100 g cm^{-1} sec^{-1}, a figure frequently quoted for the surface friction layer, the upper-level poleward current cannot exceed from 1 to 10 cm sec^{-1} without giving rise to upper-level westerlies in excess of those allowed by the thermal wind relation.

We therefore feel that Helmholtz's work, far from disproving the ideas of Ferrel and Thomson, tends to support them by showing that the absence of excessively high winds can be accounted for. At the same time, it plainly suggests how departures from zonal symmetry can be important. Ultimately it led the way to the great work of Vilhelm Bjerknes and his co-workers, and to their theories of the general circulation in which cyclones played a fundamental role.

Further objections to the symmetric theories of the general circulation were based upon observations. Routine soundings by balloons were non-existent in the nineteenth century, and many of the ideas concerning upper-level conditions had been deduced from mountain observations. However, as a result of an international conference convened by the International Meteorological Organization in 1891, a decision was made to perform a world-wide investigation of upper-level currents at different levels by observing the typical motions of different forms of clouds. All countries were invited to participate. The proposed programme became a reality in 1896 and 1897 (see Bigelow 1900).

The results of this programme, together with earlier cloud observations at selected stations, formed the basis of an assessment of the theories of the general circulation by Hildebrandsson and Teisserenc de Bort (1900). The authors concluded that the only dynamically acceptable schemes so far proposed were those of Ferrel and Thomson, but they noted that at upper levels these schemes were based upon theory rather than observation. From their cloud study they found that the average winds at the cirrus-cloud level were easterly over the Equator, becoming south-east, south-west, and finally west as the thirtieth parallel north was approached. They found no evidence at all for an upper-level poleward current extending through middle latitudes, and, although recognizing that the high-level motions were observed only when high clouds were present and no other clouds were below them, they nevertheless concluded that the schemes of Ferrel and Thomson were contradicted by the observations, and must be abandoned. They refrained from offering a scheme of their own.

Today we would regard the selection of stations as quite inadequate for eliminating the possibility of an upper poleward current. It is likely, however, that the authors were seeking a current whose strength was similar to that of the trade winds, perhaps a few metres per second, which the cloud observations may have been sufficient to eliminate. Certainly currents as weak as 10 cm sec^{-1} could not have been disproved. Yet modern observations deny the upper current just as surely as they confirm the poleward current near the surface.

Similar conclusions were drawn by Bigelow (1900, 1902), on the basis of the cloud observations over the United States. He noted an almost perfect balance between northerly and southerly wind components at high and intermediate levels and cited this as evidence against all the "canal" theories, as he termed those theories which allowed no variations with longitude. He then offered a scheme of his own, in which the flow at high levels was fairly uniform, but where cold and warm countercurrents from high and low latitudes flowed beside one another, mainly in the lowest three kilometres, and where the interaction of these currents gave rise to cyclones and anticyclones. In this respect his description seems familiarly like Dove's.

Yet Bigelow went considerably beyond Dove, who had been content to present the observations. He proposed that the warm and cold currents were the means by which the required poleward heat transport was accomplished. He moreover identified the cyclones and anticyclones as the means by which the excessively strong winds at upper levels were forestalled; specifically, he noted that the upward currents in the cyclones and the downward currents in the anticyclones would bring about a vertical exchange of momentum.

Bigelow's theory leaves some questions unanswered; there is still no specified mechanism for carrying angular momentum horizontally across middle latitudes, even though the vertical transport is present. Yet not only is the energy balance satisfied, but there is a clear statement of the necessity for a poleward heat transport, and of the mechanism through which it is accomplished. Only the more recent observations have revealed that Bigelow should not have concentrated the irregularities in the lowest kilometres.

By the turn of the century the study of the circulation had undergone a permanent change. A few years earlier the works of Ferrel and Thomson had appeared to offer a nearly complete explanation of the circulation. Now with the increasing realization that the general circulation involved more than the zonally symmetric motions, it became apparent that a full explanation was a far more difficult task than had been supposed. There was even some feeling that the circulation could not be explained at all.

From this time on few of the important new papers attempted to account for the circulation *in toto*. The most significant contributions were often confined to a single aspect. Among these was the further investigation of the role actually played by cyclones and other disturbances.

For a number of years Bigelow's ideas were often quoted but not pursued. A major advance was finally made by A. Defant (1921). In this famous paper Defant introduced the idea that the motion in middle latitudes was simply turbulence on a very large scale. He regarded the cyclones and anticyclones as the individual turbulent elements, by means of which the required amount of heat was transported from low to high latitudes. He also looked upon this large-scale turbulence as the means by which excessive wind speeds aloft were prevented from occurring.

If he had stopped at this point, he would have done no more than repeat Bigelow's ideas in a new language. Instead, he warned that his conclusions could not be considered valid unless they were quantitatively acceptable. Accordingly, he first applied the recently formulated mixing-length theory of turbulence. Assuming a mixing length of 15 degrees of latitude and an average north-south wind component of 3 m sec^{-1}, he found a horizontal Austausch coefficient — the ratio of the transport to the gradient of sensible heat — of 5×10^7 g cm^{-1} sec^{-1}, nearly a million times larger than the Austausch coefficient characterizing smaller-scale turbulence. He then estimated the Austausch coefficient by other procedures, and also found values somewhere near 10^8 g cm^{-1} sec^{-1}.

Choosing different values of the Austausch coefficient ranging from 5×10^7 to 5×10^8, he calculated the temperature distribution which should prevail north of 30°N, assuming the temperature distribution which would prevail in the absence of any horizontal exchange to be known, and he found that a value of 10^8 would lead to temperatures agreeing fairly well with observations. He thereupon concluded that his case for the circulation as a form of turbulence was established.

Whatever the general attitude may have been toward looking upon the circulation as turbulence, Defant's claim that the cyclones and anticyclones accomplished the needed transport of heat seems to have been well accepted. Perhaps it was hard to deny that north winds were colder than south winds. The conclusions which logically follow should have obviated some of the mistaken reasoning which occurred in many subsequent works. If some of the required poleward heat transport is accomplished by contrasting currents at the same level, the amount of heat remaining to be transported by other mechanisms is less than it would otherwise be. The meridional circulation will therefore not by itself transport enough heat to satisfy the balance requirements, and the zonally averaged flow will not by itself be a solution of the dynamic equations. Any attempt such as Oberbeck's to determine a zonally symmetric general circulation must therefore, if correctly completed, disclose a meridional circulation differing from the one actually occurring in the atmosphere. Any attempt to force the symmetric solution to agree with the atmosphere will fail; if it appears to succeed, it will do so only because of some unjustified steps.

The other major contribution of this period concerning the role of cyclones was made by Jeffreys (1926). Unlike any of his predecessors, Jeffreys was concerned with the manner in which angular momentum was conveyed from low to high latitudes, rather than the manner in which it was brought down from high levels after being conveyed to high latitudes. He noted that in the long run Ω-momentum need not be considered, since its net transport is proportional to the net mass transport. Thus he arrived at the conclusion that the net angular-momentum transport was proportional to the product of the eastward and northward wind components.

Assuming a zonally symmetric flow with no meridional motion except within the lowest kilometre, or the friction layer, he found that the amount of angular momentum carried northward across middle latitudes was too small by a factor of at least 20 to balance the angular momentum transferred into the Earth. He concluded that the bulk of the required transport must be accomplished by large-scale eddies, which he identified as cyclones, and which he felt should extend to considerable heights.

There are certain difficulties with Jeffreys's arguments. Because he regarded the winds as essentially geostrophic except in the friction layer, he did not consider the necessary return flow aloft, which is somewhat surprising, since he had carefully used the principle of mass continuity to eliminate the need for considering Ω-momentum. This point was eventually straightened out by Douglas (1931), who noted that if the needed return flow took place at the level of the strong upper westerlies, it would transport angular momentum equatorward and lead to even greater difficulties. He observed that a balance could be achieved if the return flow occurred immediately above the friction layer, with additional poleward flow at high levels, in the manner proposed by Ferrel and Thomson. However, he found no observational evidence for equatorward flow in the lowest four kilometres, and concluded that Jeffreys was correct in deciding that the exchange of angular momentum was carried out by currents lying side by side.

Jeffreys must be given credit for first stating the need for a horizontal angular-momentum transport and for correctly identifying the mechanism through which it is accomplished. Perhaps meteorologists found Jeffreys's notation somewhat unfamiliar, but if they had simply turned to the equations of motion and written the expression for the rate of change of angular momentum, they would have been forced to conclude that there was either a direct meridional cell operating across middle latitudes, or else a correlation between the eastward and northward wind components within latitude circles. If they had believed in an indirect cell in middle latitudes, as many of them did, they would have been forced to accept the latter conclusion. As it was, a generation had to pass before Jeffreys's ideas were generally accepted.

Another aspect of the general-circulation problem which received considerable attention was the explanation for the very existence of departures from zonal symmetry. The problem as to why cyclones exist is fundamental in cyclone theory, but with the realization that cyclones played a role in the general circulation it gained importance in general-circulation theory as well.

A number of writers maintained that a circulation without cyclones was dynamically impossible. Jeffreys believed that he had established the necessity for cyclones, and he is often quoted as having shown that a symmetric circulation is impossible. Actually, in addition to his omission of the return current aloft, his argument is based on the observed structure of the friction layer; at most he could have shown that a symmetric circulation cannot possess a friction layer like the one actually observed.

Shortly thereafter another argument began to appear in the literature. It was pointed out that zonally uniform equatorward or poleward motion would be kinematically possible, but that no zonally uniform eastward or westward pressure gradient could accompany it, since the pressure could not possibly vary in the same sense at all points of the same latitude circle. The argument was generally attributed to Exner (1925).

It is easy to see how such an argument might have arisen at a time when there was excessive reliance upon the geostrophic wind equation, since there certainly can be no zonally uniform equatorward or poleward geostrophic wind, but it is surprising that it should have been so frequently quoted as a proof that zonally symmetric flow is impossible. It is reminiscent of the remark made many years earlier by Dalton. There is however one additional feature: Exner did not make the statement with which he is credited.

Exner like many others before him was interested in explaining the absence of the excessive upper-level winds which conservation of angular momentum would seem to require. He noted that turbulent mixing would be one means of reducing the angular momentum aloft, but he felt that it would probably be insufficient, and he maintained that it was much more likely that east-west pressure gradients would develop and accomplish the same end. He observed that in this case the circulation could not be zonally symmetric.

To say that east-west pressure gradients are needed to maintain the observed flow is quite a different thing from saying that they must be present in any case. Exner did not try to show that asymmetries must exist. His important contribution was the identification of east-west pressure gradients rather than turbulent mixing as the means by which excessive upper-level winds are precluded; this contribution has been generally overlooked as a result of the misinterpretation of his remarks.

In retrospect, it is hard to understand why zonally symmetric flow should have been considered dynamically impossible. If one chooses an initial condition of complete zonal symmetry, the time-dependent solution which develops must remain zonally symmetric. This is of course true only for an idealized uniform Earth, but it was for such an Earth that the arguments had been given. The initial-value approach has become familiar with the advent of numerical weather prediction; possibly it did not often enter the dynamical thinking of an earlier era.

An alternative school of thought maintained that a symmetric general circulation might be possible, but that it would in some way be unstable, and that cyclones would develop. The proper formulation of this idea developed from the work of V. Bjerknes and his collaborators.

The early work of this group was a natural outgrowth of the meteorological work of Helmholtz. Years later, in discussing this work, Bjerknes (1933, p. 784) expressed his doubts that Helmholtz actually regarded cyclones as the disturbances which would form on unstable surfaces of discontinuity, but acknowledged that Helmholtz's work was sufficient to guide his thoughts in this direction. The eventual identification by J. Bjerknes (1919) of an observed cyclone as a wave on the polar frontal surface is a familiar chapter in meteorological history.

Although much of the earlier work was concerned with the cyclone problem, the concept of a cyclone as a disturbance growing upon a pre-existing flow pattern required a consideration of the pre-existing pattern itself. Ultimately it led to a close examination of the general circulation for its own sake.

In one discussion of the general circulation, V. Bjerknes (1921) noted the difficulties involved in deducing a zonally symmetric circulation from pure theory, and based his picture upon a combination of theory and observation; it is shown in Figure 36. He then noted that this circulation appeared to be unstable, and that cyclones should develop, leading ultimately to an unsymmetric general circulation, as shown schematically in Figure 37.

Bjerknes did not apply any specific test for stability, and his conclusion that the symmetric circulation was unstable may have been based upon his conviction that instability must be the cause of cyclones. In any event, he had not yet distinguished between the zonally averaged circulation and the symmetric circulation which would prevail in the absence of disturbances, as he was to do in a later paper.

Somewhat later, in a study of the development of fronts, Bergeron (1928) found it necessary to introduce a model of the general circulation, and he proposed a three-cell meridional circulation somewhat similar to the one which Ferrel had introduced and subsequently discarded. It is shown in Figure 38. It has been widely reproduced, and it seems to mark the beginning of the general acceptance of a three-cell pattern. The fact that it does not by itself satisfy the balance requirements is irrelevant, since Bergeron regarded the superposed eddies as a further essential part of the circulation. The middle-latitude indirect cell has since come to be known as the Ferrel cell.

Somewhat later J. Bjerknes presented a new scheme of the general circulation (V. Bjerknes *et al.*, 1933). He first described a zonally symmetric circulation which could satisfy the energy balance requirements. However he rejected the idea that the actual circulation was symmetric, and proposed an unsymmetric circulation, which is by no means a copy of Dove's of Bigelow's. In this circulation each particle travels in a circuit with an equatorward branch below and a poleward branch above, as in Hadley's theory, thus fulfilling the energy balance requirements, but the circuit is different at different longitudes. As in the symmetric theories, the Coriolis force will tend to deflect the equatorward flow westward and the poleward flow eastward, but the pressure will build up between converging deflected currents and prevent further deflection. Thus, in agreement with Exner's ideas, air can continue to flow poleward without acquiring excessive angular momentum.

Bjerknes did not describe the zonally averaged meridional motion, and went so far as to say that it would be without interest, since it would merely be a statistical average which would be close to zero. It is to be noticed that his subtropical high pressure cells have a typical ENE-WSW orientation. Possibly he had introduced this configuration simply as a result of careful observation of the real atmosphere. Nevertheless, such a pattern leads to the positive correlation between eastward and northward motion which is needed for a poleward transport of angular momentum, and it foreshadows some of the contributions which Bjerknes was to make some years later.

The explanation for the existence of cyclones which we believe to be the correct one was provided in a later paper by V. Bjerknes (1937). Here he concluded that the circulation which would prevail in the absence of departures from zonal symmetry was essentially the one given by Ferrel and Thomson, with a large direct cell occupying most of either hemisphere, and a shallow indirect cell at low levels in middle latitudes. This circulation differs considerably from his earlier picture, which was based mainly upon observations. He then maintained that this circulation was unstable with respect to small zonally unsymmetric disturbances; hence the observed circulation would contain fully developed disturbances, which would assume the form of cyclones and anticyclones. This remarkable paper, published in his seventy-fifth year, makes a fitting culmination to his contributions to this problem.

The most difficult aspect of the general circulation problem is however neither the explanation of the existence of cyclones and other disturbances nor the determination of their role, but the explanation

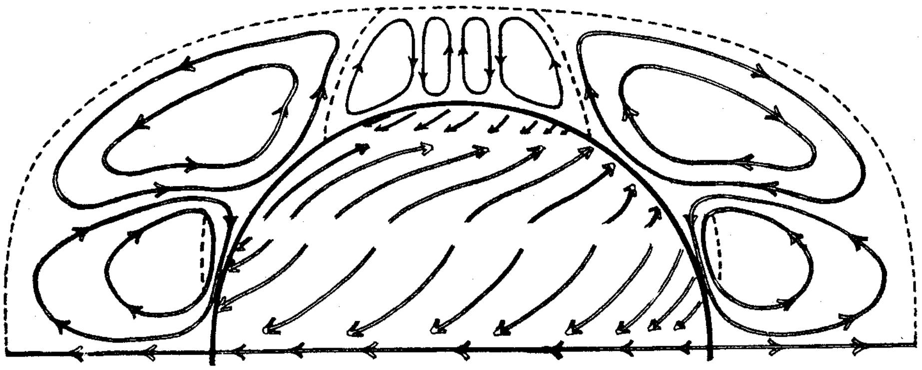


Figure 36. — A schematic representation of the zonally averaged circulation according to Bjerknes (1921)

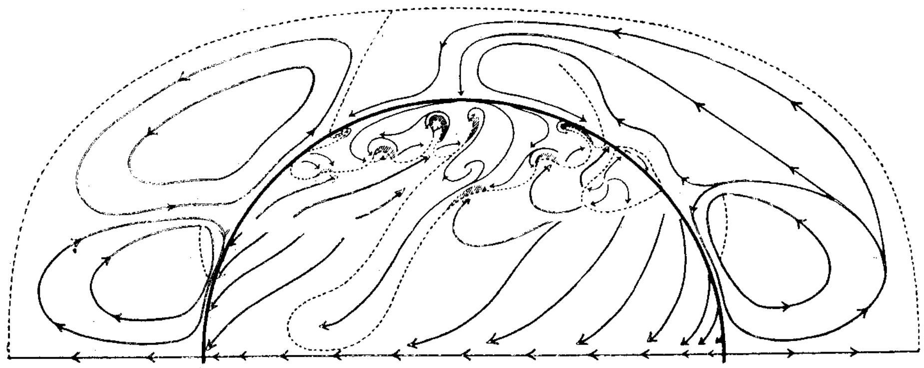


Figure 37. — A schematic representation of the general circulation of the atmosphere according to Bjerknes (1921)

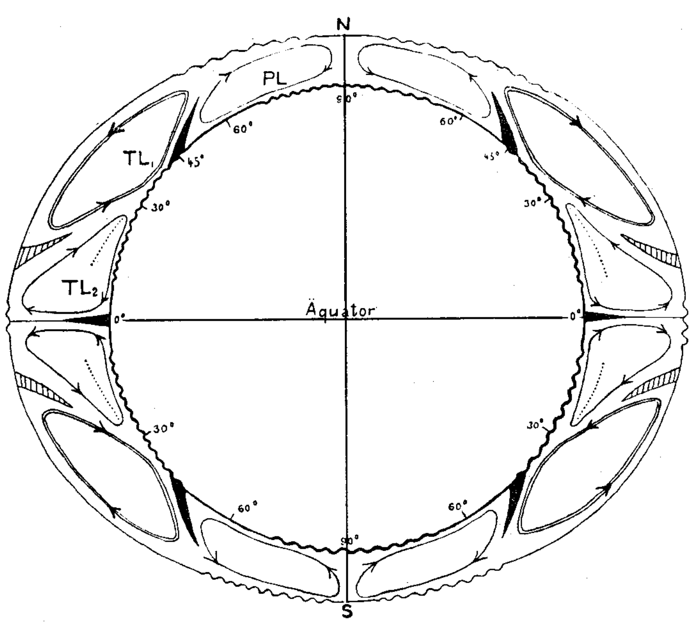


Figure 38. — A schematic representation of the meridional circulation according to Bergeron (1928)

of why cyclones behave as they do, and particularly why they transport angular momentum and energy as they do. Defant's application of turbulence theory seemed to offer a partial explanation for the heat transport, although one may legitimately question whether turbulence theory explains turbulence or merely describes it. Following the contribution of Jeffreys, it was natural to try to explain the angular-momentum transport by turbulence theory also.

In discussing the work of Jeffreys, Douglas (1931) noted that Defant's value of the Austausch coefficient applied directly to the gradient of absolute angular momentum would yield a transport a hundred times greater than that demanded by the balance requirement. He concluded that angular momentum did not diffuse in any such simple manner.

His conclusion was not universally heeded. The fact that turbulence theory appeared to yield the right sign for the transport of angular momentum in most of the atmosphere encouraged workers to pursue the matter, using a smaller Austausch coefficient than mixing-length concepts would seem to have demanded. At the time of Oberbeck's work it had appeared possible to present a complete description of the general circulation in mathematical form by solving the appropriate equations. With the realization that departures from zonal symmetry played a significant role, the possibility seemed more remote. Now, by representing the effects of cyclones and other disturbances through appropriate Austausch coefficients, it might again be possible to work with a system of equations with only latitude and elevation as independent variables.

We must therefore note that no matter how conservative absolute angular momentum may be, classical turbulence will attempt to remove internal stresses by creating a state of solid rotation, which it can do only by transporting angular momentum towards latitudes of lower angular velocity — not towards latitudes of lower absolute angular momentum. For the most part a transport towards lower angular velocity is a transport towards a region of weaker westerlies or stronger easterlies. Angular velocity is certainly not conservative, yet the notion that mixing should somehow lead to a uniform distribution of some conservative quantity is hard to dispel.

The most extensive attempts to apply horizontal-mixing concepts to the general circulation are to be found not in complete solutions of the equations but in the qualitative and semi-quantitative works of Rossby. In these works collectively we find him seeking the proper explanation; he explores one possibility, and, when it proves to be untenable, he turns to another.

The idea of an upper-level poleward current across middle latitudes had by now generally been abandoned, and one of the problems was to explain the strength rather than the weakness of the upper westerlies. In an early paper devoted to the problem, Rossby (1938a) proposed that the westerlies south of the polar front (in the northern hemisphere) could be maintained through large-scale lateral mixing with the supposedly stronger westerlies above the frontal surface, i.e. by being dragged ahead. The presence of an upper-level westerly maximum as far south as 35°N was not yet generally recognized. He assumed a surface frictional drag proportional to the square of the wind speed, and a lateral drag proportional to the square of the horizontal wind shear, and after introducing other simplifications was able to represent the motion in middle latitudes by analytic functions. In his solution the westerlies decreased exponentially with distance south of the polar front.

However his solution did not extend north of the polar front. In view of the proximity of the polar regions to the Earth's axis, it is difficult to see how any surface easterly winds there could be strong enough to supply through friction the angular momentum which in turn was supposed to be supplied to middle latitudes through mixing.

In his best-known treatment of the general circulation, which was addressed to a much wider audience than the meteorological world, Rossby (1941) modified the ideas of his earlier paper by supposing that the westerlies south of the polar front were maintained by mixing with the westerlies farther southward as well as farther northward. Thus he visualized a circulation where the necessary transports of angular momentum were accomplished by large-scale eddies, in the manner proposed by Jeffreys, but he went beyond Jeffreys in relating the transports to mixing concepts.

Not being content with a mere description, he preceded his discussion of mixing by attempting the difficult task of qualitatively deducing the zonally symmetric circulation which would develop from a hypothetical initial state of meridional motion only. He showed that something more complicated than a single direct cell should develop, but his choice of a three-cell pattern seems to have been guided by observations, whereas, as we have noted, the meridional circulation to be expected in the absence of disturbances may not contain three cells at all. The difficulty in following his reasoning at this point has undoubtedly caused many readers to overlook his contributions concerning the role of the eddies.

Still, Rossby could not reconcile the transport of angular momentum from low to middle latitudes, and supposedly from weaker to stronger westerlies, with the ideas of large-scale diffusion. Thus he was ultimately led (1947) to explore the possibility that large-scale lateral mixing is characterized by a transfer of vorticity toward latitudes of lower absolute vorticity, although he carefully noted that small-scale mixing would not have this effect. Specifically, he posed the question as to what distribution of zonal motion should develop in a thin spherical shell under the influence of lateral mixing. Under the assumption that vorticity would be transferred from a reservoir of positive vorticity in high latitudes in the northern hemisphere to a similar reservoir of negative vorticity in the southern hemisphere, he obtained a Pole-to-Pole profile of westerly wind speed which bore a fair resemblance to the observed upper-level winds in the Earth's atmosphere, with easterlies in the equatorial regions and a westerly maximum in middle latitudes in either hemisphere.

Yet the mechanism for the maintenance of this profile can hardly be equated to the mechanism prevailing on Earth, for although the needed source and sink of vorticity may be present in the surface frictional drag of the polar anticyclones, there should be at least equally intense sources and sinks in the sub-polar cyclonic cells and the subtropical anticyclones. Ultimately, Rossby and some of his collaborators were led to the conclusion that some process more complicated than classical turbulence must be present.

Fulfilment of the balance requirements

The modern era in the study of the general circulation begins with the proposals by Starr (1948), Bjerknes (1948), and Priestley (1949) that routine upper-level observations should now be plentiful and accurate enough for direct evaluation of the transport of angular momentum. Such computations might settle the question of the relative importance of the eddies and the meridional circulations. It must be noted that the ideas which Jeffreys (1926) had presented some twenty years earlier had become fairly well known but were by no means universally accepted. Starr and Bjerknes both expressed the opinion that the angular-momentum transport across middle latitudes would prove to be accomplished mainly by the eddies, as Jeffreys had maintained.

Starr observed that the required northward transport could be produced by troughs and ridges with a general NE-SW orientation. As Bjerknes also noted, elongated quasi-elliptical anticyclones with their major axes oriented WSW-ENE could produce the same effect in lower latitudes. Priestley also considered the transports of water and sensible heat, and illustrated his proposal with computations from two years of upper-level data at Larkhill, England. A few words about single-station computations are in order.

We have noted that the field of northward motion may be resolved into a meridional circulation or meridional cells and superposed eddies, and that the meridional cells may be resolved into standing or time-averaged meridional cells and superposed transient or instantaneous meridional cells, while the eddies may be resolved into standing eddies, which appear on time-averaged maps, and superposed transient or migratory eddies. The resolution is given by

$$v = [\bar{v}] + [v]' + \bar{v}^{*} + v^{*\prime}, \tag{91}$$

which follows from (80). The long-term northward transport of any quantity X may then be resolved into the amounts accomplished by the separate components of v; thus

$$[\overline{X v}] = [\overline{X}]\,[\bar{v}] + [\overline{[X]'\,[v]'}] + [\overline{X}^{*}\bar{v}^{*}] + [\overline{X^{*\prime}v^{*\prime}}]. \tag{92}$$

For brevity we may refer to the separate modes of transport as the standing-cell transport, transient-cell transport, standing-eddy transport, and transient-eddy transport.

Any long-term poleward mass flow past a single station will carry angular momentum, water, and energy with it, and thereby contribute to the poleward transports of these quantities, but this contribution will be largely cancelled by the necessary long-term equatorward mass flow at some other station. The observations at the first station alone cannot reveal the extent to which it is cancelled. Thus, as Priestley noted, the standing-eddy transport cannot be estimated from data at one station only.

Priestley recombined the terms in (92) representing the transient-eddy transports; thus

$$[\overline{Xv}] = [\overline{X}]\,[\bar{v}] + [\overline{X}^{*}\bar{v}^{*}] + [\overline{X'v'}]. \tag{93}$$

He estimated the final term in (93), which he regarded as the transport by the transient eddies, by assuming the covariances $\overline{X'v'}$ to be independent of longitude. He also estimated the standing-cell transport by assuming the departure of $[\bar{v}]$ from its vertical average to be independent of longitude. He concluded that both the meridional circulation and the eddies were important in effecting the required transports. He furthermore found no indication that the eddy transports were in agreement with mixing-length concepts.

Priestley's main objective in performing the computations with data from one station was to demonstrate the feasibility of global computations, rather than to obtain definitive measurements. More recent studies have shown that the statistical properties of the transient eddies vary considerably with longitude. Nevertheless, Priestley's results are sufficient to indicate that the eddies may play an important role.

The first transport computations extending around the globe were performed by Widger (1949), who used data for the single month of January 1946. At that time upper-level wind coverage was still not plentiful, and Widger used geostrophically estimated wind components, obtained from analysed sea-level, 700-mb, and 500-mb northern-hemisphere maps, at the intersections of standard latitudes and longitudes. This procedure automatically eliminates the transport by the meridional cells, which are entirely non-geostrophic. Widger obtained angular-momentum transports of the proper sign and order of magnitude to satisfy the balance requirements, with the major contribution coming from the 500-mb observations.

Hemispheric computations extending through the depth of the troposphere were first carried out by Mintz (1951), who used geostrophic-wind data for January 1949 extending up to 100 mb. Analysed maps above 300 mb were still a rarity, and Mintz's computations had to be preceded by a painstaking analysis of a series of upper-level maps. Mintz found that the angular-momentum transport occurring near the tropopause far outweighed the transport below 500 mb, with the jet stream apparently playing a major role.

Still, the results were not entirely convincing. The computed correlations between u and v were rather low, generally between $+0.1$ and $+0.2$, and small but systematic departures from the geostrophic wind might conceivably have reduced them to zero. The results were especially unconvincing to those who held to mixing-length concepts, since south of the westerly-wind maximum the transport was toward stronger westerlies. In any event there were no direct computations of the cell transport for comparison.

Convincing evidence of the importance of large-scale eddies in the angular-momentum balance was finally provided by Starr and White (1951), who used observed winds at a chain of sixteen stations extending around the globe in the vicinity of 30°N. They found a large poleward transport by the eddies, and such a small transport by the meridional circulation that even its sign was in doubt.

Meanwhile the eddies were proving to be important in the transports of other quantities. Since Defant's famous paper it had been generally accepted that the eddies could transport heat, but White (1951) found that geostrophically measured sensible-heat transports agreed well with the balance requirements. Benton and Estoque (1954) performed a detailed study of the flux of water vapour over North America and the adjacent oceans, and found that the hemispheric eddy transport of water vapour, as estimated from this quadrant, was sufficient to satisfy the balance requirements at these latitudes.

Subsequently Starr and White extended their computations to chains of stations at other latitudes, and finally (1954) combined these computations, along with similar ones for the transports of water and sensible heat, into a complete hemispheric study. With their computation procedure it was feasible to distinguish between the transient meridional circulation and the transient eddies, but not between the standing eddies and the transient eddies; thus their form of equation (92) was

$$[\overline{Xv}] = [\overline{X}]\,[\overline{v}] + \overline{[X]'\,[v]'} + [\overline{X^{\star}v^{\star}}]\,. \tag{94}$$

Virtually all the transport of angular momentum proved to be accomplished by the eddies, except at the southernmost latitude, 13°N, where the meridional circulation also gave an important contribution.

With the importance of the large-scale eddies firmly established, the primary purpose of subsequent transport computations became the determination of more appropriate numerical values. The most extensive collection of these computations appears in the works of the Planetary Circulations Project (formerly the General Circulation Project) at the Massachusetts Institute of Technology, which has been continually directed by V. P. Starr. Some of the earlier works of this project have already been mentioned (Widger 1949, White 1951, Starr and White 1951, 1954). The author considers himself fortunate to have been associated with this project since its inception in 1948, and is pleased to take this opportunity to present its most recent estimates of the various transports.

In the newer computations the transports have been evaluated by a uniform procedure first used by Buch (1954) in the study mentioned in Chapter III. In this study Buch used all available upper-wind data for the year 1950, at a network of 81 stations over the northern hemisphere. For each of these stations, at each of the six pressure levels 850, 700, 500, 300, 200, and 100 mb, he evaluated the statistics $\bar{u}$, $\bar{v}$, and $\overline{u'v'}$, the time-averaging being for the whole year. He also evaluated $\bar{u}$ and $\bar{v}$ separately for the summer and winter halves of the year, and computed certain other statistics, including standard deviations of u and v.

He then constructed a map of each statistic for each level, by recording the computed statistics on the map and drawing isopleths. From the analysed maps he interpolated the values of the statistics at the intersections of standard parallels and meridians, and then summed over longitude to obtain estimates of $[\bar{u}]$, $[\bar{v}]$, and $[\overline{u'v'}]$, and hence the separate terms in equation (93), the form used by Priestley. He then multiplied the terms by the appropriate latitude factor to obtain his estimates of the angular momentum transport by the meridional cells, standing eddies, and transient eddies.

The map analyses were necessarily subjective. With an average of only one station per three million square kilometres, there were some large continuous regions with no data at all. Only pilot-balloon observations were available between 60°E and 110°E, and the 200-mb and 100-mb maps could not be analysed in that sector. In fact, Buch regarded his procedure as experimental. Since that time upper-level wind measurements have become routine at many more stations, and it would be possible today to perform a similar study with several hundred stations, although some large areas with scanty data still remain.

A similar study for the southern hemisphere using 1950 data would have been out of the question, but with the inauguration of the International Geophysical Year a reasonable number of stations became established. Using all the available upper-wind data for the year 1958 at a network of 145 stations, Obasi (1963) applied the computational procedure used by Buch to the southern hemisphere.

Figure 39 compares the annual mean northward transport of angular momentum, as computed by Buch and Obasi, with the balance requirement as shown in Figure 23. In view of the various approximations involved in both curves, the agreement is remarkable. The feasibility of computing angular-momentum transports directly would appear to have been established.

Figure 40 compares the annual mean eddy transport of angular momentum with the meridional-cell transport. The general predominance of the eddies is evident; the computed eddy transport alone fulfils the estimated balance requirement as closely as does the total computed transport. It is gratifying to find that the computed cell transport conforms to the three-cell pattern in each hemisphere, but beyond this result very little reliance can be placed in the estimated values. The data are simply not adequate to give a reliable picture of the meridional circulation.

Obasi determined the transports separately for each season; the average of these is shown. Buch computed the transports only for the whole year, but from summer and winter values of $[\bar{u}]$ and $[\bar{v}]$ we have determined the cell transport for each season, and have shown the average of these. Without this modification the Hadley cell, whose position shifts with the seasons, would appear mainly as a transient meridional circulation, and the computation procedure would include the transport which it accomplishes as part of the transient-eddy transport.

Figure 41 shows the vertical distribution of the horizontal eddy transport of angular momentum. The most conspicuous feature is the extreme concentration near 200 mb and 30° latitude in either hemisphere, suggesting the great importance of the jet streams in maintaining the balance of angular momentum. Nearly half of the total transport occurs within a 200-mb layer. Both Buch and Obasi also determined the transient-eddy and standing-eddy transports separately. In the northern hemisphere the latter also shows a concentration near the region of maximum westerlies, and accounts for about twenty per cent of the total eddy transport. In the southern hemisphere, where geographical influences tend to be less pronounced, the standing-eddy transport is weaker and less regularly distributed.

Comparing Figure 41 with Figures 1-8, we see that throughout approximately half the atmosphere — the tropics and the lower temperate latitudes — the eddy transport of angular momentum is directed

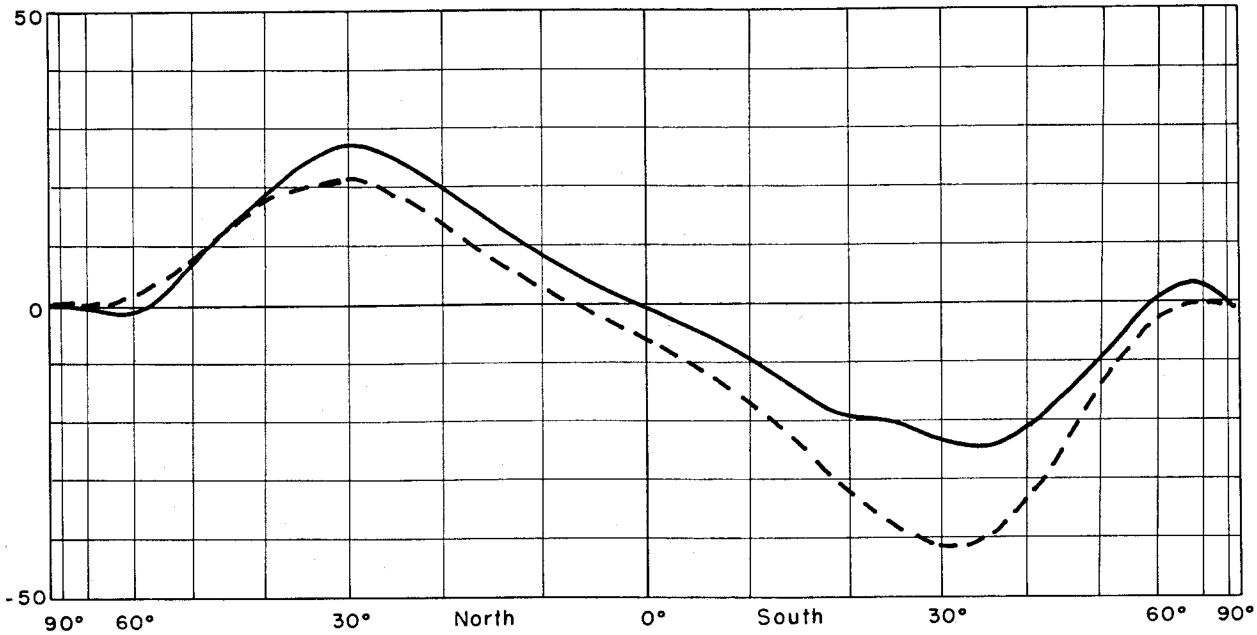


Figure 39. — The average northward transport of angular momentum (solid curve) as estimated by Buch (1954) (northern hemisphere) and Obasi (1963) (southern hemisphere), and the required transport (dashed curve) as given in Figure 23. The unit is 10^{25} g cm^2 sec^{-2} (scale on left)

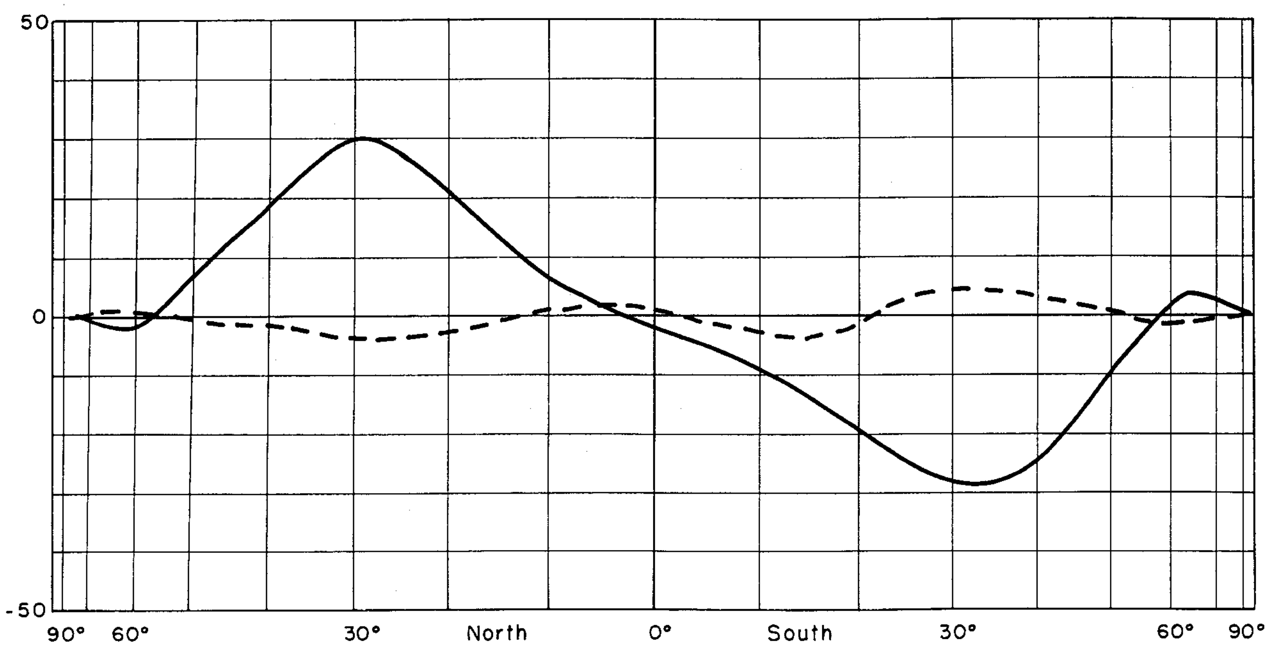


Figure 40. — The average northward transport of angular momentum by the eddies (solid curve) and by the meridional circulation (dashed curve) as estimated by Buch (1954) (northern hemisphere) and Obasi (1963) (southern hemisphere). The unit is 10^{25} g cm^2 sec^{-2} (scale on left)

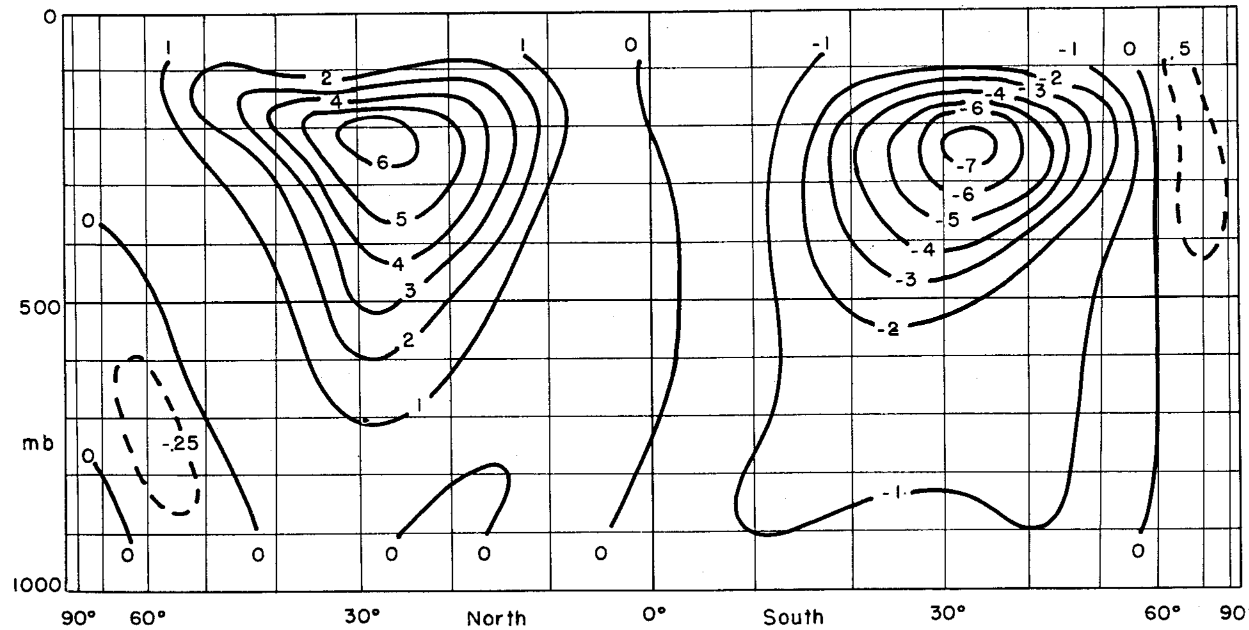


Figure 41. — The vertical distribution of the average northward transport of angular momentum by the eddies as estimated by Buch (1954) (northern hemisphere) and Obasi (1963) (southern hemisphere). The unit is 10^{25} g cm^2 sec^{-2} per 100-mb layer

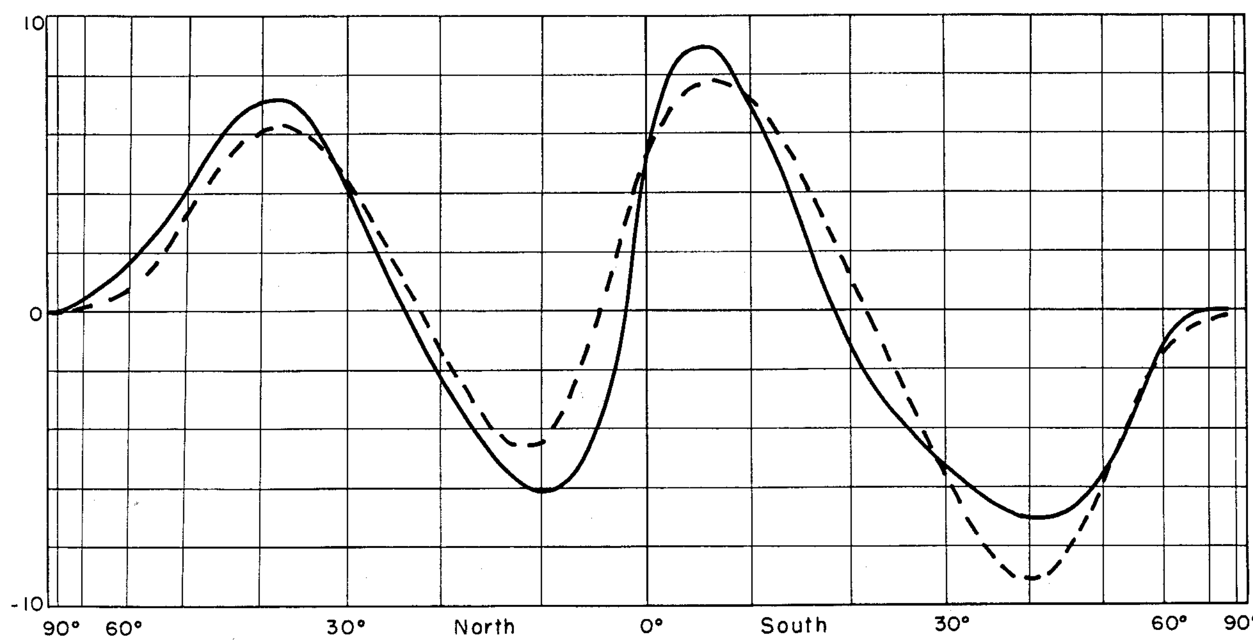


Figure 42. — The average northward transport of water (solid curve) as estimated by Peixoto and Crisi (1965) (northern hemisphere) and Peixoto (southern hemisphere), and the required transport (dashed curve) as given in Figure 21. The unit is 10^{11} g sec^{-1} (scale on left)

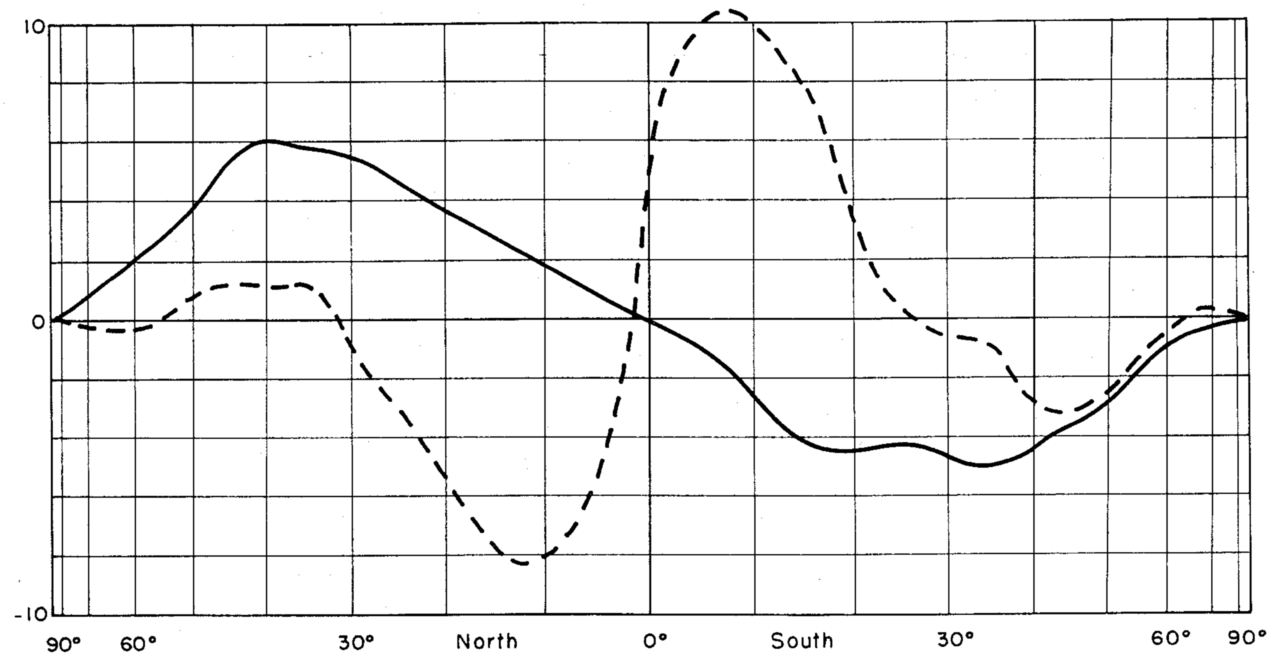


Figure 43. — The average northward transport of water by the transient eddies and transient meridional circulation (solid curve) and by the standing eddies and standing meridional circulation (dashed curve) as estimated by Peixoto and Crisi (1965) (northern hemisphere) and Peixoto (southern hemisphere). The unit is 10^{11} g sec^{-1} (scale on left)

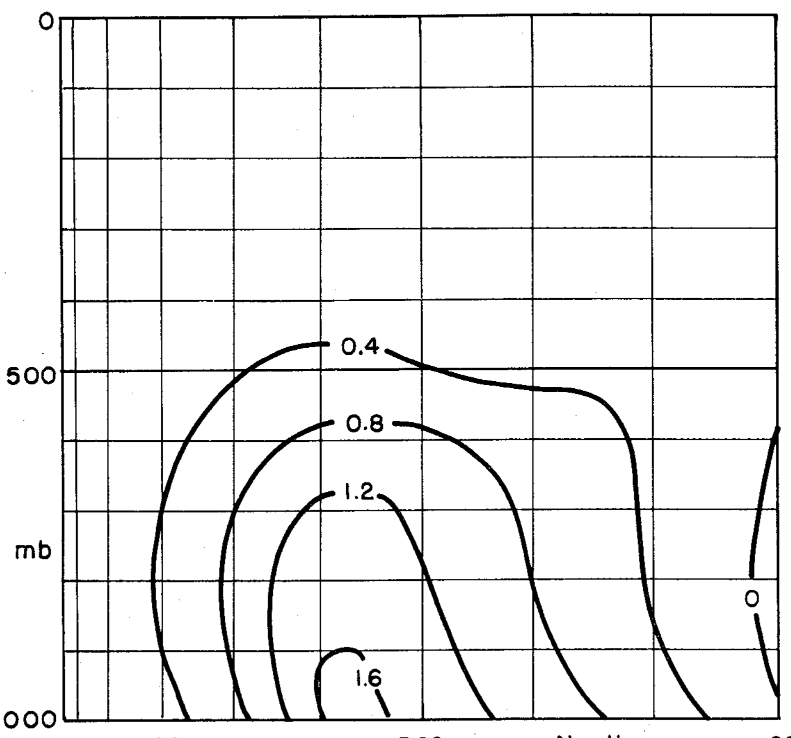


Figure 44. — The vertical distribution of the average northward transport of water by the transient motions as estimated by Peixoto and Crisi (1965). The unit is 10^{11} g sec^{-1} per 100-mb layer

toward latitudes of higher angular velocity. The weaker equatorward transport in the polar region is also countergradient. This is precisely opposite to what would be predicted by ordinary mixing-length theory. To obtain the right result, one would have to assume a negative eddy viscosity. Yet choosing a negative coefficient of viscosity for the entire atmosphere would not yield much improvement, since in the temperate latitudes the transport is directed toward latitudes of lower angular velocity. We must conclude that we are dealing with a phenomenon quite different from classical turbulence.

The water balance has been investigated by the same procedure. The computations were first carried out by Peixoto (1958) with the 1950 data, and the procedure was subsequently repeated by Peixoto and Crisi (1965) with the vastly more complete data for 1958, when 321 stations were available. Very recently, in a study not yet published, Peixoto has extended the computations to include the southern hemisphere. Fig. 42 compares the computed northward transport with the balance requirements as shown in Figure 21. Again the agreement is remarkably good at most latitudes.

Figure 43 compares the transient-eddy transport of water with the cell transport. As in the case of angular momentum, the transient eddies predominate in middle latitudes, but in the tropics the situation is quite different. Actually Peixoto and Crisi did not separate the standing-eddy transport from the cell transport, but the strong equatorward transport in the tropics appears to be due to the Hadley cell, whose lower branch is concentrated near the surface where the water vapour content is high. In fact, the entire curve is consistent with a three-cell pattern.

Figure 44 shows the vertical distribution of the horizontal transient-eddy transport of water. A notable feature is the appreciable transport at 700 mb, despite the concentration of water vapour closer to the surface. Contrary to the case of angular momentum, no countergradient flow is evident.

In considering the energy balance we should note that only the meridional circulation can transport potential energy, since the latter varies only with elevation. The direct transport of kinetic energy by either the meridional circulation or the eddies appears to be rather small. The transport of latent energy is proportional to the water transport. There remains the transport of sensible heat.

Peixoto (1060) has evaluated the eddy sensible-heat transport, again using the data for 1950. He did not compute the cell transport of either sensible heat or potential energy, since the data were quite inadequate for directly evaluating the meridional circulation.

It is nevertheless possible to estimate the long-term meridional circulation by an indirect procedure, which we shall presently describe, which makes use of previously determined values of the eddy transport of angular momentum. We have carried out the procedure, using Buch's transport values as shown in Figure 41, to estimate the meridional circulation (shown in Figure 50). We have then computed the transports of sensible heat and potential energy accomplished by this meridional circulation, using Peixoto's temperatures.

For the latter computations it is convenient to use the stream function Ψ for the mass flow. With the aid of (82) we find that the transport of any quantity X by the meridional circulation is given approximately by

$$\int_0^{p_0} 2\pi a \cos \varphi_1 \, [X] \, [v] \, g^{-1} \, dp = \int_0^{p_0} [X] \, \partial \Psi / \partial p \, dp \, . \tag{95}$$

To cope with the difficuty which arises because potential energy becomes infinite at the top of the atmosphere, we note that

$$\int_0^{p_0} [gz]\,(\partial\Psi/\partial p)\,dp = \int_0^{p_0} [RT]\,(\Psi/p)\,dp\,. \tag{96}$$

We have assumed for computation that Ψ/p remains constant above 100 mb, as it would if the northward velocity above 100 mb were uniform.

Figure 45 compares the computed total transport of sensible heat and potential energy with the balance requirement, as shown in Figure 29, while Figure 46 compares the eddy transport of sensible heat with the cell transport of sensible heat plus potential energy. Peixoto computed the standing-eddy transport for winter only; in preparing the figure we have assumed the standing-eddy transport in summer to be half as large. The meridional circulation could not be evaluated south of 10°N, but we have assumed that the Hadley cell terminates at 2°N, where Peixoto found the water transport by the cell to vanish. The transport of water across 10°N by our meridional circulation, incidentally, agrees very closely with Peixoto's value.

As in the case of the water transport, the eddies dominate in high latitudes while the Hadley cell dominates in the tropics. Although the curves in Figure 45 have certain features in common, the general agreement is no better than fair. Certainly one year of data with less than 100 stations is inadequate for transport measurements. Nevertheless, the task of determining heat exchanges between the atmosphere and the Earth is a difficult one, and the major discrepancy near 30°N might also be due to inadequate estimation of the balance requirement.

Figure 47 shows the vertical distribution of the horizontal eddy transport of sensible heat. In addition to the pronounced concentration near the surface, there is a secondary maximum in the upper

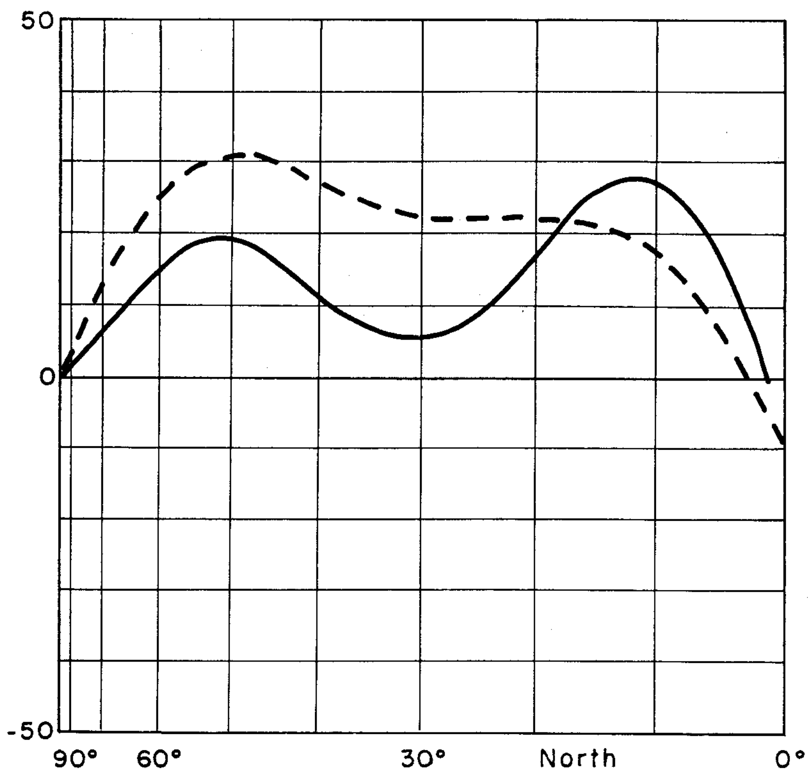
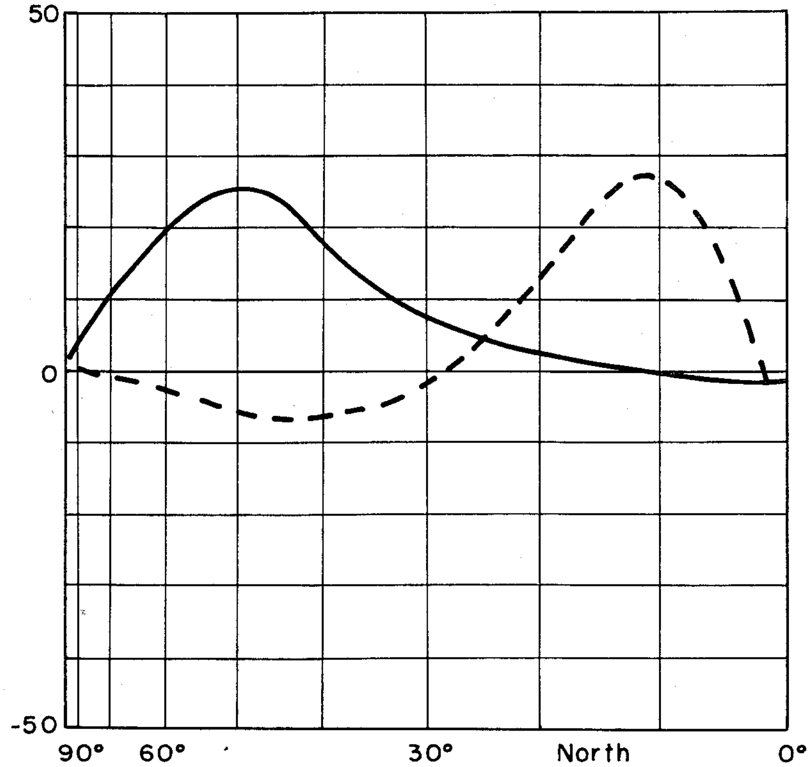


Figure 45. — The average northward transport of sensible heat plus potential energy (solid curve) as given by the sum of the curves in Figure 46, and the required transport (dashed curve) as given in Figure 21. The unit is 10^{14} watts (scale on left)

Figure 46. — The average northward transport of sensible heat by the eddies (solid curve) as estimated by Peixoto (1960), and the average northward transport of sensible heat plus potential energy by the meridional circulation (dashed curve) as determined from the meridional circulation given in Figure 50 and the temperature field estimated by Peixoto (1960). The unit is 10^{14} watts (scale on left)



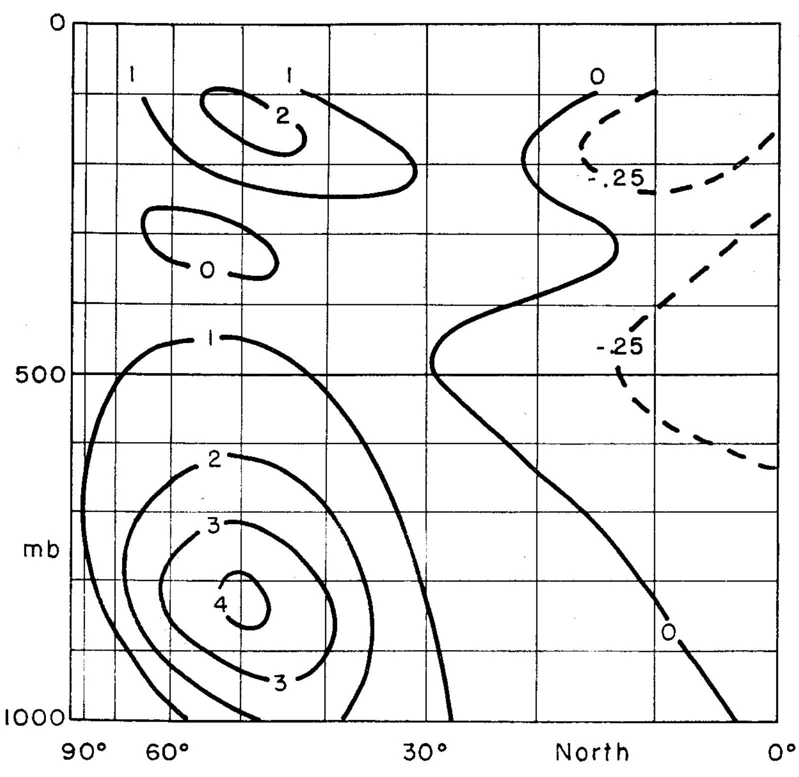


Figure 47. — The vertical distribution of the average northward transport of sensible heat by the eddies as estimated by Peixoto (1960). The unit is 10^{14} watts per 100 mb layer

troposphere. An outstanding feature is the countergradient flux throughout middle latitudes in the lower stratosphere, previously detected by White (1954) from geostrophic computations. There is also a decided countergradient flux in the tropics in the middle troposphere. These fluxes are further indications of the inadequacy of mixing-length concepts in explaining the transport processes. We are led to ask whether it is simply coincidence that turbulence theory yields the proper sign for the sensible-heat transport as often as it does.

Unlike some of the earlier studies, the ones we have just described make no use of geostrophically estimated winds. In this respect they have the obvious advantage of not depending upon an approximation which, while rather good on a point-by-point basis, could yet lead to systematic errors in the correlation of v with u, T, or q. Nevertheless, observed-wind studies seem to be more seriously affected by missing data than geostrophic-wind studies. At nearly every station some observations are missing at higher levels. As noted in the previous chapter, one of the principal reasons for missing upper-level reports is strong winds, which carry the balloon beyond the range of the observing instrument. Thus most collections of upper-level observed-wind data are biased in favour of light winds.

This bias affects the computed transports of angular momentum most seriously, since the missing observations are likely to possess extreme values of u, the transported quantity, as well as v. Recently Priestley and Troup (1964) investigated the effect of this bias by evaluating $\overline{u'v'}$ at a few stations where there were virtually no missing observations, and then determining how these values would have been altered if a few observations with the strongest winds had been missing. They found that omission of even ten per cent of the observations could drastically alter the computed value of $\overline{u'v'}$, and perhaps even reverse the sign. Their study incidentally reveals the importance of the jet stream in accomplishing the necessary transport.

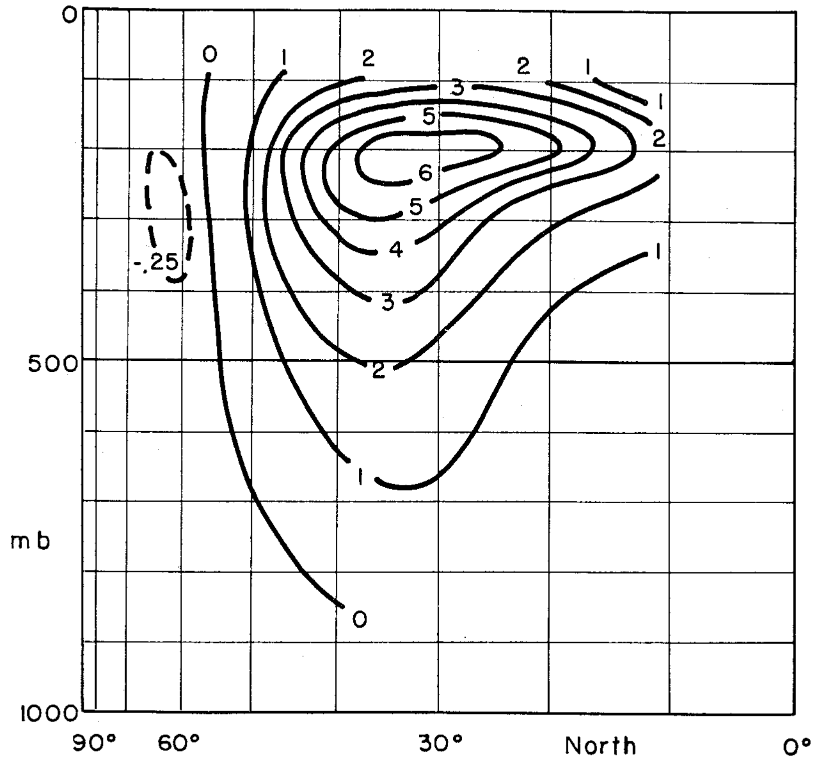
We shall therefore compare the computed transports of angular momentum with those of other investigations which may be less subject to the light-wind bias. The study using the largest sample of data is the recent one by Holopainen (1966); it is again based upon the charts compiled by Crutcher (1959). As we have noted, these charts include maps of $\bar{u}$ and $\bar{v}$ and the standard deviations of u and v at six levels; they also include maps of the correlation between u and v. From the correlations, $\overline{u'v'}$ and hence the transient-eddy transport of angular momentum may be computed, while the standing-eddy transport may be determined from the maps of $\bar{u}$ and $\bar{v}$.

Figure 48 shows the eddy-transport of angular momentum as determined by Holopainen; it is to be compared with Figure 41 showing Buch's values. There is rather good agreement; the principal difference is that Holopainen's extreme values are noticeably farther north.

Crutcher's charts were based upon observed winds wherever these were plentiful, and gradient winds where observations were scarce. On this account they may be subject to a light-wind bias, but to a lesser extent than Buch's. The computations which are least subject to the light-wind bias are those of Mintz (1955), in which all the winds have been geostrophically estimated. Figure 49 shows Mintz's results; the data are for two winter and two summer months in 1949. Again there is good qualitative agreement. Aside from the noticeably larger values, the chief feature is the absence of the much weaker transport at 100 mb than at 200 mb which appeared in the other studies. It is difficult to say whether this discrepancy results from the light-wind bias, which should be greatest at highest levels, from the use of the geostrophic-wind approximation and the inadequacy of the 1949 data, or simply because the studies used data from different years.

The principal discrepancy between the estimated transports and the estimated balance requirements occurs in the case of the energy balance (see Figure 45). Mintz (1955) obtained a curve for the eddy

Figure 48. — The vertical distribution of the average northward transport of angular momentum by the eddies as estimated by Holopainen (1966). The unit is 10^{25} g cm^2 sec^{-2} per 100 mb layer



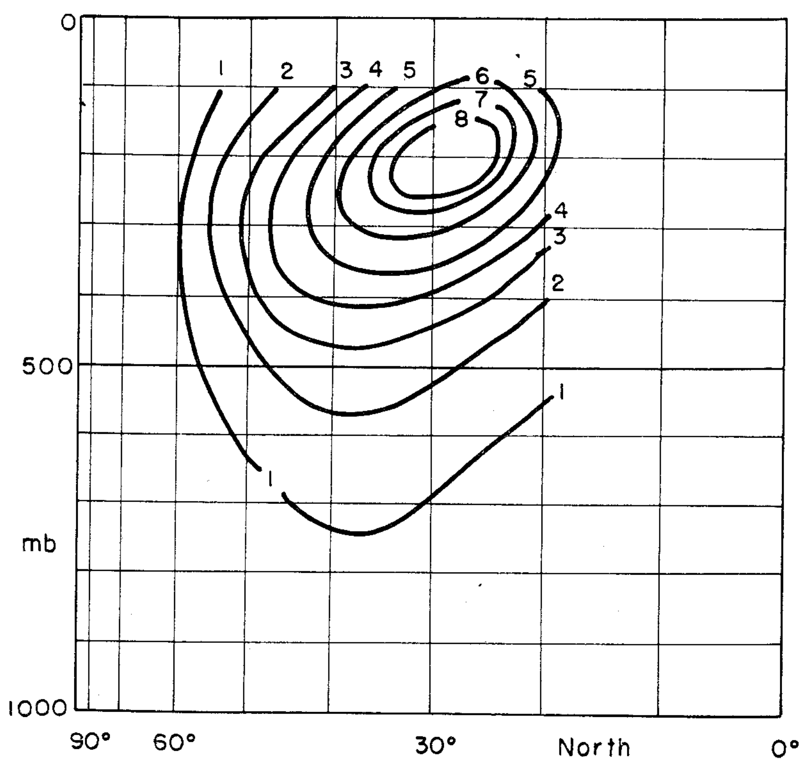


Figure 49. — The vertical distribution of the average northward transport of angular momentum by the eddies as estimated by Mintz (1955). The unit is 10^{25} g cm^2 sec^{-2} per 100 mb layer

transport of sensible heat looking very much like Peixoto's, but generally 20 per cent higher; it indicates a transport of 10×10^{14} rather than 8×10^{14} watts across 30°N. Starr and White (1954) obtained a value of 12×10^{14} watts. Likewise, the Hadley circulation deduced by Holopainen (1966) from the momentum transports shown in Figure 48 extends somewhat north of 30°N, and yields a transport of $+3 \times 10^{14}$ rather than -2×10^{14} watts. Saltzman *et al.* (1961) have estimated the direct transport of kinetic energy on the basis of 500-mb data only; it appears sufficient to add 1×10^{14} watts, and possibly more if the actual transport is concentrated near the tropopause. Nevertheless, it seems likely that the balance requirement of 22×10^{14} watts may have been overestimated; if the radiation balance and the water balance have been correctly estimated, a larger amount of energy would then be left to be transported by the oceans. Much closer agreement than indicated in Figure 45 has also been obtained by Holopainen (1965) and Palmén and Newton (1967).

The general agreement among most of the computed values is encouraging, but quantitatively we cannot regard the present pictures of the angular-momentum, water, and energy balances as the final word. When more densely distributed and more complete data finally become available, some of our estimates may be altered by as much as fifty per cent.

The vertical transports

The foregoing computations show that throughout middle and higher latitudes the required horizontal transports of angular momentum, water, and energy are accomplished mainly by large-scale eddies — the systems which were missing in the early theories of the general circulation. Only in the tropics does the meridional circulation play a dominant role. There remains the question of the vertical transports. In this respect the meridional circulation assumes an added importance.

The balance of angular momentum presents the most straightforward problem. Although the atmosphere gains or loses angular momentum only through direct contact with the ground, the horizontal transport needed to balance the exchange with the ground takes place mainly in the upper troposphere. There must therefore be a vertical transport of angular momentum within the atmosphere across intermediate elevations. This transport would appear to be upward in the tropics and downward in middle latitudes, with a slight upward flow again in the polar regions.

Here a word of caution is necessary. Angular momentum, as we have noted, may be expressed as the sum of Ω-momentum and relative momentum. In evaluating the total horizontal transport, we may disregard Ω-momentum. If, however, either the vertical transport or the vertical variation of the horizontal transport is of interest, Ω-momentum cannot be disregarded when even a weak meridional circulation is present.

Consider, for example, a direct Hadley cell confined between the Equator and 30°N. Such a cell would transport no momentum across 30°N, and would therefore not alter the need for an upward transport within the tropics. However, because Ω-momentum per unit mass decreases with increasing latitude, the vertical motion would carry a larger amount of Ω-momentum upward between the Equator and 15°N than downward between 15°N and 30°N. A cell of sufficient strength could therefore accomplish the required total upward transport. At the same time the horizontal motion would carry an intermediate amount of Ω-momentum northward across 15°N above 500 mb, and a similar amount southward across 15°N below 500 mb. The net effect would therefore be an increase of angular momentum above 500 mb, and a decrease below 500 mb, at all latitudes within the cell. Effectively, then, the cell would carry absolute angular momentum upward at all latitudes. By such means, a direct cell could accomplish a large part of the needed vertical transport at each latitude in the tropics. Likewise an indirect cell could serve a similar purpose in middle latitudes. The middle-latitude cell could be weaker, since the poleward gradient of Ω-momentum is greater there. Such cells would of course carry relative angular momentum at the same time.

The direct computation of vertical transport by the eddies, using observed vertical velocities, is impossible because vertical velocities are not observed on a global scale, even if we accept the vertical velocity deduced through the equation of continuity from the horizontal divergence as being observed. Estimates of the vertical velocity field are therefore necessarily indirect, and are based upon observations of the more readily observed non-divergent horizontal velocities and temperatures.

Three methods of deducing the vertical velocity are feasible. First there is the "adiabatic" method, based upon the thermodynamic equation. Here the field of potential temperature is assumed to be altered only by horizontal and vertical advection. The fields of horizontal advection and local change are observed, and the vertical velocity needed to achieve a balance is deduced. The procedure can be modified by introducing sources and sinks of heat, if these are known.

A somewhat analogous procedure is based upon the vorticity equation, frequently a simplified form such as (64). Again the fields of horizontal advection and local change are evaluated, and the vertical motion needed to make the remaining terms balance is deduced.

A third method uses the ω-equation, generally in a simplified form such as (73). No local changes need be evaluated, but the temperature advection and vorticity advection are both measured. These by themselves would destroy the previously existing geostrophic equilibrium. The vertical motion field and its accompanying field of horizontal divergence are assumed to be those needed to maintain geostrophic equilibrium by compensating for the effects of horizontal advection.

Using the field of vertical motion deduced by any one of these methods, it is possible to compute the vertical transport of angular momentum as a function of latitude and elevation. It is well to ask at this point what could be accomplished by such a computation.

We have noted that our direct estimates of the exchange of angular momentum between the atmosphere and the Earth are rather crude in view of our uncertainty concerning the laws of surface friction and our general failure to incorporate the effects of mountain ranges. Our estimates of the balance requirement is thus correspondingly uncertain. Direct measurement of the horizontal transport has eliminated some of this uncertainty. In addition to obtaining the intellectual satisfaction of having deduced a result by an independent method, we can place further confidence in the assumed numerical values of the surface torque.

The situation concerning the computed vertical transport is different. The deduced fields of vertical motion are at best extremely uncertain as compared to the observed fields of horizontal motion. Neglect of the effects of heating or friction will lead to incorrectly deduced vertical velocities. Additional errors can arise because the time derivatives are not well approximated by the observed 12-hour or 24-hour changes. Although it would be gratifying if the computed vertical transport should be compatible with the horizontal transport, it must be recognized that any disagreement would undoubtedly be attributed to inadequacy of the deduced vertical motions, and the vertical transport needed to balance the computed horizontal transport would still be regarded as a better estimate.

Moreover, the satisfaction of having obtained the result by independent means would be less certain, since in any event the result would not be deduced from independent data; the data used to estimate the vertical motions are the same as those used to compute the horizontal transports. It is even possible to estimate the vertical motion by a procedure based on the vorticity equation which will automatically require the computed vertical transports to agree with the computed horizontal transports.

What the computed vertical transport may indicate is the relative importance of the eddies and the meridional circulation. We mention a recent study by Starr and Dickinson (1964). Here the vertical motions were evaluated by the adiabatic method, so that there was no *a priori* reason why the computed vertical transport of momentum would have to agree with the previously computed horizontal transport. Their results indicated that the eddies were rather ineffective in transporting momentum vertically, implying that the needed vertical transport must be accomplished mainly by the cells.

Palmén and Newton (1967), on the other hand, have evaluated the vertical transport of angular momentum by the cells from the observed meridional circulation in the northern hemisphere in winter, as shown in Figure 18. They have then estimated the vertical transport by the eddies as a residual term. Again, the major portion of the vertical transport is accomplished by the cells. There appears to be a small downward eddy transport in tropical and middle latitudes. As Palmén and Newton point out, there is no way to determine from these computations what portion of the eddy transport is accomplished by cyclone-scale eddies, and what portion is accomplished by motions of much smaller scale.

The vertical transport of water requires other considerations. Since the water which returns to the Earth as precipitation falls from some intermediate elevation, rather than being transported to the ground by the motion of the atmosphere, there must be a net upward transport of water within the atmosphere as a whole.

The principal point to observe is that there is considerable precipitation even in those latitudes where evaporation exceeds precipitation, and except close to the Poles there is considerable evaporation even in those latitudes where precipitation exceeds evaporation. At each latitude the water entering the atmosphere by evaporation must be transported upward to the levels from which precipitation falls, unless it is removed by a strong divergence of horizontal transport in the lower layers. In the latter event there must be a strong convergence of horizontal transport aloft to supply the water which falls as precipitation. It does not appear that the meridional circulation can bring about this convergence, since the lower branches of the cells seem to be concentrated below the levels from which precipitation falls, while the upper branches occur above most of the water. Reference to Figure 44 reveals no convergence of eddy transport above a region of divergence, or above a region where the transport by the meridional circulation may be expected to diverge.

We are forced to conclude that there is an upward transport of water at all latitudes. This may be accomplished by the Hadley cells in the tropics, and possibly by indirect cells at higher latitudes, but between 20°N and 40°N, where the cell motion is downward, the upward transport must be accomplished by the eddies.

This eddy transport could be accomplished either by cyclones and anticyclones or by cumuliform convection, since in either type of system it is the moist air which rises and the dry air which sinks. With a knowledge of the levels from which the precipitation falls, we could deduce the vertical transport of water as a function of latitude and elevation.

The vertical transport of total energy presents further complications because of the presence of both sources and sinks at various levels in the atmosphere. Palmén and Newton have estimated the vertical transport of sensible heat across the 500-mb surface by a procedure similar to the one by which they estimated the vertical transport of angular momentum. Using the distribution of radiative heat sources and sinks as given by London (1957), the horizontal eddy-transports of sensible heat determined by Mintz (1955) and the meridional circulation of Palmén and Vuorela (1963), and partitioning the total release of latent heat into the portions above and below 500 mb by a procedure based upon a consideration of moist-adiabatic ascent of air, they have obtained the vertical eddy-transport as a residual term. They find an upward transport in lower and middle latitudes which appears to be more important than the transport by the meridional circulation. Again, there is no indication as to the scale of the eddies, but they note that, particularly in the tropics, cumuliform convection may be of major importance.

An alternative method of partitioning the total release of latent heat, based upon cloud observations, has been used by Davis (1963), who finds much smaller amounts released above 500 mb. His results would imply even larger upward eddy-transports of sensible heat.

Consequences of the transport processes

The eddy transports of angular momentum and sensible heat are the missing elements in the early theories of the general circulation. From the point of view of the zonally averaged circulation, the convergence of the eddy transport of sensible heat acts as a heat source, additional to the heating by radiation and small-scale turbulent conduction. A convergence of the eddy transport of angular momentum acts as a mechanical force, additional to surface friction and small-scale turbulent viscosity. Without these transports the three-cell structure of the meridional circulation cannot be explained.

Hadley observed that the primary effect of the low-latitude heating and high-latitude cooling would be to force a single direct meridional cell in each hemisphere. The cell would in turn necessitate eastward and westward motions, and consequently eastward and westward frictional drags at the surface. Thomson and Ferrel deduced that the frictional drag upon the surface westerlies would force an additional indirect cell in each hemisphere; Thomson and subsequently Ferrel decided that this cell should be confined mainly to the lower layers.

In an elegant treatment, Eliassen (1952) deduced the combined effects of predetermined heating and mechanical forcing upon the steady meridional motion superposed upon a general circular vortex. His equation was essentially a special case of the more recently introduced ω-equation. He found that local heating would force a flow upward along a surface of constant absolute angular momentum, while eastward mechanical forcing would force a flow outward (equatorward) along a surface of constant potential temperature.

Kuo (1956) applied this approach to the atmosphere, and found that a three-cell pattern was demanded. If the eddies, no matter how intense, transported no angular momentum nor sensible heat, the meridional circulations might be more or less as deduced by Thomson and Ferrel. The supposedly irrelevant disturbances would be irrelevant indeed. With the transports as shown in Figures 41 and 47 the situation is different. The strong divergence of angular-momentum transport in low latitudes, particularly at the jet-stream level, and the strong convergence in middle latitudes force converging poleward and equatorward currents at high levels. These meet in the subtropics and descend to form a direct low-latitude cell and an indirect middle-latitude cell, which are superposed upon the meridional motion which would otherwise exist. The weaker divergence of angular-momentum transport in the polar regions gives rise to a high-latitude direct cell. In addition, the divergence of sensible-heat transport extending well into middle latitudes, and the convergence closer to the Pole, force upward motion in higher middle latitudes and downward motion in lower latitudes, further intensifying the indirect cell.

Just as individual vertical velocities may be estimated from the vorticity equation or the thermodynamic equation alone instead of the ω-equation, so the meridional circulation may be deduced from the momentum transports or the energy transports alone. In short, any convergence of eddy transport of angular momentum, or energy, which is not balanced by friction, or heating, must be balanced by the cell transport. In view of the possibility of significant unknown vertical eddy transports of energy and the general uncertainty as to the distribution of heating, the most reliable results should be obtained from the angular-momentum transport data.

The procedure was first used by Mintz and Lang (1955). In brief, if the absolute angular momentum is known on each of four sides of a "rectangle", say ACDB in the upper left of Figure 50, if the horizontal eddy-transport of angular momentum through sides AB and CD is known, if there is assumed to be no vertical eddy-transport or frictional transfer through sides AC and BD, and if the flow of mass through two adjacent sides AB and AC is known, the flow of mass through sides BD and CD is easily deduced from continuity considerations. Starting at the upper corner of Figure 50, with the boundary condition that no mass flows across the top of the atmosphere or the "90th parallel", we can evaluate the mass flow across any segment if the complete field of horizontal eddy transport is known. In the friction layer, where the above assumptions are no longer valid, the circulation is deduced from mass continuity.

Mintz and Lang used the geostrophically estimated angular-momentum transport values which Mintz (1955) had previously determined, and they assumed no mass flow across the 200-mb level. Because of the crudeness of some of their assumptions they regarded their result as a model of the meridional circulation rather than an evaluation of it. Nevertheless, it appears to be as reasonable as any estimate which was then available. The same procedure has since been used by Holopainen (1966).

The procedure becomes even simpler if A and B, and also C and D, instead of lying on a vertical line, lie on a nearly vertical line of constant absolute angular momentum. One may then begin at any latitude and work downward from the top of the atmosphere. We have carried out modified procedure using Buch's values of the angular-momentum transport and assuming that the friction layer extends to 850 mb. The resulting meridional circulation is shown in Figure 50. It contains well developed Hadley and Ferrel cells, and it is very much like the winter circulation of Mintz and Lang, except that both cells appear farther south.

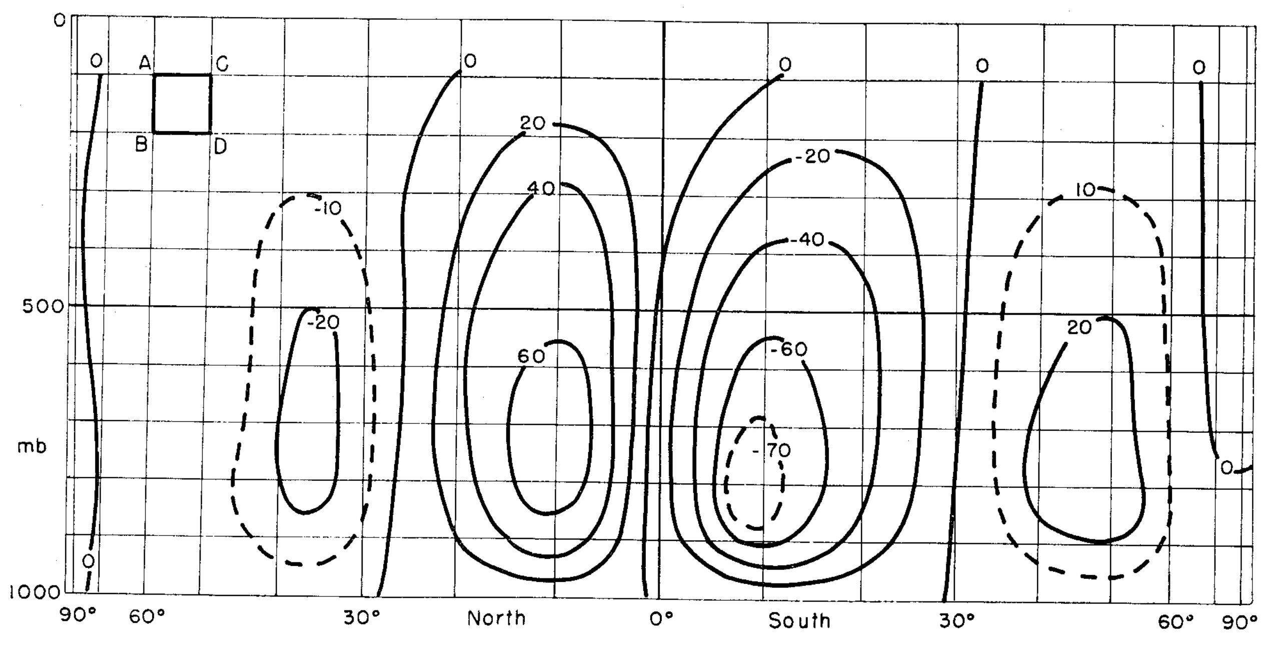


Figure 50. — The average meridional circulation as determined indirectly from the eddy-transport of angular momentum given in Figure 41. The southern hemisphere estimate is by Gilman (1965). The unit for stream function $\overline{\Psi}$ is 10^{12} g sec^{-1}

Gilman (1965) has computed the summer and winter meridional circulations for the southern hemisphere by the modified procedure, using the transport values of Obasi. The average of his summer and winter circulations is also shown in Figure 50. Southern hemisphere data are still quite inadequate for a direct computation of the meridional circulation, and Gilman's indirect computation represents the best picture so far available. Near the Equator the procedure breaks down, and we have joined the two hemispheres by simple continuity, together with the assumption that the opposing currents should meet slightly north of the Equator.

Having obtained a picture of the horizontal and vertical transports of angular momentum, water and energy by the eddies and the meridional cells, we may now ask what is implied concerning the fields of the transported quantities, or, alternatively, the fields of zonal motion, specific humidity, and temperature. It is sometimes stated that the strong upper-level westerly winds are maintained by a convergence of the horizontal transport of angular momentum. In a sense this statement is true; there is convergence of the horizontal transport where the westerlies reach their maximum. Yet there is no simple relation through which the westerly wind speed may be deduced from the field of angular-momentum transport.

Since in the long run any convergence of the horizontal or vertical transport of angular momentum must be balanced by turbulent friction, the time-averaged field of friction may be deduced from the field of angular-momentum transport. The westerly-wind field may therefore be deduced from the angular momentum transport field only to the extent that it may be deduced from the field of friction.

It is reasonable to believe that friction tends to diminish the westerly winds in the regions where they are strongest, but for the atmosphere as a whole the precise relation between wind speed and turbulent friction, if one exists, is but poorly known. The most obvious instance of a direct relation between wind speed and friction occurs at the Earth's surface, where the drag opposes the surface wind at a rate which, if not exactly determined by the wind speed, at least tends to increase with increasing wind speed. The long-term convergence of the vertically integrated horizontal transport of angular momentum in middle latitudes therefore demands a westerly surface wind in these latitudes; if no surface westerlies were present, and if the convergence of the transport persisted, angular momentum would continue to accumulate in middle latitudes. Since the thermal-wind relation would continue to be approximately satisfied, and the horizontal temperature gradient would not become infinite, westerlies would ultimately appear at the surface. The easterlies in low latitudes are similarly demanded by the divergence of the horizontal transport of angular momentum. Of course, the processes required to maintain geostrophic balance might themselves alter the transport of angular momentum. It is only if we postulate that the convergence of angular momentum transport must continue that we can deduce that surface westerlies must appear and persist.

Similar considerations apply to the transport of water. A net convergence of the total water transport does not imply a high value of specific humidity; it implies precipitation in excess of the gain of water by evaporation and turbulent transfer. To a first approximation it may imply a high average relative humidity. Without a knowledge of the temperature field as well, a complete knowledge of the horizontal and vertical transport of water would tell us rather little about the distribution of specific humidity.

The horizontal and vertical transport of energy may be more revealing. A convergence of the transport requires a net loss of energy by radiation and turbulent conduction. To some extent, at least, the amount of energy lost by radiation depends upon the temperature, although the presence of water vapour and clouds can complicate the picture. With a continued convergence of the transport of energy, the temperature should therefore rise until such a time as the increased effect of radiation and conduction can balance the effect of the transport. The air should therefore be warm, or at least warmer than would be indicated by radiation considerations alone.

A complete knowledge of the field of energy transport should therefore give a fair first approximation to the horizontal and vertical distribution of temperature. A knowledge of the field of angular-momentum transport would indicate the distribution of surface easterlies and westerlies, but would by itself give a rather poor indication of the westerly wind field aloft. With a knowledge of the transports of both energy and angular momentum, we should be able to infer the upper-level wind field from the fields of temperature and surface wind, using the thermal wind relation. The procedure would be most satisfactory in the case of an idealized dry atmosphere. For the real atmosphere we should not expect the fields so deduced to be very realistic, because of the complicating effects of water upon absorption and emission of radiation.

For the idealized atmosphere, an explanation of the transport processes would therefore amount to an explanation of the zonally averaged circulation. Nevertheless, since the transport processes are themselves affected by field of motion, an independent explanation of the transport processes which could then be used to explain the zonally averaged motion does not appear possible. On the other hand, any complete explanation of the zonally averaged circulation must contain, explicitly or implicitly, an explanation of the transport processes.

CHAPTER V

THE ENERGETICS OF THE ATMOSPHERE

In his famous account of the trade winds and monsoons, Halley (1686) identified the sun as the cause of the motions of the atmosphere. Although the bulk of Halley's theory is no longer regarded with favour, it is still generally accepted that the ultimate source of atmospheric energy is the sun. The direct effect of solar radiation is to heat the atmosphere and the underlying ocean and land, and thereby produce internal energy. The motions of the atmosphere, on the other hand, represent a great supply of kinetic energy. This supply is being continually dissipated by friction. One of the main problems in general-circulation theory concerns the manner in which some of the internal energy produced by solar heating is ultimately converted into kinetic energy to replenish the supply thereof.

In the previous chapter we examined in some detail the maintenance of the spatial distributions of wind velocity, temperature, and moisture. In the present chapter we shall examine in further detail the manner in which the total amounts of kinetic, potential, and internal energy represented by wind, temperature, and moisture fields are maintained, but we shall be only incidentally concerned with the geographical distributions of these fields. In all probability we cannot completely explain the maintenance of the total amounts without explaining the geographical distributions as well. Nevertheless, by considering the production, transformation, and dissipation of energy separately from the remaining aspects of the circulation, we may gain further insight into the roles played by some of the physical processes.

Two fundamental quantities to be considered are the rate at which solar energy reaches the extremity of the atmosphere, and the rate at which new kinetic energy must be produced to offset the dissipative effects of friction. The former rate is observed to be about 1.8×10^{17} watts, or on the average about 350 watts per square metre of the Earth's surface. Various estimates place the latter rate at about one hundredth of the former. If the atmosphere is regarded as a heat engine, producing kinetic energy, the ratio η of these rates, about one per cent, is a measure of its efficiency. The determination and explanation of the efficiency η constitute the fundamental observational and theoretical problems of atmospheric energetics.

Basic energy forms and conversions

Since the bulk of the incoming solar radiation heats the underlying ocean or land instead of heating the atmosphere directly, we need to examine the energetics of the atmosphere-ocean-Earth system, or at least that part of the system which directly or indirectly exchanges significant amounts of energy with the atmosphere. We may disregard the hot interior of the Earth, since the heat received from it is negligible, except locally in regions of volcanic activity. We may likewise disregard the deep oceans, although we should recognize the possibility that heat stored there may reach the surface years later through slow overturning, and influence the long-period atmospheric fluctuations (see Rossby 1957).

The forms of energy which play a significant role are kinetic energy (KE), potential energy (PE), and internal energy (IE). Thermodynamically both thermal internal energy and the latent energy of condensation and fusion of water are forms of IE, but sometimes they are more conveniently treated as separate forms of energy. Some writers prefer to treat the kinetic energy of small-scale turbulent motions as a form distinct from *KE*. In the present treatment, where the small-scale motion is not regarded as a part of the circulation, and where no distinction is made between turbulent and molecular friction, it seems most logical to treat turbulent kinetic energy as neither *KE* nor a form by itself, but as a portion of the *IE*, thus effectively grouping the kinetic energy of small-scale motions with the kinetic energy of molecular motions.

Other forms of energy are not directly or indirectly converted into *KE*, *PE*, or *IE* in large amounts, although they may be important on a local scale. The electrical energy converted into *IE* through lightning discharges may, for example, play an important part in the dynamics of thunderstorms and tornadoes, but the total amount of electrical energy in the atmosphere is minor. By contrast, there is a vast supply of nuclear energy, but, fortunately for humanity, the natural processes for releasing it are virtually absent.

The atmosphere-ocean-Earth system exchanges total energy with its environment only through radiation. In so doing, the system gains or loses only *IE*. Since the system does not undergo any net long-term change in total energy, the heating by incoming radiation must in the long run balance the cooling by outgoing radiation.

Within the atmosphere-ocean-Earth system, *IE* may be transferred from one location to another, and in particular from the atmosphere to the underlying surface or vice versa, through radiation and conduction. Again, the net heating of the system by these processes is zero.

Only those processes involving a force can produce or destroy *KE*. Motion of the atmosphere (or the ocean) with or against the force of gravity, and hence downward or upward, converts *PE* into *KE*, or *KE* into *PE*. The process is adiabatic and thermodynamically reversible. *KE* is the only immediate source or sink for *PE*.

Likewise, motion of the atmosphere with or against the pressure-gradient force, and hence across the isobaric surfaces toward lower or higher pressure, converts *IE* into *KE*, or *KE* into *IE*. Again the process is reversible and adiabatic. Motion of the atmosphere against the force of friction, and the frictional heating which accompanies it, also convert *KE* into *IE*. By its very nature the process is irreversible, since friction must on the average oppose the motion. The frictional heating produces the necessary increase in entropy. The only remaining force, the Coriolis force, acts at right-angles to the motion and does not add or remove *KE*.

It follows that the conversion of *IE* into *KE* by the pressure-gradient force, although thermodynamically reversible in that it can proceed equally well in either direction, does not proceed to the same extent in either direction. In the long run it produces as much *KE* as is dissipated by friction.

It also follows, since there is no long-term net heating by radiation and conduction, and since the remaining energy-conversion processes other than friction involve no heating at all, that the net total heating of the atmosphere-ocean-Earth system equals the net frictional heating. The total heating of the system is therefore positive, not zero.

If a distinction is made between the thermal and latent forms of *IE*, the processes of evaporation and melting and the reverse processes of condensation and freezing convert thermal *IE* into latent *IE*, and vice versa. In particular, evaporation from the ocean surface removes thermal *IE* from the ocean and adds latent *IE* to the atmosphere. It is possible, however, not to include latent *IE* as a form of atmospheric energy, provided that the release of latent heat, which must inevitably occur regardless of the attitude which one takes toward latent energy, is treated as a special form of heating by the environment, rather than an internal quasi-adiabatic process. If this convention is adopted, the atmosphere will be assumed to gain IE not when water evaporates from the ocean, but when the water subsequently condenses within the atmosphere.

In this manner one may treat the energetics of the atmosphere by itself. Aside from surface friction, the total influence of the environment upon the atmosphere may be treated as the addition and removal of equal amounts of thermal IE by heating and cooling. Surface friction need not be distinguished from internal friction, since both processes convert KE into IE irreversibly. The only remaining conversions from one form to another are then the reversible processes within the atmosphere, involving KE and PE, or KE and IE. Throughout the remainder of this chapter we shall consider the energetics of the atmosphere alone, from this point of view.

The fact that the atmosphere remains very nearly in hydrostatic equilibrium places certain constraints upon the energy conversion processes which may actually occur. When, for example, heating adds IE to the atmosphere, upward expansion occurs, and the upward motion converts some IE into KE and an equal amount of KE into PE. It is easily shown that under hydrostatic equilibrium the PE contained in a vertical column extending throughout the depth of the atmosphere is proportional to the IE, in the ratio R/c_v, or approximately 2/5, although the result is strictly true only for a dry atmosphere extending upward from sea-level. The amounts of PE and IE per unit mass are given by gz and c_vT, while an element of mass of a column of unit cross-section is ρdz; thus

$$\int_0^\infty gz\rho dz = \int_0^{p_0} zdp = \int_0^\infty pdz = \int_0^\infty RT\rho dz\,, \tag{97}$$

where p_0 is the surface pressure. Thus PE and IE increase or decrease together, and it is convenient to regard them as a single form of energy, called total potential energy (TPE) by Margules (1903). It is meaningless to speak of the TPE at a particular point, but within a vertical column the average amount of TPE per unit mass is given by the average value of c_pT, which is simply the sensible heat per unit mass.

It follows that whenever the horizontal motions by themselves convert seven units of IE into KE, the vertical motions which must accompany them in order to maintain hydrostatic equilibrium convert two units of KE into IE, and two units of PE into KE. The net result is a conversion of five units of IE and two units of PE, or seven units of TPE, into KE. Thus the vertical motions alone do not alter KE or TPE. Effectively, horizontal motion of the atmosphere with or against the pressure-gradient force, and hence across the isobars toward lower or higher pressure, converts TPE into KE, or KE into TPE. This is now the only conversion process which we need to consider, in view of the convention concerning latent energy which we have adopted.

Mathematical expressions for the conversion processes may be obtained from the equations presented in Chapter II. If $\{X\}$ denotes the total amount within the atmosphere of a quantity whose value per unit mass is denoted by X, and if, as in the previous chapter, $\overline{X}$ denotes a long-term time average of X, it follows that $\{dX/dt\} = \partial\{X\}/\partial t$. Hence, from (40) and (41),

$$\partial\{\Phi + I\}/\partial t = H - C\,, \tag{98}$$

$$\partial\{K\}/\partial t = C - D\,, \tag{99}$$

where

$$H = \{Q\} \tag{100}$$

is the total heating of the atmosphere,

$$D = -\{\mathbf{U}\cdot\mathbf{F}\} \tag{101}$$

is the total dissipation, and C is the rate of conversion of TPE into KE by reversible adiabatic processes, given by the alternative expressions

$$C = -\{\omega\alpha\} = g\{\omega\partial z/\partial p\} = -g\{z\partial\omega/\partial p\} = g\{z\nabla\cdot\mathbf{U}\} = -g\{\mathbf{U}\cdot\nabla z\}. \tag{102}$$

It is to be observed that no matter which expression is used for C, only the divergent part of the wind is involved.

It follows that the long-term averages $\overline{H}$, $\overline{C}$, and $\overline{D}$ are equal, and, as previously noted, are positive, since D is always positive. Further restrictions upon the field of Q follow because there are no net long-term changes of entropy and potential temperature. From (15), (14), and (100),

$$\overline{\{Q/T\}} = 0\,, \quad \overline{\{Q/p^{\kappa}\}} = 0\,, \quad \overline{\{Q\}} = \overline{H} > 0\,. \tag{103}$$

Since T and p^{κ} are necessarily positive, it follows that Q is negatively correlated with $1/T$ and $1/p^{\kappa}$.

This statement applies equally well if Q is replaced by the non-frictional heating $Q_n = Q - Q_f$, since the frictional heating Q_f, is positive everywhere. Thus

$$\overline{\{Q_n/T\}} < 0\,, \quad \overline{\{Q_n/p^{\kappa}\}} < 0\,, \quad \overline{\{Q_n\}} = 0\,. \tag{104}$$

Within the limited sense of these inequalities, the heating must occur at a higher temperature and a higher pressure than the cooling.

The process which converts TPE into KE, represented by $C = -\{\omega\alpha\}$, is often colloquially described as a sinking of colder air and a rising of warmer air at the same elevation. This interpretation seems reasonable in view of the strong negative correlation between ω and w. Certainly it is correct for the case of a simple Hadley circulation. Yet $\{\omega\alpha\}$ cannot be converted without further approximation into an expression involving w and T alone, and the interpretation cannot be rigorously defended.

We have previously identified the measurement of the efficiency η, or equivalently the determination of $\overline{H}$, $\overline{C}$, or $\overline{D}$, as the fundamental observational problem of atmospheric energetics. Early estimates of the efficiency, which ranged as high as 20 per cent were estimates of the classical thermodynamic efficiency, which is considerably greater than η. The thermodynamic efficiency may be defined as the ratio of the net heating to the heating at the heat source. In this ratio the numerator may be identified with the net heating of the atmosphere, but the denominator is not the net solar heating. It is probably best identified with the difference between the incoming and outgoing radiation, summed over only those regions where this difference is positive. It follows that the classical thermodynamic efficiency exceeds η by a factor of perhaps 4 or 5. The thermodynamic efficiency cannot exceed the ratio of the temperature difference between the heat source and the heat sink to the temperature of the heat source, whence an upper bound can be estimated from the temperature field alone.

Direct evaluation of $\overline{D}$ is difficult, in view of our inadequate knowledge of friction, especially at higher elevations. An early estimate by Sverdrup (1917) was 1.3×10^{15} watts, or 2.55 watts per square metre of the Earth's surface, equivalent to a value of η of 0.007. This value was based upon empirically determined coefficients of viscosity in the relatively steady trade winds.

Brunt (1920) estimated a value of 5 watts m^{-2}, which would make $\eta = 0.014$; this he obtained by computing a value of 3 watts m^{-2} in the surface friction layer, and assuming that the remainder of the atmosphere ought to contribute nearly as much. For some time Brunt's value was regarded as the best estimate attainable. Subsequent estimates were substantially lower, and in a comprehensive review of previous estimates, Oort (1964a) chose 2.3 watts m^{-2} as the most reasonable value.

Direct evaluation of $\overline{C}$ requires measurement of the divergent part of the wind in one manner or another, which is also difficult to perform accurately. Nevertheless, it would appear to offer an independent method of determining η.

Examination of Sverdrup's procedure reveals, however, that he based his coefficients of viscosity on observations of cross-isobar flow in the trade winds. Likewise, Brunt's estimate was based upon typical cross-isobar flow in the friction layer. It thus appears that both Brunt and Sverdrup were actually evaluating $\overline{C}$ rather than $\overline{D}$.

The very recent estimates of $\overline{D}$ by Holopainen (1963) and Kung (1966) are also based upon evaluation of the cross-isobar flow $-g\mathbf{U}\cdot\nabla z$. The procedure is of considerable interest. Holopainen used only a few weeks of winter data at a few stations in England. Kung used a year of data at a fairly dense network over North America, but his study was still far from global. Since the conversion $-g\mathbf{U}\cdot\nabla z$ varies widely and even changes sign from one point to another, its average over a limited region would not likely represent a global average, even if errors in observation could be eliminated. The less easily computed dissipation $-\mathbf{U}\cdot\mathbf{F}$ can be expected to be positive everywhere, at least when vertically integrated, and an estimate from a limited region might be fairly acceptable. Accordingly the remaining term $d\,(\mathbf{U}\cdot\mathbf{U}/2)/dt$ in the kinetic-energy equation, which would automatically vanish in a long-term global average, was also evaluated over the limited region, and was subtracted from $-g\mathbf{U}\cdot\nabla z$ to yield the estimate of $-\mathbf{U}\cdot\mathbf{F}$.

Holopainen found a value of 10.4 watts m^{-2}; Kung's more extensive data indicated an annual mean of 7.1 watts m^{-2} and rather little seasonal variation, thus raising η to 0.02. In Kung's computations the atmosphere was divided into twenty layers. The greatest dissipation was found in the lowest 100 mb, but, after a relative minimum near the 500-mb level a secondary maximum was reached near the tropopause. Undoubtedly the final estimate of $\overline{D}$ has yet to be made, but it is not unlikely that generally accepted values will have to be revised upward.

Direct evaluation of $\overline{H}$ might appear to offer a third method of determining η, but this method proves to be useless. Heating of all kinds includes heating by friction, and the evaluation cannot be independent of the evaluation of $\overline{D}$. The total non-frictional heating must be zero, and there is nothing further to learn about this total by computing any of its parts.

It is nevertheless possible to estimate $\overline{H}$ from the spatial distribution of Q by taking advantage of the restriction that heating does not in the long run alter the mean entropy. From (103) it follows that for any constant T_1,

$$\overline{\{Q\,(1-T_1/T)\}} = \overline{H}. \tag{105}$$

If T_1 is chosen so that $1/T_1 = \overline{\{1/T\}}$, $\overline{\{1-T_1/T\}}$ vanishes, and $\overline{H}$ is seen to depend only upon the covariance of heating and temperature. It may therefore be estimated moderately well from estimates of Q which are not sufficiently accurate to estimate $\overline{\{Q\}}$. That is, errors in Q will largely be cancelled by errors in $Q\,T_1/T$.

This is in essence the method of evaluation used by Lettau (1954), who thereby obtained a value of 2 watts m^{-2}. Dividing a hemisphere into six zones, each covering fifteen degrees of latitude, Lettau estimated average values of Q and T for each zone, and obtained an estimate of $\overline{H}$. Only horizontal variations of Q and T entered his computations, and he pointed out that his value was probably an underestimate on this account.

It would be equally possible to take advantage of the restriction that heating does not in the long run alter the mean potential temperature. Hence, again in view of (103), for any constant p_1,

$$\overline{\{Q\,(1 - p_1{}^{\kappa}/p^{\kappa})\}} = \overline{H}. \tag{106}$$

To use this equation effectively, a knowledge of the vertical but not the horizontal variations of Q would be needed. Apparently neither equation (105) nor (106) makes full use of the known restrictions upon the distribution of Q. For this purpose the concept of available potential energy has proven advantageous.

Available potential energy

In most of the recent work in atmospheric energetics, TPE has been further resolved into available potential energy (APE) and unavailable potential energy (UPE). With the conventions which we have adopted, the reversible adiabatic processes which convert TPE into KE preserve the potential temperature of each parcel of air, and therefore preserve the statistical distribution of potential temperature. Among those hypothetical states of the atmosphere possessing the same statistical distribution of potential temperature as the existing state, there is one state, commonly known as the reference state, which possesses the least TPE. In the reference state the surfaces of constant pressure and constant potential temperature are horizontal, and the potential temperature never decreases with increasing elevation.

Following Lorenz (1955, 1960), the UPE of any state of the atmosphere is defined as the TPE of the corresponding reference state, while the APE is defined as the excess of TPE over UPE. In his famous paper on the energy of storms, Margules (1903) introduced a quantity similar to APE, which he called available kinetic energy *(die verfügbare kinetische Energie)*, but he did not apply the concept to the general circulation.

Since the reversible adiabatic processes which convert TPE into KE do not alter the reference state, they do not affect the UPE. The conversion of TPE into KE, whose rate is given by C, may therefore equally well be described as a conversion of APE into KE.

In the long run the APE removed by conversion must be replaced by heating, at the same rate at which TPE is replaced by heating. However the individual modes of heating — radiation, conduction, condensation, friction — each affect the UPE, and hence need not affect APE and TPE alike.

We have observed that the net long-term production of TPE by friction alone is equal to the net production of TPE by all modes of heating; the net production by non-frictional heating is zero. In converting KE into TPE, however, friction raises the potential temperature of some portions of the atmosphere, and does not lower the potential temperature of any. It must therefore raise the TPE of the reference state, i.e. the UPE, very likely by about as much as it raises the TPE of the existing state. The APE produced by friction, if any, must therefore be less than the TPE produced, or the KE dissipated, and presumably it is much less.

Herein lies the principal importance of APE. Since as much APE as TPE is produced by heating of all kinds, there must be a net production of APE by heating other than friction. An estimate of the rate at which heating generates APE therefore affords an estimate of η which is independent of those estimates involving the concept of friction.

Since UPE cannot be altered by reversible adiabatic processes, APE is in a sense a measure of the portion of the TPE available for conversion into KE; hence the name. Nevertheless there is no requirement in the definition of the reference state that a reversible adiabatic process leading from the existing to the reference state actually exists, since conservation of potential temperature is only one of the requirements which a dynamically possible reversible adiabatic process must fulfil. In general the reference state cannot be reached, so that *APE* is only an upper bound, and not a least upper bound, for the amount of energy available for conversion into *KE*. The fact that not all of the *APE* is "available" is not however of great consequence, since during any time interval when a major portion of the *APE* could be converted into *KE*, additional *APE* will be produced by heating.

The sum of *APE* and *KE* resembles negative entropy in that friction decreases it and internal reversible adiabatic processes leave it unaltered, while heating is needed to increase it. It is nevertheless a separate concept from entropy, since it involves the field of motion, while entropy depends only upon the thermodynamic state. Strictly speaking, internal radiative transfer can sometimes increase the *APE*: this is not true of negative entropy.

The generation of *APE* by heating, the conversion of *APE* into *KE* by reversible adiabatic processes, and the dissipation of *KE* by friction may be regarded as the three steps in the basic energy cycle of the general circulation. They are indicated schematically in Figure 51.

Before making any direct estimate of the generation of *APE* by heating, it is almost essential to obtain analytic expressions for *APE* and the rate at which it is produced, in order to avoid the errors of omission and oversimplification which are so easily made when verbal arguments are applied to rather intricate phenomena. The definition of *APE* involves the reference state, which is most readily described in a co-ordinate system in which potential temperature θ rather than z or p is used as the vertical co-ordinate. Within any vertical column of unit cross-section, barring superadiabatic lapse rates or appreciable departures from hydrostatic equilibrium, the pressure p corresponding to a given potential temperature θ is simply the weight of the air whose potential temperature exceeds θ. This statement will be true even for values of θ less than the surface value θ_0 provided that the definition of p is extended so that $p(\lambda, \varphi, \theta) = p_0(\lambda, \theta)$ when $\theta < \theta_0$, where p_0 is the surface pressure. It is possible to define a quantity $P(\theta)$ whose value at any point in the atmosphere equals the average value of p on the isentropic surface passing through that point; thus

$$P(\theta) = S \int_S p(\lambda, \varphi, \theta)\, dS, \qquad (107)$$

where $dS = a^2 \cos\varphi \, d\lambda \, d\varphi$ is an element of horizontal area, and the integration extends over the area S of the Earth. The ratio $P(\theta)/P(o)$ is then the probability that a randomly selected mass of air will

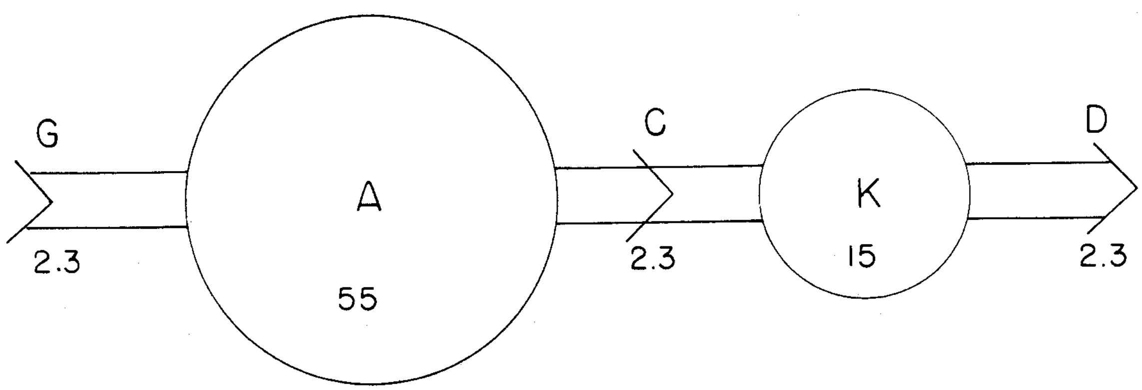


Figure 51. — The basic energy cycle of the atmosphere as estimated by Oort (1964 a). Values of available potential energy and kinetic energy are in units of 10^5 joules m^{-2}. Values of generation, conversion, and dissipation are in watts m^{-2}

have a potential temperature exceeding θ. If now any quantity characterizing the existing state is expressed in terms of $p\,(\lambda, \varphi, \theta)$, the same quantity for the reference state is obtained by replacing p by P.

This assumption neglects the topography. Because elevated land masses fill some of the space which would otherwise be occupied by air, pressure is not a perfect measure of total mass; the mass of air having a pressure between 1000 and 900 mb, for example, is less than the mass having a pressure between 600 and 500 mb.

Neglecting topography, the TPE of a vertical column is given by

$$c_p\, g^{-1} \int_0^{p_0} T dp = (1+\kappa)^{-1}\, p_{00}{}^{-\kappa}\, c_p\, g^{-1} \int_0^{\infty} p^{1+\kappa}\, d\,\theta\,. \tag{108}$$

The term involving $p_0{}^{1+\kappa}\,\theta_0$ which would otherwise appear upon integration by parts has been eliminated by choosing $\theta = 0$ as the lower limit of integration and taking advantage of the extended definition of p.

The UPE may now be obtained by replacing p by P in (108) and integrating horizontally, and the APE may then be obtained by subtraction. At a particular point or even within a particular column APE is not defined, but the APE of the whole atmosphere is seen to be given by

$$A = (1+\kappa)^{-1}\, p_{00}{}^{-\kappa}\, c_p\, g^{-1} \int_S \int_0^{\infty} (p^{1+\kappa} - P^{1+\kappa})\, d\theta\, dS\,. \tag{109}$$

Equation (109) is the so-called exact formula for the APE, although actually it contains the hydrostatic approximation and neglects the presence of topography. It is the appropriate form to use in further theoretical developments. Nevertheless, unless one is thoroughly conditioned to thinking in a θ-co-ordinate system, the features of the atmosphere which are associated with significant amounts of APE may not be apparent. A number of approximate expressions have therefore been developed; the original approximation of Lorenz (1955) is

$$A = \frac{1}{2}\, c_p \left\{ \Gamma_d\, (\Gamma_d - \widetilde{\Gamma}\,)^{-1}\, \widetilde{T}^{-1}\, \widetilde{T''^2} \right\}. \tag{110}$$

Here $\Gamma = -\partial T/\partial z$ is the vertical lapse rate of temperature, and $\Gamma_d = g/c_p$ is the dry-adiabatic lapse rate (about 10° per kilometre). The tilde ($\sim$) denotes an average over an entire isobaric (or approximately horizontal) surface, and the double prime ($''$) denotes a departure from this average. Lorenz obtained the approximation by observing first that since $p > 0$, $\kappa > 0$, and P is an average of p, the integral of $p^{1+\kappa} - P^{1+\kappa}$ is positive definite for each isentropic surface, and may be approximated in terms of the variance of p on the isentropic surface. Second, provided that the isentropic surfaces are not too greatly inclined to the horizontal, the variance of p on an isentropic surface may be approximated in terms of the variance of θ, and hence of T, on an isobaric surface.

Van Mieghem (1956) obtained a somewhat similar approximation by assuming that the reference state could evolve from the existing state by a dynamically possible adiabatic process, and then using a variational procedure to compute the gain of KE during the envisioned process. The general agreement between the expressions might not have been expected in view of the usual absence of a process leading to the reference state. However, APE depends upon the field of mass alone, while the existence of the envisioned process depends also upon the field of motion. Apparently, corresponding to any existing field of mass, at least within reasonable limits, there exists a hypothetical field of motion, in general differing from the existing field of motion, such that, if the state of the atmosphere consisted of the existing field of mass and the hypothetical field of motion, the reference state would subsequently occur. The expressions should therefore agree.

From expression (110) it follows that the APE may be approximated by a weighted average of the *horizontal* variance of temperature, the weighting function being inversely proportional to the horizontally averaged static stability, as measured by $\Gamma_d - \widetilde{\Gamma}$. The approximation may be shown to be most acceptable when $\Gamma_d - \widetilde{\Gamma}$ is large, and it is worthless if $\widetilde{\Gamma}$ is near Γ_d, since the APE does not become infinite.

This approximation is consistent with the approximate rule that KE is produced when cold air sinks and warm air rises at the same level. In order that such a process may occur at all, there must first of all be different temperatures at the same level. If the stratification is stable, the temperature at a fixed elevation will rise in the sinking air and fall in the rising air, and the process will thereby reduce the horizontal temperature contrast and finally eliminate it. Moreover, the less stable the stratification, the farther the cold air must sink and the warm air must rise in order to eliminate the temperature contrast. Thus more KE is attainable, and so more APE is present, when the horizontal temperature contrast is greater and when stratification is less stable.

According to (110), there should be two principal methods by which heating can produce APE; first, by heating the warmer regions and cooling the cooler regions at the same elevation, thereby increasing the horizontal temperature variance, and, second, by heating the lower levels and cooling the upper levels, thereby decreasing the static stability. The former process is essentially the one explicitly considered by Lettau in his estimate of the efficiency.

From the exact expression (109) for APE, it follows that

$$\partial A / \partial t = G - C, \tag{111}$$

where

$$G = \{ Q \, (1 - p^{-\kappa} P^{\kappa}) \}. \tag{112}$$

The quantity $N = 1 - p^{-\kappa} P^{\kappa}$ appearing in (112) may be regarded as an efficiency factor, which represents the effectiveness of heating at any point in producing APE. Where N is negative, cooling will produce APE.

Figure 52 presents a hypothetical distribution of potential temperature, which is based upon the average temperature field shown in Figure 10. Possible variations of surface pressure have been neglected. The pressure which a given potential-temperature surface would assume in the reference state is proportional to the area above this surface; these pressures are indicated by the numbers in parentheses, the pressure at the Earth's surface being taken as 1000 mb. With this numbering, Figure 52 becomes a chart of the distribution of P. From Figure 52 the distribution of the efficiency factor N has been evaluated; it is shown in Figure 53.

If this distribution of N is at all typical, the APE generated by friction could not exceed 8 per cent of the kinetic energy dissipated, even if friction were confined to low levels in the tropics. If friction were uniformly distributed throughout low levels, the amount could not exceed 3 per cent. If it is true that there is considerable dissipation near the tropopause, where N is negative, friction may generate no APE at all. The assumption that a direct estimate of $\overline{G}$ is independent of an estimate based upon friction therefore seems to be justified.

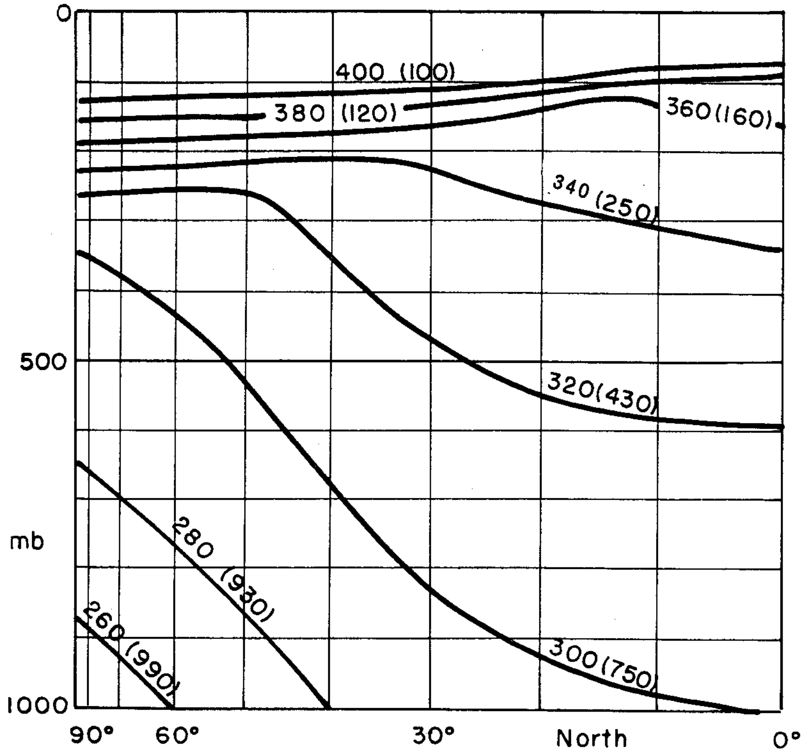


Figure 52. — A hypothetical distribution of potential temperature θ, and corresponding distribution of $P(\theta)$. Values of θ are in degrees K, and values $P(\theta)$ (in parentheses) are in millibars

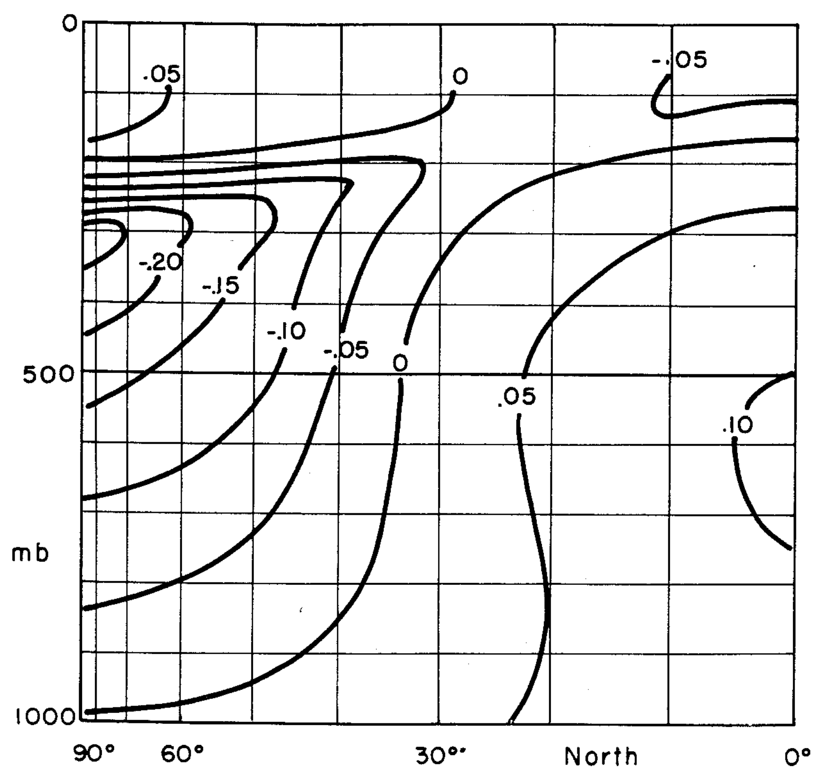


Figure 53. — The distribution of efficiency factor N corresponding to distribution of θ in Figure 52

From the approximate expression (110) for A, Lorenz derived the approximation

$$G = \left\{ \Gamma_\mathrm{d} \, (\Gamma_\mathrm{d} \widetilde{\Gamma}) \, \widetilde{T}^{-1} \, \widetilde{Q}'' \, \widetilde{T}'' \right\}. \tag{113}$$

This formula has formed the basis for many of the subsequent observational studies. Only the influence of heating upon the factor $\widetilde{T}''^2$ in (110) has been included. A closer approximation would include the influence of heating on $\widetilde{\Gamma}$ and $\widetilde{T}$, but it would still not agree with the exact expression (112).

On the basis of this expression Lorenz estimated a value of 4 watts m^{-2} for $\overline{G}$. This estimate was based upon assumed average values of Q and T as functions of latitude only. It is of interest to compare the result with Lettau's lower estimate. There was little difference in the data used, and the formulae are nearly the same except for the lapse-rate factor $\Gamma_\mathrm{d} \, (\Gamma_\mathrm{d} - \widetilde{\Gamma})^{-1}$ and the restriction to the horizontal covariance in (113).

With a "normal" lapse rate of 2/3 of the dry-adiabatic, the lapse-rate factor acquires a value of 3. This more than accounts for the difference in the estimates, and seems to represent the degree of the underestimation whose existence Lettau pointed out. The heating must decrease with elevation, and this decrease must contribute to a covariance of Q and T in formula (105) if T also decreases with elevation. Formula (113) effectively takes into account the necessary vertical decrease of Q by means of the lapse-rate factor.

The fundamental theoretical question now arises as to why η should be as low as one or two per cent, or, for that matter, why it should not be even lower. In attempting to answer this question, Lorenz (1960) sought the maximum possible value of η. Since the generation of APE depends essentially upon the horizontal covariance of heating and temperature, there can be no generation with no temperature contrast. If on the other hand the contrast is so great that radiative equilibrium prevails, there is no net heating and again no generation. The maximum generation therefore accompanies a somewhat weaker temperature contrast. The thermodynamic efficiency is then not particularly great, since even the cold source is not excessively cold, while η falls considerably short of the thermodynamic efficiency, since there is considerable outgoing radiation at the heat source.

On the basis of a crude model, Lorenz found a maximum η of not much greater than two per cent. This led him to speculate that the atmosphere might be constrained to operate at nearly maximum efficiency; specifically, when several modes of behaviour satisfy the dynamic equations, the less efficient modes may be unstable and give way to the more efficient modes.

Further considerations suggest that the maximum efficiency may really be considerably higher; for example, Lorenz's model applies to a dry atmosphere, where the outgoing radiation is greatest at the warmest latitudes, thus acting to destroy APE. In the real atmosphere outgoing radiation is more nearly independent of latitude. Nevertheless, the notion that the general circulation should act so as to maximize or minimize some basic quantity is attractive, and perhaps not unreasonable. Dutton and Johnson (1967) have recently attempted to apply principles similar to the principle of least action to the general circulation. The introduction of such a principle might seem to be an overspecification, but, since the governing equations apparently possess a nearly infinite variety of sets of long-term statistics, some such principle may be just what is required to single out the set which actually prevails.

Zonal and eddy energy

The knowledge that APE is continually converted into KE at a certain average rate does not by itself reveal the types of weather patterns which are primarily responsible for the conversion. The dominant process could conceivably be a general sinking of cold air in higher latitudes and a rising of warm air in lower latitudes, as in Hadley's circulation. It could equally well be a sinking in the colder portions of cyclones and anticyclones and a rising in the warmer portions at the same latitudes. Most of the recent studies in atmospheric energetics have been concerned with the further resolution of KE and APE, and of the accompanying generation, conversion, and dissipation processes, into the portions associated with separate modes of motion.

Most frequently KE and APE are resolved into the amounts associated with the zonally averaged fields of motion and mass and the amounts associated with the eddies. The KE may be resolved into zonal kinetic energy (ZKE), the amount of KE which would exist if the existing zonally averaged motion but no eddy motion were present, and eddy kinetic energy (EKE), the excess of KE over ZKE. Because KE is a quadratic function of velocity, EKE is also the KE which would remain if the existing eddy motion but no zonal motion were present. If the velocity field $\mathbf{U}$ is resolved into the zonally averaged motion $[\mathbf{U}]$ and the eddy motion $\mathbf{U}^\star$, the total amounts of ZKE and EKE become

$$\mathrm{K_z} = \frac{1}{2}\left\{[\mathbf{U}]\cdot[\mathbf{U}]\right\}, \tag{114}$$

$$\mathrm{K_E} = \frac{1}{2}\left\{\mathbf{U}^\star\cdot\mathbf{U}^\star\right\}. \tag{115}$$

It is to be observed that the term "zonal kinetic energy" does not refer to the kinetic energy of the zonal motion (which would be $u^2/2$) nor the zonally averaged kinetic energy (which would be $[\mathbf{U}\cdot\mathbf{U}]/2$).

Likewise, the APE may be resolved into zonal available potential energy ($ZAPE$), the amount of APE which would exist if the field of mass were replaced by its zonal average, and eddy available potential energy ($EAPE$), the excess of APE over $ZAPE$. It is not at all certain how the zonal average of the mass field is best defined. In order to use the exact formula (109), it would be best to define $[p]$ as the average pressure along a latitude circle on an *isentropic* surface, and to let $p^\star = p - [p]$. In this case the total amounts of $ZAPE$ and $EAPE$ become

$$A_z = (1+\kappa)^{-1}\, p_{00}{}^{-\kappa}\, c_p\, g^{-1} \int_S \int_0^\infty ([p]^{1+\kappa} - P^{1+\kappa})\, d\theta\, dS, \tag{116}$$

$$A_E = (1+\kappa)^{-1}\, p_{00}{}^{-\kappa}\, c_p\, g^{-1} \int_S \int_0^\infty (p^{1+\kappa} - [p]^{1+\kappa})\, d\theta\, dS. \tag{117}$$

Because the integrand in (109) is approximately quadratic in p, the $ZAPE$ is approximately equal to the amount of APE which would exist if p were replaced everywhere by $P + p^\star$.

In most computations the approximate expression (110) for APE has been used in place of (109). The corresponding approximations for $ZAPE$ and $EAPE$ are

$$A_z = \frac{1}{2}\, c_p \left\{\Gamma_d\, (\Gamma_d - \widetilde{\Gamma})^{-1}\, \widetilde{T}^{-1}\, [T]''^2\right\}, \tag{118}$$

$$A_E = \frac{1}{2}\, c_p \left\{\Gamma_d\, (\Gamma_d - \widetilde{\Gamma})^{-1}\, \widetilde{T}^{-1}\, T^{\star 2}\right\}, \tag{119}$$

where, according to the usual convention, $[T]$ is the average temperature along a latitude circle on an isobaric surface, and $T^\star = T - [T]$.

The recognition of additional forms of energy makes it possible to investigate a more detailed energy cycle. We must first observe that when a complicated physical process affects several forms of energy, it is not always possible to define the rate at which one form is converted into another by this process. If however the process can be resolved into simpler processes, each of which affects only two forms of energy, the conversion rate by each process can be defined. The resolution of the original process into simpler processes may at times seem somewhat arbitrary, but ordinarily it will be no more arbitrary than the resolution of energy which made it necessary.

Accordingly, the process which generates APE may be resolved into a heating at warmer latitudes and a cooling at colder latitudes, which generates $ZAPE$, and a heating of warmer regions and cooling of colder regions at the same latitude, which generates $EAPE$. Likewise, the conversion process may be resolved into a sinking in colder latitudes and rising in warmer latitudes, converting $ZAPE$ into ZKE, and a sinking of colder air and rising of warmer air at the same latitude, which converts $EAPE$ into EKE. Finally, the dissipation may be resolved into dissipation of ZKE by the zonally averaged friction and dissipation of EKE by the departure of friction from its zonal average.

With sufficient resolution of the physical processes, there is no process which converts $ZAPE$ into EKE, or $EAPE$ into ZKE (see Lorenz, 1955). However, there remain the processes which can convert $ZAPE$ into $EAPE$ without affecting KE, and ZKE into EKE without affecting APE.

The latter process consists of a horizontal or vertical transport of absolute angular momentum by the eddies toward latitude circles of lower angular velocity, in the direction which would be expected if the large-scale eddies behaved in the manner of classical turbulence. Likewise, the former process consists of a horizontal or vertical transport of sensible heat by the eddies toward latitude circles of lower temperature (actually lower $[T]''$), again in the manner suggested by classical turbulence theory.

Presumably both ZKE and EKE are dissipated by friction and therefore require sources. There must be a net conversion of $ZAPE$ into ZKE or $EAPE$ into EKE, but it is not necessary that both conversions proceed in this direction, since one form of KE could serve as the needed source for the other. Likewise there must be a net generation of $ZAPE$ or $EAPE$ by heating, but it is not necessary that both forms be generated, since one form of APE could serve as the source for the other, or KE could serve as the source for one form of APE. Thus, unlike the basic energy cycle, the directions in which the various steps of the more detailed energy cycle proceed cannot be deduced in any simple manner from existing theory, and have been ascertained only from observations.

The numerous observational studies of the energy cycle are in fair qualitative agreement. First, the heating at low latitudes and the cooling at high latitudes are very effective in generating $ZAPE$. Whether $EAPE$ is produced or destroyed by heating is less certain. Next, the eddy transport of heat is mainly toward colder latitudes, as classical turbulence theory would suggest, so that $ZAPE$ is converted into $EAPE$. On the other hand, the eddy transport of angular momentum is on the average toward latitudes of higher angular velocity, so that EKE is converted into ZKE. This result is just the opposite of what would be predicted by classical turbulence theory. Well before the introduction of the concept of APE, it was deduced by Kuo (1951) on the basis of observed winds over North America.

It follows that $EAPE$ must be converted into EKE, since there is no other source for EKE. Whether $ZAPE$ is converted into ZKE or vice versa is less certain. The low-latitude Hadley cells must act to convert $ZAPE$ into ZKE, while the middle latitude Ferrel cells have the opposite effect. It seems likely that the effect of the weak Ferrel cells is greater than that of the stronger Hadley cells, because the former occur in regions of stronger horizontal temperature contrast.

The conversion of *EKE* into *ZKE* implies that in a sense the large-scale eddying motion is an unmixing process. Any attempt to deduce the circulation by treating the over-all mechanical effect of the eddies as large-scale turbulent friction would fail unless it assumed a negative coefficient of turbulent viscosity. The phenomenon of negative viscosity has been one of the most unexpected and perhaps one of the most important recent meteorological discoveries; some evidence for it has subsequently been found in other physical systems, ranging in size from small laboratory models to Jupiter and the Sun.

Figure 54 illustrates the detailed energy cycle. The directions of the arrows indicate the directions in which the various processes proceed, according to the consensus of investigators.

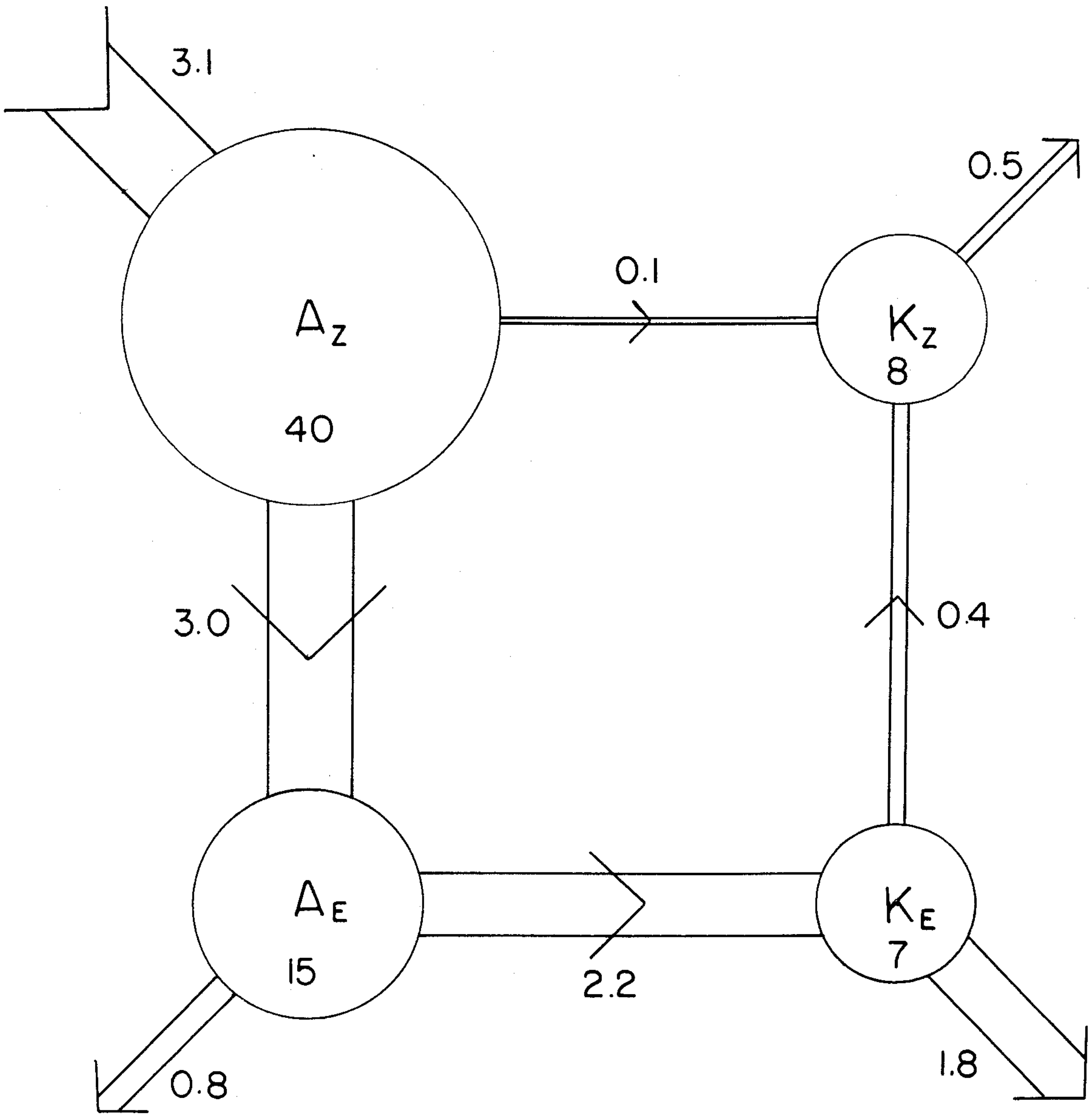


Figure 54. — The energy cycle of the atmosphere as estimated by Oort (1964a). Values of energy are in units of 10^5 joules m^{-2}, and values of generation, conversion, and dissipation are in watts m^{-2}. The estimated value of conversion from A_Z to K_Z is smaller than the probable error of the estimate

To compute the numerical values we require suitable mathematical expressions. Most of the estimates so far performed have been based upon the approximate expressions given by Lorenz. We first let

$$\partial A_z/\partial t = G_z - C_z - C_A\,, \tag{120}$$

$$\partial A_E/\partial t = G_E - C_E + C_A\,, \tag{121}$$

$$\partial K_z/\partial t = C_z - C_K - D_z\,, \tag{122}$$

$$\partial K_E/\partial t = C_E + C_K - D_E\,. \tag{123}$$

Equations (113), (102) and (101) express G, C, and D as covariances; the corresponding zonal and eddy generations, conversions, and dissipations are simply covariances of zonally-averaged quantities or eddy quantities. Thus

$$G_z = \left\{\Gamma_d\,(\Gamma_d - \widetilde{\Gamma})^{-1}\,\widetilde{T}^{-1}\,[Q]''\,[T]''\right\}, \tag{124}$$

$$G_E = \left\{\Gamma_d\,(\Gamma_d - \widetilde{\Gamma})^{-1}\,\widetilde{T}^{-1}\,Q^\star\,T^\star\right\}, \tag{125}$$

$$C_z = -\left\{[\omega]\,[\alpha]\right\}, \tag{126}$$

$$C_E = -\left\{\omega^\star\,\alpha^\star\right\}, \tag{127}$$

$$D_z = -\left\{[\mathbf{U}]\cdot[\mathbf{F}]\right\}, \tag{128}$$

$$D_E = -\left\{\mathbf{U}^\star\cdot\mathbf{F}^\star\right\}. \tag{129}$$

The two new processes are the conversion from ZKE to EKE,

$$C_K = -\left\{a\cos\varphi\left([u^\star\,v^\star]\,\frac{1}{a}\,\frac{\partial}{\partial\varphi} + [u^\star\,\omega^\star]\,\frac{\partial}{\partial p}\right)\left[\frac{u}{a\cos\varphi}\right]\right\}$$
$$-\left\{a\cos\varphi\left([v^{\star 2}]\,\frac{1}{a}\,\frac{\partial}{\partial\varphi} + [v^\star\,\omega^\star]\,\frac{\partial}{\partial p} + \frac{\tan\varphi}{a}\,[u^{\star 2} + v^{\star 2}]\right)\left[\frac{v}{a\cos\varphi}\right]\right\}, \tag{130}$$

and the conversion from $ZAPE$ to $EAPE$

$$C_A = -\,c_p\left\{\widetilde{T}^{-1}\widetilde{\theta}\left([T^\star\,v^\star]''\,\frac{1}{a}\,\frac{\partial}{\partial\varphi} + [T^\star\,\omega^\star]''\,\frac{\partial}{\partial p}\right)\left(\Gamma_d\,(\Gamma_d - \widetilde{\Gamma})^{-1}\,\widetilde{\theta}^{-1}\,[T]''\right)\right\}. \tag{131}$$

Since $[v]$ tends to be small, the second line in (130) may ordinarily be disregarded. In deriving (131), we have considered the influence of the eddies only on the factor $\widetilde{T''^2}$ in the approximate expression (110) for A.

The numerical values in Figure 54 are those given by Oort (1964a). They are based upon a comprehensive assessment of all previously published numerical values.

Undoubtedly Oort's figures will not be the final word. Dutton and Johnson (1967), for example, have noted the dangers in using the approximate formula (113) to evaluate G. In particular, they note that surface heating in higher latitudes should produce APE, since N is slightly positive, whereas according to (113) surface heating will destroy APE, since T'' is negative. Accordingly they have estimated G from the exact formula (112), and their value is 5.6 watts m^{-2}. Since their estimates of N and Q are based upon seasonal averages, they assume that their result represents the contribution of G_z, and, on the basis of Kung's estimate of G, which exceeds Oort's by a factor of 3, they assume that G_E is positive.

Figure 53 also indicates the long-term average amounts of the various forms of energy. The numerical values are again those of Oort. The total KE, 15×10^5 joules m^{-2}, is equal to the amount which would exist if the wind had a speed of 17 m sec^{-1} everywhere. Even on such a basic quantity there is a lack of agreement; some studies suggest that it is nearly twice as large, while Holopainen's computations (1966) based upon Crutcher's charts (1959) indicate that it is 20 per cent smaller.

A number of investigators have recently subjected APE and KE to still further resolutions, in order to gain further evidence as to how the circulation operates. The number of possible resolutions seems to be almost limitless. Oort (1964a) has investigated the resolution of KE and APE into the amounts associated with the time-and-longitude averaged fields of motion and temperature and the amounts associated with the remaining fields, noting that this resolution facilitates the computation procedure. Qualitatively the energy cycle is much the same, but quantitatively it is different. Wiin-Nielsen (1962) has resolved KE into the amounts associated with the vertically averaged flow and the departure from the vertical average, and has found that APE is converted primarily into the latter form of KE.

A number of studies have been concerned with the resolutions of $EAPE$ and EKE and the processes affecting them into the amounts associated with each wave number. The wave numbers are defined by means of a Fourier analysis with respect to longitude. The appropriate equations have been presented by Saltzman (1957).

The most complete study of this sort in terms of length of record has been performed by Saltzman and Teweles (1964), who have used nine years of daily 500-mb data to resolve the conversion C_K into the amounts accomplished separately by wave numbers 1 to 15 inclusive. They find that all wave numbers contribute negatively to C_K, with a peak contribution at wave numbers 2 and 3 and another peak at wave numbers 6 and 7. The former peak is the dominant one in winter and is nearly absent in summer, while the latter peak appears at every season of the year.

Wiin-Nielsen, Brown and Drake (1964) performed similar computations at five levels, using a total of eight months of data, and found that the great majority of conversions by individual wave numbers for individual months were positive, although there were some notable exceptions. Their results indicated that the conversion processes might not be very reliably determined from data at only one level. In addition they performed a resolution of the conversion C_A, and found that all wave numbers contributed positively.

Measurements of the conversion C_E are hindered by our lack of knowledge of the true vertical-motion field, but Saltzman and Fleisher (1961) used six months of daily values of ω for the 850–500 mb layer, as evaluated by the ω-equation, and obtained a resolution of the conversion process. Again all wave numbers contributed positively to a conversion of $EAPE$ into EKE, with the peak contribution occurring at wave number 6.

The study of the energetics of a particular portion of the atmosphere is complicated by the possibility of a flow of mass as well as energy across the boundary. Much speculation has arisen concerning the maintenance of the circulation of the stratosphere following the discovery by White (1954) of the counter-gradient flux of sensible heat, which acts in the proper direction to maintain the equatorward temperature decrease. Oort (1964b) has studied the energy cycle of the layer of the stratosphere between 100 and 30 mb, using one year of data, and has found that within this layer the processes contribute to a net conversion of EKE into ZKE, EKE into $EAPE$, and $EAPE$ into $ZAPE$. There is therefore no source within the region for EKE, and Oort concludes that the stratospheric eddies must be mechanically forced by the air above or below, and presumably by the tropospheric eddies. Radiation also serves as a sink for $ZAPE$. The lower stratospheric energy cycle therefore seems to differ considerably from that of the atmosphere as a whole, and the stratosphere appears to act as a heat engine in reverse, or a thermodynamic refrigerator, converting KE into APE.

In considering the basic energy cycle we faced the problem of explaining its intensity. With the recognition of the more detailed energy cycle there is the further problem of accounting for the directions in which the various steps proceed. Here we may again invoke the hypothesis that the general circulation is constrained to operate at nearly maximum efficiency.

The energy cycle can operate at its maximum rate only if the cross-latitude temperature contrast is considerably less than the contrast which would occur under thermal equilibrium — perhaps half as great. As we noted in the previous chapter, in the absence of eddies the meridional circulation would have to be extremely weak in order not to lead to upper-level westerly winds in excess of those permitted by the thermal wind relation. It could therefore transport only enough energy from warm to cold latitudes to reduce the temperature contrast slightly below its thermal-equilibrium value. If the energy cycle is to proceed at nearly maximum efficiency, eddies must therefore occur.

We do not offer this suggestion as an alternative to the hypothesis that the presence of eddies is attributable to instability. It is more properly an alternative statement of the same hypothesis. Instability is still the mechanism through which the eddies would develop and allow the cycle to proceed more efficiently.

It would seem reasonable to pass to the conclusion that the eddies, whose presence is demanded because the unaided meridional circulation cannot accomplish the needed energy transport, will now form the mechanism for producing the transport, and will therefore convert $EAPE$ into $ZAPE$. Such a conclusion does not necessarily follow. As we also noted in the previous chapter, as long as eddies are present the meridional circulation need not be weak. Conceivably the meridional circulation could accomplish the needed heat transport, and the function of the eddies could be to prevent the angular momentum aloft from reaching excessive values. The conversions would then be from $ZAPE$ to ZKE to EKE to $EAPE$, just the opposite of those shown in Figure 54. It is only through observations that we know that this is not the case. We are led to conclude that the problem of explaining the direction of the energy cycle is more complicated than it might at first seem to be.

CHAPTER VI

LABORATORY MODELS OF THE ATMOSPHERE

The theoretical study of the Earth's atmosphere is often facilitated by introducing various idealizations. Some of these consist of suppressing certain supposedly irrelevant details, so that the important processes may be more readily examined. In other instances certain features which may not be irrelevant at all are nevertheless omitted to render the theoretical treatment less awkward. In the present chapter we shall consider certain real fluid systems other than the Earth's atmosphere which share certain properties with it, and which may serve in one way or another as still further idealizations.

Perhaps the most obvious systems of this sort are the atmospheres of other planets. Unfortunately, definitive measurements of the motions of other planetary atmospheres are hard to obtain. Our ideas have come mainly from tracking identifiable features, which may or may not actually move with the large-scale flow. Clouds have been followed on Mars, and on Jupiter there have been abundant observations of the motions of spots, which themselves appear to be large atmospheric disturbances. The available observations are sufficient to reveal that the circulations of other atmospheres in our solar system do not resemble that of our own atmosphere very closely.

Since the planets differ so greatly in their solar distances and angular velocities and in the compositions of their atmospheres, it is not surprising that the resulting circulations also differ considerably. Nevertheless it is of interest to observe that many of the earlier theories of the general circulation of the Earth's atmosphere made no use of any numerical values. Taken at face value, these theories would then have predicted similar circulations in all of the planetary atmospheres, provided simply that the planets were rotating, with their equatorial planes not too greatly inclined to their orbital planes. Whatever else we may learn from a brief consideration of the atmospheres of other planets, we are forcefully reminded of the importance of quantitative considerations.

There is much to be learned about our own atmosphere from studying our oceans. The analogy between the Gulf Stream and the jet stream is especially revealing. Nevertheless the oceanic circulation as a whole is not a particularly good model of the atmospheric circulation.

In the absence of other easily observed natural systems resembling the atmosphere, further analogues must be confined to man-made systems. Chief among these are the laboratory models which have been specifically designed to simulate the atmosphere. These models consist of differentially heated rotating cylinders containing a fluid. The cylinder represents a hemisphere of the Earth, the fluid represents the atmosphere, and the motion of the fluid is intended to represent the atmospheric circulation.

The primary importance of the models stems from the fact that the experiments, being performed in the laboratory, are subject to a certain degree of control. The shape and size of the container, the nature and amount of the fluid, the distribution and intensity of the external heat sources, and the rate of rotation can all be chosen in advance. Such internal parameters as the average speed of the fluid relative to the container can often be set to desired values by adjusting the controllable parameters on a trial-and-error basis.

The earliest experiments appear to have been those of Vettin (1857), whose apparatus was a rotating cylinder about 30 centimetres in diameter and 5 centimetres deep, containing air. In one experiment ice was placed at the centre. The resulting motion of the air was in its essential features like the atmospheric circulation envisioned by Hadley. Apparently Vettin's work had no immediate successors.

If the arguments advanced by the early writers on the general circulation should be applicable to the atmospheres of other planets, they should be equally applicable to suitably designed laboratory experiments. This fact was recognized by Thomson, who in his Bakerian Lecture (1892) proposed some experiments with an apparatus very much like Vettin's, but containing water. He also noted the desirability of varying some of the controllable parameters. Any plans he might have had to carry out his suggestions were halted by his death soon afterward.

A few more experiments were performed in the early twentieth century. For a more detailed account the reader is referred to Fultz (1951), or to the monograph of Fultz *et al.* (1959).

The dishpan experiments

The modern era of laboratory simulation began shortly after World War II with the experiments performed at the Hydrodynamics Laboratory of the University of Chicago (see Fultz *et al.* 1959). The experiments most nearly duplicating the atmosphere are the dishpan experiments. The central piece of apparatus is a cylindrical container — often an ordinary aluminium dishpan — containing water. It is mounted on a rotating turntable. A heat source is provided near the rim, and in the more refined experiments there is a cold source near the centre. The apparatus may easily be duplicated if only qualitative results are desired, but extreme care must be used if useful numerical data are to be obtained. Among the requirements in the latter instance are an almost exactly constant rate of rotation. The motion at the free surface can be made visible for an indefinite length of time by particles of aluminium powder or some other tracer, while internal motions can be temporarily followed by injecting a dye.

Velocities at the free surface are readily measured photographically. If the camera rotates with the apparatus, and a time exposure is made, the tracer particles will appear as streaks. The lengths of the streaks will indicate the speed of the flow. Temperature measurements may be obtained by means of thermocouples.

The various details of the experiments are conveniently designated by their atmospheric counterparts; the centre of the dishpan becomes the North Pole, the rim becomes the Equator, the direction toward the centre becomes north, and the direction toward which the dishpan rotates becomes east. A period of revolution becomes a day. A typical day is actually a fraction of a minute, a typical radius is about 15 cm, and the water is typically about 2 cm deep.

The external parameters which are readily altered without changing the apparatus or adding or removing water are the rate of rotation and the contrast between the heat and cold sources. Over a considerable range of these parameters, the flow which develops in the dishpan is perfectly symmetric about the axis of rotation, within the limits of observational error. In other experiments differing only in the values of the external parameters, a set of large-scale wave-like disturbances develops, and the pattern at the free surface bears considerable resemblance to an upper-level weather map. There are thus two qualitatively different régimes of flow, a zonally symmetric régime and a zonally asymmetric régime, which Fultz has called the Hadley régime and the Rossby régime. Typical circulation patterns occurring in the Hadley and Rossby régimes are shown in Figures 55 and 56, which are photographs of the free surface.

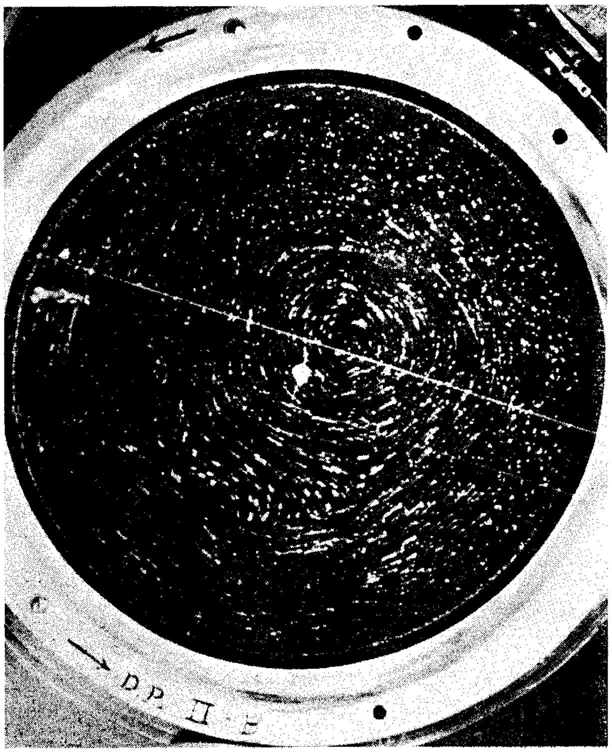


Figure 55. — A photograph of the upper surface of fluid in a rotating dishpan, showing a nearly symmetric circulation pattern. The photograph is a time exposure, so that particles of tracer on the upper surface appear as streaks indicating direction and speed of flow (from Starr and Long, 1953)

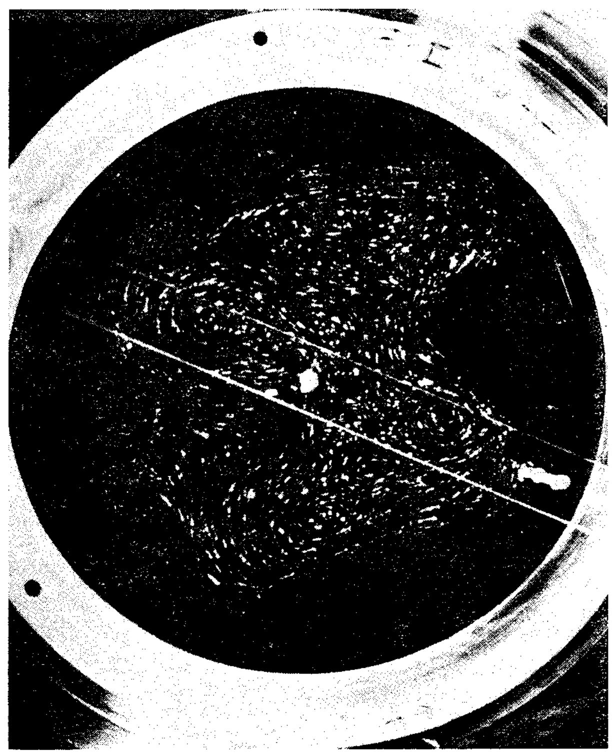


Figure 56. — A photograph of the upper surface of fluid in a rotating dishpan, showing an unsymmetric (Rossby-régime) circulation pattern. The rate of rotation of the dishpan is higher than in Figure 55 (from Starr and Long, 1953)

In the earlier experiments it was found that with sufficiently slow rotation only Hadley flow would develop. With faster rotation, Hadley flow would also develop under a strong heating contrast, but with a weak heating contrast, Rossby flow occurred. Later it was found that Hadley flow could also be produced with extremely weak heating.

We have already noted that the arguments presented by Hadley (1735) and some of the nineteenth century meteorologists ought to be equally applicable to the dishpan. It is therefore of considerable interest to note that the dishpan may contain a flow bearing considerable resemblance to the one described by Hadley. Perhaps more than any other discovery, the dishpan has exonerated Hadley's ideas from some of the criticisms raised against them. Hadley's paper is a nearly correct account of what sometimes occurs in the dishpan, although it can never be a demonstration that this type of flow must occur in preference to some other one.

From the point of view of the general circulation, perhaps the most important question to be answered is whether the resemblance of the Rossby-régime flow in the dishpan to the circulation of the atmosphere is more than superficial. Certainly under proper conditions the free surface looks like a weather map at the tropopause level. There is generally a large irregular circumpolar vortex, surrounded by a narrow meandering jet stream. The wave-like disturbances in the westerly flow progress about the centre in much the same manner as the upper-level waves in the atmosphere.

Further similarities appear in the flow below the free surface. By injecting a dye, Fultz (1952) found evidence for the existence of fronts and migratory cyclones, which moreover appeared to be properly located with respect to the upper-level waves. Using a much larger apparatus, Faller (1956) was able to observe extensive frontal surfaces, which possessed families of wave disturbances whose structure and evolution closely resembled the classical Norwegian cyclone model.

Evidence that the dishpan resembles the atmosphere in other respects was provided by Starr and Long (1953). Using velocity measurements at the free surface obtained from a sequence of 108 photographs, they evaluated the northward transport of angular momentum at six different latitudes. They found an average northward transport at all latitudes, with a peak value in low latitudes occurring near the latitude of maximum westerly "wind". Furthermore nearly all of the transport was accomplished by the eddies. It thus appears that eddies in the dishpan and in the atmosphere play similar roles in the angular-momentum balance.

The production of kinetic energy in the dishpan must be accomplished by the pressure forces, yet the differences in pressure are so minute that direct measurements would be difficult to obtain. The hydrostatic and geostrophic relations therefore cannot be directly confirmed. Temperature measurements are easily made, however, and one can confirm the geostrophic relation by confirming the thermal wind relation and assuming that hydrostatic equilibrium must prevail in any case. In the dishpan the thermal-wind relation assumes the form

$$\partial \mathbf{U}/\partial z = \frac{1}{2}\, \varepsilon g\, \Omega^{-1}\, \mathbf{k} \times \nabla T\,, \tag{132}$$

where $\varepsilon = d\,(1n\, a)/dt$ is the coefficient of thermal expansion. Measurements in a dishpan with a radius of 15 centimetres, where the rim-to-centre contrast may be about 10°, show that nearly half of this contrast may occur across a jet stream about 1 cm wide; furthermore, under the assumption that $\mathbf{U} = 0$ at the bottom, the thermal-wind relation is found to be very closely satisfied.

The circulation in the dishpan therefore resembles the atmospheric circulation in many important respects. Some features of the atmosphere of course are not reproduced. Notable among these is the tropical circulation; the dishpan has a constant Coriolis parameter, and the tropical and temperate-latitude circulations cannot be modelled simultaneously. The stratosphere, which depends upon radiative heating in the ozone layer, is not present in the dishpan, and phenomena depending upon changes of phase of water are not reproduced.

According to the usual conditions for dynamic modelling, the transition between the Hadley and Rossby regimes should depend upon the values of certain dimensionless ratios involving the rotation and the heating contrast, rather than directly upon dimensional values of these quantities. The governing equations in dimensionless form contain a number of dimensionless parameters; among these, the values of the Taylor number T_a and the thermal Rossby number R_{OT} seem to have the greatest influence on the results of an experiment.

The Taylor number is defined as

$$T_a = 4\,\Omega^2\,h^4\,\nu^{-2}\,, \tag{133}$$

where h is the depth of the fluid and ν is the kinematic viscosity. The square root of T_a is a measure of the ratio of Coriolis forces to viscous forces, and has been called the "rotation Reynolds number". Because of the great depth and low molecular viscosity of the atmosphere, the atmospheric Taylor number cannot be conveniently duplicated in the dishpan, and it had been argued that for this reason the dishpan could not simulate the atmosphere. When the Chicago experiments were undertaken despite this warning and proved successful, an explanation was needed. The most plausible explanation seems to be that the effective viscosity in the atmosphere is the turbulent viscosity, which is generally at least 10^5 times as great as molecular viscosity. Defined in terms of turbulent viscosity, a typical atmospheric Taylor number is about 10^7, which is readily obtained in the dishpan.

The thermal Rossby number may be defined as

$$R_{OT} = \frac{1}{2}\,g\varepsilon\,h\,\Omega^{-2}\,a^{-2}\,\Delta T\,, \tag{134}$$

where ΔT is the rim-to-centre difference of the vertically averaged temperature. It is evident from (134) that R_{OT} depends upon both heating and rotation; the dimensionless parameter which depends upon temperature contrast alone is the product $T_a\,R_{OT}$.

In a more general rotating fluid system the Rossby number R_0 is often defined as the ratio of a suitably chosen relative velocity to a suitably chosen absolute velocity. Such a ratio was first used in connection with the atmosphere by Kibel (1940). In the dishpan it is convenient to let

$$R_o = u_1\,\Omega^{-1}\,a^{-1}\,, \tag{135}$$

where u_1 is the average value of $[u]$ with respect to latitude, $[u]$ being the average of u with respect to longitude, at the free surface. The thermal Rossby number R_{OT} has been defined in such a way that it will reduce to the Rossby number R_0 if the thermal wind relation is valid, and if the flow near the bottom is negligibly small; this may be seen by comparing (132), (134), and (135).

It should be noted that R_{OT} is strictly speaking an internal parameter; it is used because of the obvious difficulties in measuring a meaningful external temperature contrast when the heating element may be a Bunsen burner or an electrical coil. Thus a pre-chosen R_{OT} cannot be directly entered into an experiment, although it can generally be established by a trial-and-error procedure. Nevertheless it is often convenient to think of an experiment as being varied by varying the thermal Rossby number.

A typical atmospheric value of R_0 is about 0.03. Except for low values of T_a, Hadley flow is generally not observed in the dishpan for values of R_{OT} below 0.3. The experiments therefore definitely imply that the atmospheric circulation should lie within the Rossby régime.

The annulus experiments

At the time that the early experiments at Chicago were in progress, Hide (1953) was performing somewhat similar experiments at Cambridge University, using a deep annular container rather than an open dishpan. Hide was interested in simulating the Earth's core, but he recognized the meteorological significance of some of his results, and ultimately his experiments influenced the work of all others in the field.

From the meteorological point of view Hide's most significant discovery was probably that of the occurrence of flow patterns containing chains of identical waves, which progressed about the axis at a uniform rate without altering their shape, in sharp contrast to the patterns in the open dishpan, which seemed no more regular or periodic than the atmosphere itself. Theoretical meteorologists had been using such waves as mathematical idealizations since Rossby (1939) had introduced them in his famous paper, and the idea was sometimes expressed that since the patterns in the atmosphere were always irregular, Rossby's sinusoidal waves were no more than fantasy. Hide's experiments definitely showed that Rossby's ideas could apply to real fluid systems, which, although lacking the unpredictability of the atmosphere, were at least driven by an analogous mechanism.

More important from the practical point of view was the opportunity provided for detailed experimental measurements. Although motions at the free surface of the dishpan can be measured photographically, it is generally not feasible to scatter thermometers or "anemometers" throughout the interior, and the temperatures are measured at only a few interior points at once, while the internal flow may not be measured at all. When the flow is non-periodic, only the long-term statistics can be measured in detail, and these require an experiment of extended duration, sufficient to allow measurements during a representative collection of "weather situations". There is a slight advantage over real atmospheric studies in that the long-term statistics should not vary from one longitude to another.

Waves moving without changing their shape form a steady-state flow in a co-ordinate system moving with the waves. Aside from experimental errors, a single photograph is sufficient to measure the free-surface velocities, while an instantaneous temperature field can be constructed from temperature measurements taken a few at a time. This possibility has been exploited by Riehl and Fultz (1958), who have determined three-dimensional distributions of temperature and motion in considerable detail. Even when only the statistics are desired, the labour is greatly reduced by the elimination of the major sampling fluctuations.

The flow patterns discovered by Hide form a sub-régime of the Rossby régime. This may be called the "steady Rossby régime", although it must be noted that this term has also been used to describe an ordinary Rossby régime which has attained a statistically steady state following the modification of some external parameter.

When all the waves are identical, there must be a definite number of waves. The wave number is then one characteristic of the flow. It is possible within the same annular apparatus to change the number of waves by changing the external conditions. In general, lowering the thermal Rossby number increases the wave number. There are well-marked transitions from one wave number to another, so that the steady Rossby régime may be further divided into sub-régimes, one for each wave number.

Figure 57 shows the Hadley-Rossby transition, and the transitions between wave numbers, as determined experimentally by Fultz *et al.* (1964) using an annular apparatus like the one first used by Hide. (We shall presently consider the somewhat similar Figure 58.) To obtain the transitions, the rotation rate was held fixed, while the heating contrast, starting at zero, was increased in steps until the Hadley régime was re-established. The procedure was then repeated for different fixed rotation rates. It is noteworthy that the wave-number transition curves are marked by nearly constant thermal Rossby numbers. (For an annulus, the factor a^{-2} in (134) must be replaced by $a^{-1}(a-b)^{-1}$, where b is the inner radius.)

A further striking discovery of Hide's was a phenomenon which he named "vacillation". Here the waves are not steady in a moving co-ordinate system, but they alter their shape and speed of progression in a regular periodic manner, returning to their original configuration at the completion of a vacillation cycle. This phenomenon aroused immediate interest among meteorologists because of its resemblance to the fluctuations of the zonal index, a quantity introduced, incidentally, by Rossby in the paper just cited. Like the steady waves, the idea of an index cycle had been criticized as an over-idealization. It was pointed out, for example, that the variations of the circulation were more nearly random than cyclic. Hide's experiments clearly showed that regular predictable fluctuations in real fluid systems were by no means preposterous.

Since a vacillating pattern repeats at regular intervals, it is possible in this case also to measure the three-dimensional temperature field at any phase of the vacillation cycle, and thence to obtain long-term statistics for a non-steady circulation, without the usual sampling difficulties. The three-dimensional wind field is most readily determined by measuring the wind at the free surface, and deducing the wind below from the thermal wind equation.

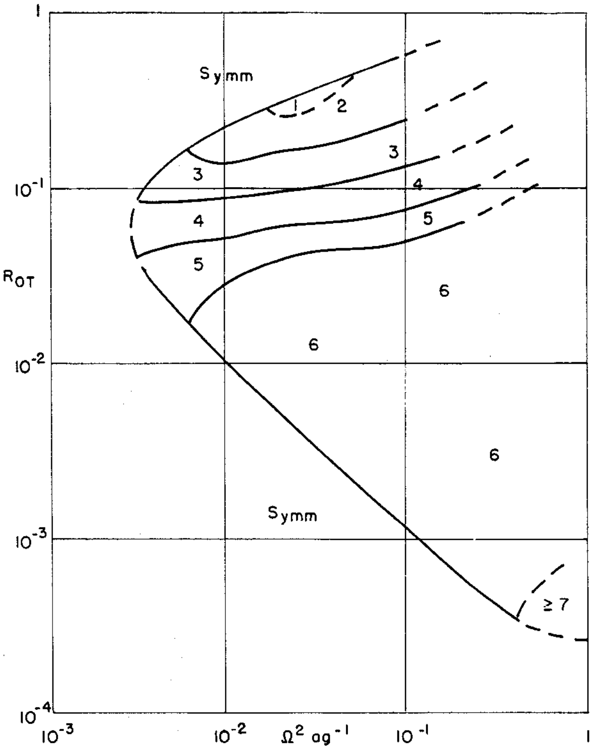
Figures 59-62 show the appearance of the free surface at four phases of a vacillation cycle in an experiment recently performed by Fultz. (A somewhat similar experiment is illustrated by Fultz *et al.*, 1959, pp. 94-95.) The vacillation period in this case was 16¼ revolutions; the successive photographs are separated by 4 revolutions. Almost as striking as the pronounced change in the shape of the waves from one phase to another is the nearly identical shape possessed by the five separate waves at each phase.

In Figure 59 the nearly due-north winds to the left of the troughs, together with the south-westerly winds to the right, indicate a strong northward transport of angular momentum. By the time of Figure 60 this transport has ceased and reversed its sign. The increased westerlies at lower latitudes and decreased westerlies at higher latitudes, which result from the southward momentum transport, manifest themselves in the transformation of the open troughs into closed cyclonic centres in Figure 61. By this time the southward transport of momentum has ceased, and the transport has become decidedly northward in Figure 62, so that four days later, when the pattern is again as in Figure 59, the troughs have opened up again. Measurements have revealed that the poleward transport of heat undergoes similar fluctuations during the vacillation cycle.

Implications of the experiments

Entirely aside from any resemblance which they may bear to the atmosphere, the experiments pose a basic theoretical problem which demands an explanation. This problem concerns the reasons for the existence of separate régimes of flow, and the abrupt transitions between them. Of main interest is the transition between the Hadley and Rossby régimes. For definiteness the "upper transition" appearing in the upper portion of Figure 57 may be considered. The following discussion must be regarded as largely descriptive rather than truly explanatory.

Figure 57. — Transitions between Hadley (symmetric) and Rossby régimes, and transitions between wave numbers within the Rossby régime, occurring in rotating-annulus experiment of Fultz *et al.* (1964), as the thermal Rossby number R_{OT} is *increased* while the Taylor number remains fixed. Dashed curves indicate that the location of the transition is somewhat uncertain. Parameter $\Omega^2\, a\, g^{-1}$ is proportional to the Taylor number for a given apparatus with a given amount of fluid



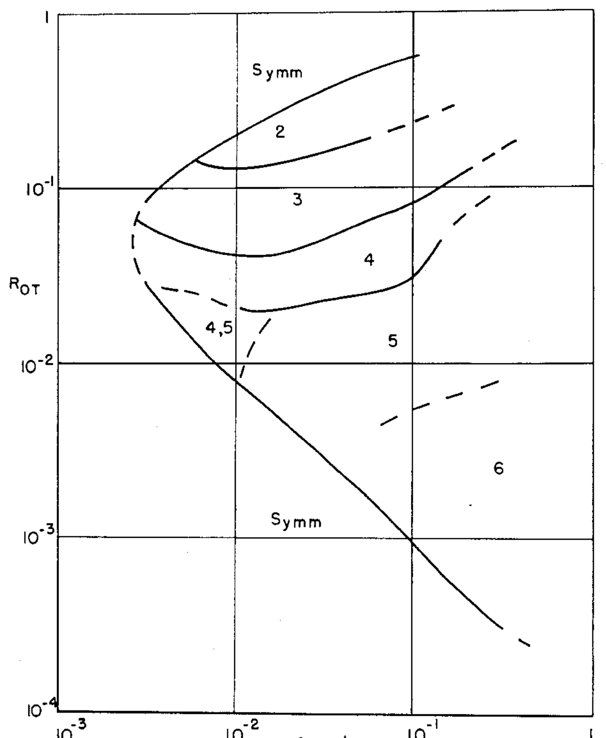


Figure 58. — Transitions between Hadley (symmetric) and Rossby régimes, and transitions between wave numbers within the Rossby régime, occurring in same rotating-annulus experiment as in Figure 57, as the thermal Rossby number is *decreased* while the Taylor number remains fixed

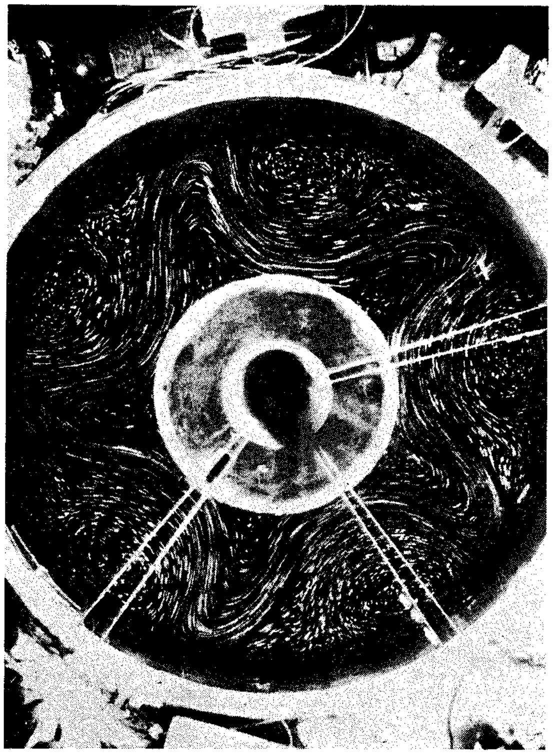


Figure 59. — Photograph of upper surface of fluid in a rotating annulus, showing a five-wave Rossby-régime circulation pattern. The photograph is a time exposure, so that particles of tracer on the upper surface appear as streaks indicating the direction and speed of flow. The circulation pattern is vacillating with a period of $16\frac{1}{4}$ revolutions (photo by Dave Fultz)

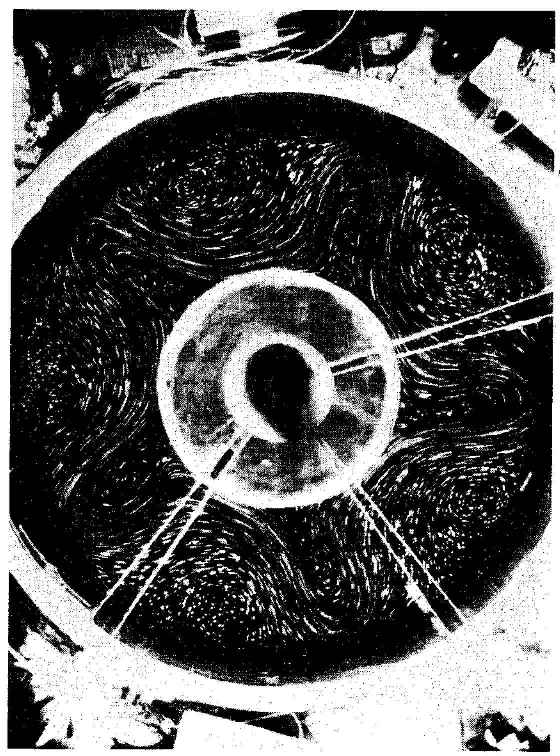


Figure 60. — The same as Figure 59, four revolutions later

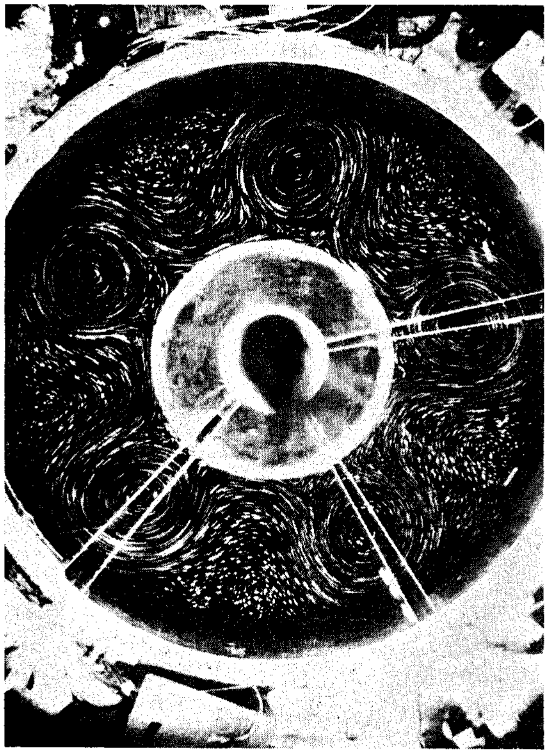


Figure 61. — The same as Figure 59, eight revolutions later

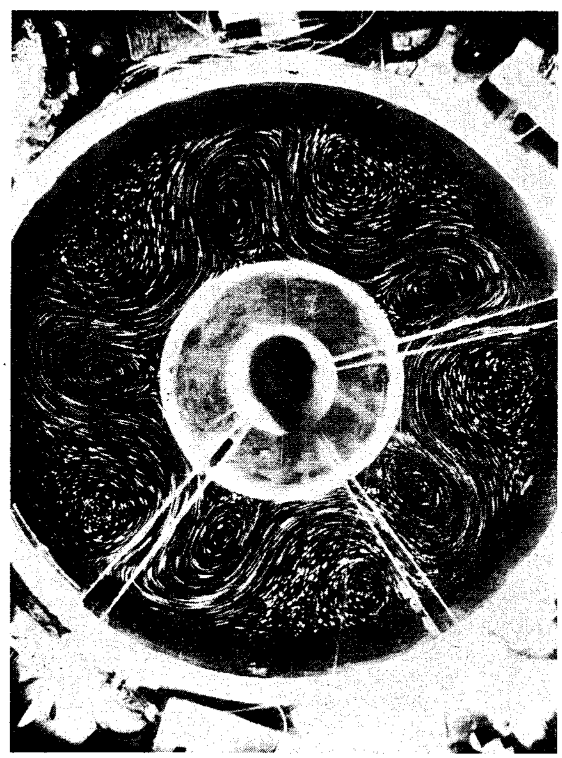


Figure 62. — The same as Figure 59, twelve revolutions later

It is evident that for any set of conditions under which zonally symmetric flow is observed, the flow must constitute a symmetric solution of the mathematical equations, and that moreover this solution must be stable with respect to sufficiently small perturbations. For any set of conditions where the Rossby régime is observed, there appear to be three different possibilities:

1. Symmetric flow is mathematically impossible; there is no steady-state symmetric solution of the equations.
2. Steady symmetric flow is mathematically possible, but it is unstable with respect to small asymmetric perturbations; these perturbations therefore develop into finite waves when the critical conditions are exceeded.
3. Steady symmetric flow is mathematically possible and is also stable, but the system is intransitive, possessing at least two "general circulations"; unsymmetric flow is possible also.

The first of these possibilities had been suggested during the course of the early experiments, but the arguments in favour of it do not seem very strong. It may be recalled that once Hadley flow was found not to be present in the atmosphere, the arguments advanced to show the impossibility rather than the instability of Hadley flow proved fallacious. In any event either steady or oscillatory symmetric flow must be mathematically possible; one need but choose zonally symmetric initial conditions and the symmetry will be preserved.

The second possibility, which is much like the argument presented by Bjerknes (1921) to account for the large disturbances in the atmosphere, was proposed in connection with the dishpan experiments by Lorenz (1956). He noted that for symmetric flow the primary effect of stronger heating would be to produce a stronger temperature contrast, with accompanying stronger vertical wind shear. According to the usual criteria for baroclinic instability (which is discussed in more detail in Chapter VIII), this should favour instability rather than stability. He therefore sought a second-order or non-linear effect through which stronger heating could increase the stability. Such an effect appears to be the transport of heat by the meridional circulation, which conveys warm fluid across the top and cold fluid across the bottom, and thereby creates a stable stratification, which favours baroclinic stability. Since the vertical stability is proportional to both the strength of the meridional circulation and the temperature contrast, it is proportional to the square of the heating contrast, neglecting still higher-order effects, and for sufficiently strong heating it should offset the destabilizing effect of the vertical shear.

However correct this explanation may be in some instances, more recent experiments have shown that for certain combinations of rotation and heating the third explanation is the proper one. Fultz *et al.* (1959, p. 78) describe one case in detail. In the particular apparatus, the Rossby régime ordinarily occurred when R_{OT} fell below about 0.5. However, when R_{OT} was decreased very carefully by increasing the rotation rate in small steps, allowing time for adjustment of the flow between steps, the Hadley régime was preserved until R_{OT} fell to about 0.3. Between these two values, the symmetric circulation was stable in the ordinary sense, since moderate disturbance would not destroy it. Yet when the water was stirred violently with a rod for about one second, a pattern of three waves developed and remained. This was evidently the same flow which would have developed in any case if less care had been taken in decreasing the rotation rate.

Initial conditions consisting of a symmetric flow plus small but not infinitesimal disturbances have a finite rather than a zero probability of being selected by chance. More irregular initial conditions leading to the Rossby régime also have a finite probability of being selected by chance. The experiment therefore provides concrete evidence that thermally forced rotating fluid flows may be intransitive, and incidentally suggests that one is not necessarily on safe ground in assuming that the atmosphere is transitive.

Rossby flow is therefore not restricted to those instances when the mathematically possible Hadley flow is unstable. It may occur mainly when Hadley flow is unstable, but for some external conditions Rossby flow is simply an alternative to Hadley flow. The instability of Hadley flow is a sufficient condition for Rossby flow, but not a necessary condition.

In the open dishpan the principal abrupt transition is from the Hadley to the Rossby régime. The Rossby flow tends to be irregular and non-periodic, and transitions from one set of statistics to another, as the external conditions change, seem to be slow and continuous rather than abrupt. In the annular experiments, where many qualitatively distinct regular periodic flows occur, the transitions may be rather sharp. Here also, intransitivities have been experimentally observed.

The curves in Figure 57 show the transitions to lower wave numbers which occur in one particular apparatus as the heating is increased, while the rotation rate remains fixed. The transitions to higher wave numbers as the heating is decreased again are shown in Figure 58. These invariably occur at lower values of R_{OT}; thus there are external conditions under which either of two consecutive wave numbers may occur.

Like the Hadley-Rossby transition, the transitions between wave numbers may be described in terms of stability and instability. It is not sufficient in this case to consider only the stability of a steady-state symmetric flow; the stability of a time-dependent flow with respect to still further disturbances is involved. Consider two adjacent wave numbers, say numbers three and four. An equilibrium Hadley flow may be unstable with respect to disturbances having wave number three only, or four only, or it may be unstable with respect to both disturbances, or neither. If it is unstable with respect to both wave numbers, a Rossby flow containing three waves, and another one containing four, will each be mathematically possible. These Rossby flows, as opposed to the mathematically possible Hadley flow alone, may be unstable with respect to further disturbances having respectively four or three waves. If the three-wave flow is unstable with respect to four-wave disturbances, and vice versa, neither three-wave nor four-wave flow can exist by itself experimentally, and (barring still further wave numbers) the resulting flow will be somewhat irregular, as in the open dishpan. If the three-wave and four-wave flows are each stable with respect to disturbances of the other wave number, either pattern can occur and persist, and the system is intransitive. This is evidently what happens in the annulus near the wave number transitions. If only one flow is unstable with respect to disturbances having the other wave number, only the stable flow will be observed. This evidently occurs in the annulus away from the transitions.

Similar considerations seem to characterize the phenomenon of vacillation. When a wave pattern is observed to move without changing its shape, the pattern is stable with respect to further disturbances. Under other conditions, steady waves may still be mathematically possible but unstable. If the pattern is unstable with respect to further disturbances having the same wave number but a different shape, vacillation may be mathematically possible. In that event it will be possible to observe the vacillating flow, unless it is also unstable with respect to still further disturbances, in which case the resulting flow will presumably be rather irregular.

The flow in the dishpan and the annulus has received its most extensive theoretical treatment from Kuo (1957) and Davies (1959). Rather complicated mathematics is involved. That the pertinent physical processes may nevertheless be rather simple is indicated by a study by Lorenz (1962), who was able to establish a highly simplified system of equations giving a crude description of the experiments. The system consists of only eight ordinary differential equations, and it can be solved analytically for the Hadley flow under all conditions, and for the Rossby flow under those conditions where such a flow exists. He obtained a Hadley-Rossby curve looking very much like Fultz's, although it was difficult to assign an absolute scale. As he had postulated earlier, the vertical stability was the controlling factor in rendering the Hadley flow stable for strong heating. Moreover, the intransitivity of the system near the Hadley-Rossby transition was reproduced.

With the addition of four more variables, Lorenz also obtained curves for the transitions between pairs of consecutive wave numbers. These bore a fair resemblance to the experimentally determined transitions shown in Figures 57 and 58. The intransitivity near the wave-number transitions was not reproduced; instead there were conditions where two wave numbers occurred simultaneously. Most likely the Rossby flow was too crudely represented for its stability with respect to further disturbances to be properly described.

The laboratory experiments presumably tell us more about planetary atmospheres in general than about the Earth's atmosphere in particular. They indicate the variety of flow patterns which can occur, and the conditions favourable to each of these. Since regular flow patterns, other than Hadley flow, occur mainly in the annular experiments, and since an open dishpan is presumably a better planetary analogue than an annulus, the experiments suggest that large-scale planetary flow patterns should be confined to Hadley flow and irregular Rossby flow.

Among the controlled experiments which deserve special mention is one designed to explore the original hypothesis of Halley (1686) to the effect that the trade winds are produced by the diurnal progression of the most strongly heated region about the Equator, rather than by any deflective effect of the rotation. Fultz *et al.* (1959, pp. 36-39) describe an experiment in which a flame was moved in a circular path underneath the rim of a stationary dishpan. In due time westerly winds, i.e. motion opposed to the direction of the flame, developed at the upper surface. At the bottom these were easterlies near the rim and westerlies near the centre, while a single direct meridional cell occupied the entire dishpan. Presumably the corresponding effect in the atmosphere is insignificant by comparison with the deflective effect of the rotation, but qualitatively the effect hypothesized by Halley seems to be verified.

Perhaps the most important contribution of the laboratory experiments to the theory of the atmosphere has been the separation of the essential considerations from the minor and the irrelevant. Condensation of water vapour, for example, may yet play an essential role in the tropics, where the circulation has not been well modelled, but in temperate latitudes it appears to be no more than a modifying influence, since systems occurring in the atmosphere, including even cyclones and fronts, occur also in the dishpan, where there is no analogue of the condensation process. Similar remarks apply to the topographic features of the Earth, which are intentionally omitted in most of the experiments. The so-called β-effect — the tendency for the relative vorticity to decrease in northward flow and increase in southward flow because of the variability of the Coriolis parameter — now appears to play a lesser role than had once been assumed. Certainly a numerical weather forecast would fail if the β-effect were disregarded, but the β-effect does not seem to be required for the development of typical atmospheric systems.

On the other side of the ledger, the experiments emphasize the necessity for quantitative considerations; at the very least these must be sufficient to place the atmosphere in the Rossby régime. The most that a completely qualitative treatment can do is to establish the separate properties of the Rossby and Hadley régimes, and then state that the atmosphere conforms to one or the other of these.

CHAPTER VII

NUMERICAL SIMULATION OF THE ATMOSPHERE

The recent development of laboratory models, which has provided a new tool for the investigation of the atmosphere, has been followed closely by the development of a further tool — the numerical model. Actually a numerical model is a system of mathematical equations, designed to resemble the equations governing the atmosphere, and arranged in a form suitable for solution by numerical methods. A numerical experiment consists simply of the determination and examination of a particular time-dependent solution of the equations, starting from some chosen initial conditions. Yet the procedure for performing a numerical experiment has become highly specialized, and it has become customary to regard a numerical model as a new physical system whose behaviour simulates that of the atmosphere, rather than simply an approximate mathematical means for studying the atmospheric system itself.

Numerical experiments are almost invariably performed with the aid of high-speed digital computing machines. The required amount of computation in most meaningful experiments is so great that slower methods are wholly inadequate.

Size and time limitations

Although it may be argued that the atmosphere is really a finite system containing about 10^{44} molecules, and that it may therefore be represented by a finite collection of numbers, the usual representation of the atmosphere as a continuum is far more realistic than any approximation which attempts to represent the atmosphere by a small finite set of numbers. The equations governing the continuous atmosphere may therefore be regarded as the exact equations. On the other hand a digital computer has a finite capacity, and operates at a finite speed. It therefore cannot describe the atmosphere as either a continuous or a continuously varying system.

The state of the atmosphere at any one time t must somehow be represented by a finite set of numbers, say m numbers $X_1, \cdots, X_m$, if the equations are to be solved numerically as an initial value problem. In most models these numbers have been the values of the meteorological variables at a previously chosen three-dimensional grid of points, with the values between the points being implied by some interpolative or extrapolative scheme. The interpolations are not actually carried out, but, as a part of the model, the exact equations are replaced by approximate expressions for the time derivatives $dX_1,/dt, \cdots, dX_m/dt$ at the grid points, in terms of the values $X_1, \cdots, X_m$ at the grid points.

Other schemes afford more economical representations of the state of the atmosphere, but the corresponding modified equations are generally more awkward to handle. In the most common alternative scheme the field of each variable is expressed as a linear combination of a set of previously chosen functions, such as spherical harmonics, and the m numbers are the coefficient appearing in these combinations.

The m numbers cannot be varied continuously during an experiment but must be changed in a finite number of steps, say n steps. Generally a time increment Δt is chosen, and the values of the m numbers at time $t + \Delta t$ are approximated by a formula such as

$$X_j(t + \Delta t) = X_j(t - \Delta t) + 2\,\Delta t\, dX_j/dt\,, \tag{136}$$

or often by a more sophisticated scheme. If an average of k arithmetic operations is required to compute one time derivative, the complete experiment requires a total of kmn arithmetic operations.

The mn numbers so generated may then be treated as observational data. They may be used to construct a series of simulated weather maps, or to compute any desired statistic such as a long-term average. For the latter purpose they have one obvious advantage over real data in that there should be no missing observations.

If an investigator wished to perform only one numerical experiment, he might be willing to let the machine compute for a year. Certainly comparable times have been spent in analysing the numerical data after they have been generated. Yet greater benefits are to be anticipated from a series of experiments, where certain factors may be varied from one experiment to another. The day when large computers will become so inexpensive that a meteorological centre can afford to own several of them and perform several experiments simultaneously does not seem to be near at hand. More likely there will be additional demands upon a single computer. A few weeks of computation therefore appears to be a practical upper limit for most numerical experiments.

The fastest computers in use today can perform a single arithmetic operation, such as the addition or multiplication of two several-digit numbers, in about 10^{-6} seconds, and can thus perform about 10^{12} operations in the course of two weeks, if they are kept running continuously. With further improvement in computer technology this figure is likely to increase. By comparison, a person working unaided would be unlikely to perform more than 10^4 operations during the same period.

The practical limit for the product kmn is therefore about 10^{12}. In a typical experiment about 100 operations are needed to compute a single time derivative of a single quantity, so that mn is limited to about 10^{10}. Assuming that an experiment must simulate at least a few months of atmospheric behaviour — an interval generally needed to obtain meaningful statistics from real observations — and assuming tentatively that the time increment Δt may be a few hours, a lower limit for n is about 10^3, whence an upper limit for m is about 10^7.

In the most elaborate numerical model so far investigated, Manabe *et al.* (1965) have represented the state of the atmosphere by about 50 000 numbers. If one can imagine using the upper limit of 10^7 numbers, these might reasonably be chosen as the values of five atmospheric variables at each of 20 elevations at each point in a horizontal grid of 10^5 points. There would then be one grid point for each 5000 square kilometres of the Earth's surface.

Undoubtedly there are particular solutions of the exact equations in which the fields of the variables are so smooth that they would be adequately described by interpolation between the 10^5 grid points, but observations show that these are not the solutions chosen by the real atmosphere. The air is filled with such disturbances as thunderstorms, cumulus clouds, and smaller turbulent eddies. The detailed structures of these systems cannot be described by the values of the variables at a coarse grid. Any numerical solution performed today must therefore apply to an atmosphere which has been idealized to the extent of omitting all motions of thunderstorm scale or smaller.

It might appear that with an eventual increase in the speed of computing machines by a factor of 10^3, which is not out of the question, it would become possible to introduce one grid point for every five square kilometres, and include at least the larger cumulus clouds. There is a further reason, however, aside from the large number of dependent variables demanded, why these features must be omitted. In any realistic solution each variable at each grid point will undergo oscillations about some mean value.

An instantaneous time-derivative affords a good approximation to the average time-derivative during a small fraction of a period of oscillation, but it gives a very poor approximation for a full period, let alone several periods. Even under the more sophisticated computation schemes, small inconsequential irregularities which should rapidly die out will be described as intensifying, and ultimately dominating the circulation, if Δt is as great as one fourth of the period of oscillation of these irregularities. The time increment must therefore be made small enough to accommodate the oscillations of all the systems which are retained, and it becomes as essential to eliminate rapidly oscillating systems, to keep n small, as it is to eliminate small-scale systems, to keep m within bounds.

The oscillations at an individual point may result from the propagation of waves or wave-like motions through the air, or from the simple displacement of some irregularity by the air motion. Local fluctuations due to turbulence are largely of the latter type; the period of oscillation is at least comparable to the time required for the smoothed wind-field to carry a turbulent element through its own length. Inclusion of even the crudest representation of cumulus clouds would therefore limit Δt to about a minute, while smaller-scale turbulence could limit it to a fraction of a second.

With Δt reduced to a minute or less, n would be raised to 10^5 or more. If even the largest cumulus clouds are to be carried as part of the global circulation in a numerical model, the speed of computation must increase by a factor of at least 10^5 above its present maximum; smaller cumulus clouds would require an even greater increase.

Nevertheless, cumulus clouds and smaller-scale eddies are instrumental in the vertical transfer of angular momentum, water, and energy, and the numerical solution must somehow incorporate their effects if it is to serve its purpose. This is ordinarily accomplished by relating the effects of these systems to the large-scale motions on which they are superposed, through the use of coefficients of turbulent viscosity and conductivity. Since the intensity of the smaller-scale systems is in turn affected by the larger-scale fields, the coefficients are better approximated by functions of the numbers at the grid points than by constants. Determination of suitable approximations is a problem which is still far short of solution.

Since cyclones and other disturbances of similar size, which are instrumental in the horizontal transports of angular momentum, water, and energy, generally do not follow one upon another by less than a day, it might appear that without the smaller-scale systems the solution could proceed in six-hour time increments. In practice this is not the case. Very small random errors, such as those introduced by round-off, will be interpreted by the computational scheme as small-amplitude disturbances, to be either carried along with the motion of the air or propagated through the air as waves. Once introduced, these disturbances will remain as part of the numerical solution. If they fail to grow they will cause no difficulties, but if Δt is greater than about one fourth of the period with which these fictitious disturbances ought to oscillate, according to the governing equations, the errors will amplify and finally dominate the field.

In general the errors will be interpreted as superpositions of simpler disturbances (normal modes: see Chapter VIII) having lengths as short as four grid intervals. (Waves with lengths between two and four grid intervals will appear in a spectral analysis, but under the usual differencing processes they will not oscillate as rapidly as those which are just four grid intervals long.) It follows that the maximum allowable value of Δt is the time required for the air to move one grid interval, or for a wave to travel one grid interval. This restriction is the well-known Courant-Friederichs-Lewy criterion for computational stability. For a more rigorous treatment the reader is referred to the book of Thompson (1959) or a shorter summary article by Phillips (1960).

The limitation upon Δt is dictated first of all by those waves which move much more rapidly than the speed of the wind. The most rapidly propagating waves are sound waves and external gravity waves. These waves are of questionable importance in the total circulation, but they soon become very important in an unstable computational scheme. It is possible to introduce further approximations which effectively modify the equations so that these waves are incapable of being propagated.

The most troublesome waves are vertically travelling sound waves, since the grid interval in the vertical is necessarily small compared to the horizontal interval. Even with minimum vertical resolution (two layers) they would limit Δt to about half a minute. As noted in Chapter II, they are completely eliminated by using the hydrostatic equation in place of the exact vertical equation of motion. This approximation is in general use in any case.

Gravity waves may be eliminated in a number of ways, the simplest of which is the use of the geostrophic approximation, in the form which equates the vorticity of the wind to the vorticity of the geostrophic wind, in place of the divergence equation. The more cumbersome but more realistic equation of balance would have a similar effect. Only the external gravity waves move at the most extreme speeds, and they may be eliminated, while internal gravity waves are retained, by simply requiring the vertically averaged divergence to vanish.

Whereas the use of an extremely short time-increment would suppress only the spurious sound and gravity waves and leave the ones which should actually occur, the hydrostatic and geostrophic approximations, with or without a very short time-increment, remove all sound and gravity waves. The approximations can therefore be justified only if these waves have very little effect upon the component of the circulation which is being studied. This indeed appears to be the case for sound waves; for gravity waves it has been claimed that this is the case, but the conclusion is less certain.

Even with the hydrostatic and geostrophic approximations, a six-hour time-increment is not possible in practice. Cyclones ordinarily travel more slowly than the wind, while fictitious disturbances other than sound and gravity waves, but not having the typical structure of cyclones, can be carried with the wind. Equally important, the minimum horizontal resolution needed to represent the presence and propagation of cyclones is considerably coarser than that needed for a good representation. Smagorinsky *et al.* (1965) found that a 250-kilometre grid interval gave noticeably more realistic results than a 500-kilometre interval. This was apparently not so much because somewhat smaller-scale features were admitted, but because the larger-scale features were more accurately depicted. With a reasonable horizontal resolution a one-hour time-increment appears to be near the upper limit.

Other frequently used idealizations afford minor savings in computation time, but their main purpose is simplicity rather than economy. One of these is the treatment of air as an ideal gas. Water in its various phases, with its consequent thermodynamic and radiational effects, is completely omitted. Another approximation treats the Earth's surface as completely homogeneous. Oceans and continents with their contrasting thermal capacities, and mountains and valleys with their contrasting mechanical influences, are eliminated. A still further approximation regards the solar heating as a function of latitude only. A hypothetical average sun is introduced, and the diurnal and annual variations about this average are disregarded. In many of the earlier models, the beta-plane approximation (see Chapter II) has been used.

Numerical weather prediction and the first experiment

The earlier stages of the development of numerical simulation depended almost completely upon the prior development of numerical weather prediction — the direct application of the governing equations to the problem of weather forecasting. The procedures for numerical simulation are almost the same, with two principal differences. First, in numerical weather prediction the initial conditions must be chosen to represent the current weather situation, while in numerical simulation they may be chosen for maximum convenience. Second, in numerical weather prediction the solution is usually not extended beyond a few days, while in numerical simulation it must cover a few months at the least.

The latter distinction has considerable bearing upon the forms of the equations which may be used to advantage. There is no reason why a system of equations which gives no meaningful results at long range cannot give rather good short-range predictions. Consequently, for several years following the first reasonably successful numerical weather forecast, heating and friction were omitted altogether from the equations. To be suitable for simulating the global circulation, these equations would have to be modified at least to the extent of adding terms representing heating and friction.

The first serious attempt at numerical weather prediction was the remarkable work of Richardson (1922). Using a rather complete formulation of the primitive equations, Richardson performed a single six-hour forecast for the European area. His forecast completely failed to agree with the observed development, and he attributed the failure to inaccuracies in the initial wind measurements. It is true that the wind measurements were inadequate, but Richardson was also unaware of the phenomenon of computational instability.

Digital computers were unknown at that time, and Richardson spent many months preparing his single forecast. He visualized the establishment of a meteorological centre where 64 000 persons working together would be able to predict the weather as fast as it occurred. It is only recently, incidentally, that a single machine has attained the speed of 64 000 persons.

For a number of years afterward potential investigators were discouraged by Richardson's lack of success. It even appeared that sufficiently accurate initial wind-measurements might be impossible. A promising procedure in which accurate wind measurements were not required was eventually proposed by Charney (1947), and soon afterward Charney, Fjørtoft, and von Neumann (1950) produced the first moderately successful numerical weather forecast with a digital computer.

The forecast was based on a "one-level" model (see equation 58) in which the wind field appeared at only one level, and the temperature did not appear as a dependent variable at all. It would have been possible to modify the model by adding friction, but with no thermodynamic equation it would not have been impossible to add thermal forcing. With the introduction by Phillips (1951) of a "two-level" model, in which the wind field appeared at two levels, while a single temperature field was related to the difference of the wind fields through the thermal wind relation, the stage was set for numerical experiments of extended duration.

The original numerical simulation of the global circulation was the famous experiment of Phillips (1956). Phillips used a two-level model which contained the hydrostatic and geostrophic approximations and all of the other idealizations cited above. He chose a region bounded by parallel walls 10 000 kilometres apart, and within the region he constrained each variable to repeat itself every 6000 kilometres in the east-west direction. The external heating was simply a linear function of latitude, and friction was a linear function of the wind field extrapolated to the surface.

Phillips first suppressed all variations with longitude, and allowed a Hadley-type circulation to develop. The circulation lacked the contrasting trade winds and prevailing westerlies, and instead possessed weak easterlies everywhere at the bottom. These also would have disappeared if the solution had been allowed to attain full equilibrium, since, in conformity with the geostrophic approximation, the transport of momentum by the divergent component of the wind had been suppressed, and nothing was left to balance the surface frictional stress.

As initial conditions for the main experiment, Phillips chose the Hadley circulation plus small random perturbations which varied with latitude and longitude. In the course of about five days a system of organized waves developed, and then gradually increased in intensity. During the course of the integration the extreme wind speed also increased, and to maintain computational stability Phillips had to reduce the time-increment, which had at first been two hours, to one hour and then half an hour. After about thirty days the errors introduced by the computational procedure seemed to render the solution rather meaningless.

The experiment was remarkably successful in reproducing certain features of the circulation. At the surface easterlies soon appeared in low and high latitudes, with westerlies in between, and persisted throughout the experiment. Aloft the westerlies showed a distinct tendency to culminate in a jet stream. Moreover, the detailed energy cycle was qualitatively like the one observed in the atmosphere, and the various conversion processes possessed the right orders of magnitude, even if not the precise numerical values.

Unlike the real atmosphere there was no essential difference between the trade winds and the polar easterlies, but actually this similarity was demanded by the simplicity of the model. Certain symmetries had been built into the equations. Specifically, corresponding to any time-dependent solution, there exists another time-dependent solution in which the field of the eastward wind component is the mirror image, in the line midway between the extreme latitudes, of the same field in the former solution. Hence, if the equations are transitive, the trade winds and the polar easterlies must have equal average intensities. If they are intransitive, and one set of statistics possesses stronger trade winds, another set will possess stronger polar easterlies.

Yet the geometry alone does not demand easterlies and westerlies where they occur; the westerlies might have formed at high and low latitudes, with easterlies between. Thus, to the extent allowed by the constraints, the model duplicates the trade winds and prevailing westerlies, and suggests that their basic physical cause may not have been eliminated by the various approximations.

The experiment revealed a form of computational instability which had not been suggested by the shorter-range numerical forecasts. If the system of equations had been linear, a disturbance could have been carried along by a preassigned flow, but not by a flow which was itself a variable of the system. A suitable time-increment could then have been chosen once and for all. For a non-linear system the appropriate Δt decreases whenever the maximum wind increases. If for computational reasons the system continues to gain kinetic energy, even at a slow rate, a computation which has appeared to be stable for perhaps many days will suddenly become unstable as the critical wind speed is exceeded.

Before numerical experiments could be carried out over indefinitely long intervals, this non-linear instability had to be eliminated. Phillips (1959) has discovered one way in which the horizontal differencing procedure can produce a spurious gain in kinetic energy, and has found that the instability may be eliminated by periodically subtracting all disturbances having wavelengths of less than four grid intervals. A somewhat similar method consists of introducing a non-linear viscosity, which is very effective in damping the smaller-scale systems while having rather little effect on the larger ones. Another successful solution to the problem has been provided by Arakawa (1966), who has chosen a finite-difference formulation of the horizontal advective processes which strictly conserves the kinetic energy. Leith (1965) has solved the problem by a Lagrangian formulation of the advective processes, which seeks the location at time $t - \Delta t$ of a point which will be carried by the flow to a standard location at time t.

Recent numerical experiments

Some of the more recent simulations have attempted to remove the less realistic simplifications used by Phillips. Smagorinsky (1963) returned to the primitive equations, although still excluding "external" gravity waves. He used a spherical geometry, but restricted the flow to a region bounded by the Equator and 64.4° north (= arcsin 0.9). He extended the integration for 60 days in 20-minute time increments.

This model yielded a considerably more realistic representation of the trade winds and zonal westerlies. Irregular systems of five or six well-developed waves appeared in middle and higher latitudes, while the tropics contained many more disturbances of much smaller amplitude and horizontal scale.

A considerable advance was made by Leith (1965), who integrated a six-layer primitive-equation model containing evaporation, condensation, and precipitation. A novel feature of his numerical solution was a computer-produced motion picture, where the growth, modification, and decay of individual disturbances could readily be followed.

By far the most detailed numerical simulations to date are the experiment by Smagorinsky *et al.* (1965) with a dry atmosphere, and an accompanying experiment by Manabe *et al.* (1965) with a moist atmosphere. The former study permits external as well as internal gravity waves, and covers an entire hemisphere, but its principal refinement is the representation of the vertical structure by nine levels, which, with respect to pressure, are most closely packed at the bottom and top of the atmosphere. This allows a more realistic treatment of the surface friction layer, and also allows the effects of radiation, including ultra-violet absorption by ozone in the highest layers, to be incorporated in a more sophisticated manner. The latter experiment contains, in addition to all the features of the former, a simplified hydrological cycle, including evaporation, advection of water vapour, and precipitation, although the radiational heating is still based upon the normal rather than the current distribution of clouds and water vapour.

The experiments were remarkably successful in duplicating many features of the real circulation. The model develops its own tropopause, and the complete temperature distribution is not unlike that actually observed, including a poleward increase in the lower stratosphere. The energy cycle is in qualitative and reasonable quantitative agreement with the real atmosphere, and, as appears to be the case in reality, the kinetic energy of the stratosphere is maintained through mechanical interaction with the troposphere. The latter model yields a realistic over-all precipitation rate, although the maximum in the tropics is too intense. Altogether the energetics are satisfactory, although there seems to be too little eddy kinetic energy. The improvements to be expected from further refinements are largely quantitative, although the refinements themselves will necessarily have to be qualitative.

Another outstanding large numerical simulation is that of Mintz (1964). In conjunction with the models of Smagorinsky *et al.* and Manabe *et al.* it is especially valuable in that it includes many basic features which the other models omitted. Mintz used only two layers in the vertical, but he included the large-scale topography of the Earth, and also the oceans and land surfaces. The oceans were assumed to have infinite heat capacity, so that the ocean surface temperatures were prespecified quantities. The land, and also the sea ice, were assumed to have zero heat capacity, so that any heat received by them was immediately transferred to the atmosphere.

Mintz first performed an experiment without the mountains, and then at a certain point suddenly introduced them. According to Mintz (1964, p. 146), "As soon as this was done the air began to pour down the mountain sides (as would water down the sides of an island emerging from the sea), producing large gravity waves. After some days these large external gravity waves died out and only the familiar meteorological motions remained..."

Mintz found that the mountains had little effect upon the general type of behaviour of the atmosphere, and in particular upon the total kinetic energy, but that, together with the land-ocean contrast and the predetermined sea-surface temperatures, they had a considerable influence upon the geographical locations at which various features occurred. The time-averaged sea-level pressure field agreed rather well with observations, as did the upper-level temperatures averaged over longitude and time.

A number of other investigators have recently become engaged in numerical experiments. An excellent comparative account of the work thus far performed in this field has been given by Gavrilin (1965).

From a comparison of the large numerical experiments we gain the impression that we can duplicate the behaviour of the atmosphere as closely as we wish simply by making the equations more and more realistic. The principal remaining physical problems are the proper representation of small-scale motions and the proper treatment of water in the atmosphere; these problems are compounded in the problem of cumulus convection. There seems to be no obvious reason why acceptable solutions cannot eventually be found.

It is therefore in order to ask what would be gained, other than the satisfaction of completing a challenging task, if we should eventually reproduce the general circulation in all its relevant details. Since we know in any case, by definition, that the correct equations, solved in a correct manner, will correctly reproduce the general circulation if the atmosphere is transitive, or will reproduce the correct circulation from a wide range of initial conditions even if the atmosphere is intransitive, the immediate result would be to reassure us as to the correctness of the equations and the method of solution. Our understanding of the general circulation would be increased only because many properties of the real circulation are not easily observed and measured, whereas an essentially complete numerical solution would contain the numerical value of every relevant variable, from which any desired statistics could be evaluated. Such questions as the details of the energy cycle could be settled once and for all.

The solution would not, however, necessarily increase our physical insight. The total behaviour of the circulation is so complex that the relative importance of various physical features, such as the Earth's topography and the presence of water, is no more more evident from an examination of numerical solutions than from direct observations of the real atmosphere.

As in the case of the laboratory models, the greatest value of the numerical models should lie in the opportunity for controlled experiments. The control may be of the type which is readily introduced in the laboratory, namely, the variation of one or more parameters, such as the intensity of the heating. The added power of the method lies in the possibility of altering not only physical features, such as the Earth's topography, but also entire physical processes, such as the propagation of sound and gravity waves, since the equations are readily modified so as to describe a system which is completely unrealizable in nature. In the laboratory, sound and gravity waves may be subdued by various means, but they cannot be completely eliminated. In a numerical model the hydrostatic and geostrophic approximations eliminate them altogether. The significance of these waves for the remainder of the circulation may then be readily assessed.

The performance of controlled comparative numerical experiments appears to have only begun. Most of the effort to date has been devoted to perfecting individual experiments. Perhaps we may anticipate the time in years to come when we may as a matter of course perform several experiments of the scale of the one performed by Manabe *et al.* whenever we wish to test a hypothesis. In the meantime the effort devoted to reproducing the atmosphere rather than changing it is also well spent. The more closely we are able to duplicate the atmosphere when such is our purpose, the more confidence we can later place in controlled numerical experiments in which systematic departures from the real atmosphere have been introduced.

CHAPTER VIII

THEORETICAL INVESTIGATIONS

The physical laws which govern the behaviour of the atmosphere, and the mathematical equations which represent these laws, form the basis for every theoretical study of the circulation. The mathematical techniques for handling these equations are numerous. With the recent outburst of interest in numerical simulation, and the growing anticipation that this method may eventually supply an answer to almost any question which may be asked of it, it is easy to lose sight of the fact that theoretical studies were an established part of meteorology long before the advent of the first digital computer.

Because numerical simulation seems to show promise of achieving a position of unique and long-lasting importance, and because, in contrast to the analytic expressions yielded by more classical mathematical techniques, the results of a numerical experiment consist of a collection of numbers to be processed further in the manner of observational or experimental data, we chose to discuss numerical simulation separately in the previous chapter, following the chapters on observational and experimental studies. In the present chapter we shall examine some of the more conventional procedures through which the governing equations may be applied to the study of the circulation, with particular emphasis upon those procedures for which a computing machine is not essential.

Analytic solutions of the dynamic equations

Perhaps the ultimate theoretical achievement would be the discovery of the general analytic solution of the exact equations. At present such an achievement is precluded by our inability to formulate such processes as friction and condensation in an adequate manner, but the discovery of the general solution of an idealized system of equations would be an almost equally satisfying accomplishment. Unless the equations have been so highly simplified that non-periodic solutions no longer exist, this accomplishment also appears to be impossible; the non-periodic functions characteristic of the atmosphere cannot be explicitly expressed in a finite number of symbols.

It is perhaps a matter of opinion as to how much simplification of the equations is permissible, but we feel that the following minimum requirements must be satisfied by the general solution of any system of equations, if this solution is to be regarded as depicting the general circulation and the processes which maintain it. First, there must be an energy cycle, in which heating produces available potential energy, reversible adiabatic processes convert available potential energy into kinetic energy, and kinetic energy is dissipated by friction. Second, there must be eddies, or departures from zonal symmetry. These eddies must for the most part transport sensible heat toward higher latitudes, and thereby act to reduce the poleward temperature gradient; they must also gain available potential energy from the zonally averaged field of mass. Likewise the eddies must for the most part transport angular momentum toward higher latitudes, and thereby contribute to the maintenance of low-level trade winds and prevailing westerlies; they must also give up some kinetic energy to the zonal circulation. To complete the angular-momentum balance, direct meridional circulations must appear in low latitudes and indirect circulations must appear in middle latitudes; these must also transport additional sensible heat and potential energy. It need hardly be mentioned that the transports of sensible heat and angular momentum are non-linear processes, and that the equations must be non-linear.

Possibly there is a system of equations whose general solutions is periodic and yet satisfies the above requirements. If so, it might be possible to obtain the general solution in analytic form. An intensive study of the solution might then disclose the basic reasons why the processes which it depicts also prevail in the real atmosphere.

In the absence of any assurance that the general solution of any suitable system of equations can be found, a less ambitious and more realistic goal is the determination of special analytic solutions. If the atmosphere is idealized at least to the extent of omitting all geographical features and all variations of external heating with time and longitude, one of these solutions is likewise independent of time and longitude. This solution describes the Hadley flow. It is not a good representation of the flow observed in the real atmosphere, but it is of interest for its own sake, and it plays an important role in current ideas regarding the general circulation.

Oberbeck (1888) was the first to attempt to obtain the Hadley solution in analytic form. In so doing he specified the temperature field rather than the field of external heating. Qualitatively his solution resembles the flow envisioned by Hadley (1735), with a single direct cell, but, as we noted in Chapter IV, the resemblance seems to be accidental.

Many years later Kropatscheck (1935) attacked the same problem, and obtained a meridional circulation looking very much like the flow visualized by Thomson (1857) and Ferrel (1859), with a shallow indirect cell in middle and higher latitudes and a single Pole-to-Equator direct cell filling the rest of the troposphere. Like Oberbeck he specified the temperature field, but his equations also included the thermal wind relation; thus his westerly winds were of the proper intensity, and his direct cell was of the proper strength for maintaining them. He adjusted the surface zonal winds so that the surface frictional torque would balance the vertically integrated transport of angular momentum.

For the friction layer, however, he assumed in advance that wherever there were surface easterlies, or westerlies, there would be a substantial equatorward, or poleward, drift superposed upon the meridional circulation which would otherwise prevail. Thus he was forced to obtain the shallow indirect cell, and his work cannot be taken as a demonstration that Thomson's or Ferrel's picture of the meridional circulation is preferable to Hadley's.

The reader will notice that we have not yet stated whether the meridional circulation which would actually prevail if large-scale eddies could be prevented from developing, would be the circulation envisioned by Hadley or the one pictured by Thomson and Ferrel, or perhaps some other circulation. This omission is intentional; we do not know the answer.

Since the real Earth possesses geographical irregularities which would render a circulation without eddies impossible in any case, the question can only pertain to the atmosphere of an idealized Earth. At the very least, to make the answer determinate we must specify whether the Earth's surface consists entirely of land or entirely of ocean. Certainly the field of heating will be profoundly affected by this choice.

Let us note, then, that in a certain sense any meridional circulation is possible. We may choose any temperature field, and an accompanying zonal wind field which approximately satisfies the thermal wind relation. We may superpose any meridional circulation upon this. We may then solve for the fields of friction and heating.

In general the friction and heating so determined will form rather unrealistic patterns. If we demand that the friction must bear a known relation to the motion, we can modify our procedure by choosing any meridional circulation and solving a linear equation, with variable coefficients, for the zonal circulation and the accompanying temperature. Alternatively we can choose a zonal circulation, which must satisfy the constraint that the total frictional torque is zero upon any region bounded by the Earth's surface and two surfaces of constant angular momentum, and solve for the meridional circulation. In either event we can then determine the heating from the thermodynamic equation.

The remaining problem will be that of finding an initial choice of the meridional or zonal circulation which leads to a reasonable field of heating. It seems altogether possible that some circulation like Hadley's and also some circulation like Thomson's or Ferrel's will serve this purpose.

In any event, a zonally symmetric solution does not describe the observed circulation satisfactorily. Among the special solutions of the idealized equations which may offer better representations, the simplest are those which describe a flow containing waves which progress without changing their shape, as in the steady Rossby régime in the laboratory experiments. These are steady-state solutions in a moving co-ordinate system, and it may well be possible to obtain them by analytic procedures. Next in simplicity are the vacillating solutions, where the waves alter their shape and speed of progression in a regular periodic manner. These doubly periodic solutions become simply periodic in a suitably moving co-ordinate system, and it does not seem unrealistic to hope that they also may be determined analytically.

Like the Hadley solution, the steady-wave solutions and the vacillating solutions, if they exist at all, are unstable with respect to still further disturbances, and are not the solution chosen by the atmosphere. If they are to be offered as a representation of the atmosphere, they should satisfy the same minimum requirements which we have demanded of the general solution. We cannot be certain that they will do so. The real atmosphere contains many irregularly shaped eddies. Some of these transport angular momentum and sensible heat poleward across middle latitudes, while others transport these quantities equatorward. The fact that the net effect of the eddies is a poleward transport is no assurance that the eddies in the steady-wave or even the vacillating solutions will transport angular momentum and sensible heat in this direction.

Nevertheless, it seems plausible that the physical processes which control the configurations of the eddies in the irregularly fluctuating atmosphere, and thereby determine the transports which they accomplish, may also act in a similar manner upon the eddies appearing in the unstable but mathematically possible steady-wave solutions. It is even more reasonable to assume that the average transports of angular momentum and sensible heat by the eddies in the vacillating solutions will be similar in direction and even in amount to the transports accomplished by the irregular eddies in the real atmosphere. If this can be shown to be the case, the discovery of the most general vacillating solution would approach the desired goal.

In view of the difficulties involved in determining even the Hadley solution analytically, the task of obtaining steady-wave solutions would appear to be rather formidable. Some of the successful attempts to solve equations which have been simplified beyond the point of describing the maintenance of the general circulation reveal some of the problems involved.

The most celebrated analytic solution of a dynamic equation in meteorological literature is undoubtedly Rossby's (1939) solution of the two-dimensional vorticity equation on the beta-plane. This solution exhibits the propagation of steady waves in the westerly-wind belt, and bears considerable resemblance to the subsequently discovered flow in some of the laboratory experiments. Nevertheless it is not intended to offer an explanation of the general circulation, since there is no heating nor friction, and the amplitude of the waves must be prescribed in advance.

Haurwitz (1940) extended Rossby's solution to include friction and forcing, but since the solution was two-dimensional, the forcing had to be mechanical rather than thermal. Haurwitz also obtained solutions for the complete sphere. Both Rossby and Haurwitz had used a linearized form of the vorticity equation, but Craig (1945) noted that their solutions also satisfied the non-linear equation. More recently Kuo (1959) has obtained three-dimensional analytic solutions for a simplified non-linear system of equations, again without heating and friction.

A feature of these non-linear solutions is that the eddies consist of a single mode of motion, i.e. they are of such a configuration that the advection which they accomplish produces no distortion. In the more general field of motion, superposed eddies of more than one mode will distort one another and thereby produce still further modes. Solutions possessing more than one mode of motion therefore possess an infinite number.

If the eddies do not distort the field of motion, they certainly do not alter the field of zonally averaged angular momentum. It follows that the solutions of the frictionless equations containing eddies of a single mode possess no convergence of eddy angular-momentum transport.

We know of no analytic solutions of even the most simplified equations which yield the proper eddy transports of sensible heat and angular momentum. In attempting to represent some of the dishpan experiments, Lorenz (1962) found the general non-transient solution of a highly simplified system and obtained a proper transport of sensible heat, but the system admitted no transport of angular momentum at all, either by the eddies or the meridional circulation. When more degrees of freedom were added to the system and an eddy transport of angular momentum was allowed, Lorenz (1963) could obtain the solution only by numerical integration.

It would therefore greatly facilitate the theoretical study of the circulation if the statistical properties of the large-scale eddies could in some manner be represented in terms of the zonally averaged motion upon which they are superposed. The procedure which most naturally suggests itself consists of assuming that the horizontal eddy transports of angular momentum and sensible heat are proportional to the gradients of zonally averaged angular velocity and temperature, the factors of proportionality being suitably chosen Austausch coefficients. Such a procedure would reduce the system of equations to one not much more complicated than the system which must be solved if the Hadley solution is to be obtained.

The observational studies have revealed, however, that throughout about half of the atmosphere the eddies transport angular momentum toward latitudes of higher angular velocity, while in some regions they transport sensible heat toward latitudes of higher temperature. Thus a procedure based upon the introduction of large-scale Austausch coefficients would yield the wrong answer. A clear statement of the theoretical reasons why the eddy-transport of angular momentum cannot be expected to conform to classical turbulence theory has been given by Eady (1954).

We feel that an analytic solution of the dynamic equations, if it is to be offered as an explanation of the circulation, cannot by-pass the problem of explaining the configurations of the eddies. Meanwhile, most theoretical studies actually carried out, except those which have sought the Hadley circulation, have had the more modest aim of demonstrating that some observed feature of the circulation must occur after some other observed feature has been assumed to occur. With this more limited goal, it is frequently possible to simplify the equations by methods which are not allowable when a complete solution is sought.

The perturbation equations

The technique of handling otherwise intractable non-linear equations which is found most frequently in meteorological literature is that of linearization, i.e. converting the non-linear equations into linear equations by treating certain variable factors as constants. The most systematic use of linearization occurs in the familiar "perturbation method". If two time-dependent solutions of the same system of non-linear equations differ only slightly from one another at some initial time, the departure of one solution from the other is governed to a close approximation by a system of homogeneous linear equations, during such time as this departure remains small. This principle finds its greatest use when one of the two solutions is already known, in which case the other solution may be found by solving the linear system. Usually the known solution is in some way simpler than the solution being sought, and most frequently it is because of its greater simplicity that it has become known. Consequently, the flow represented by the simpler solution is often called the "basic flow". In the majority of studies in meteorological literature the basic flow is independent of time, and in most of these cases it is also independent of longitude. The basic flow could, for example, be a complete Hadley circulation. There is no real necessity, however, for the known solution to be independent of longitude or time, or to be any simpler than the unknown solution.

In any event, the coefficients in the linear equations depend not only upon the original non-linear equations but also upon the particular known solution. In some instances where the basic flow is sufficiently simple, the coefficients are constants and the equations are readily solved. In other important cases the coefficients vary with latitude and elevation. If the basic flow is time-variable, the coefficients also vary with time. The perturbation method has received its most extensive meteorological development by Bjerknes *et al.* (1933).

The most important property of homogeneous linear equations is superposability; the sum of any two solutions is also a solution. It is this fact which has made it possible for the mathematical theory of these equations to become so highly developed. In many instances the general solution may be expressed as a superposition of simpler solutions or "normal modes". When the basic flow is independent of time and longitude each normal mode represents a pattern which progresses without changing its shape, while its amplitude may grow exponentially, remain constant, or decay exponentially.

The perturbation method is most easily justified when it is used to investigate the stability of a basic flow, i.e. to determine whether small disturbances superposed on the basic flow will amplify or decay. If for arbitrary initial conditions the solution consists of a superposition of normal modes, the basic flow is unstable if at least one normal mode amplifies. If all of the normal modes decay, it is stable. It is often considered neutral rather than stable if at least one mode retains its amplitude, while none amplify. An unstable flow may be stable with respect to certain classes of disturbance, and unstable with respect to others. A general solution of the linear equations reveals not only the stability or instability of the known solution but also the forms of the amplifying and decaying modes. When the general solution is not readily found, the stability may often be determined by examining the behaviour of the total energy of the eddies.

A disturbance or "eddy" superposed upon a steady zonally symmetric flow represents a supply of energy over and above that contained in the steady flow. The linearized equations do not conform to the law of conservation of energy since they frequently allow a disturbance to amplify without limit at an exponential rate. Nevertheless, the non-linear equations from which the linear equations are derived, and which they closely resemble during the period when the disturbances are small, require a source of energy for the growing disturbances; this source must be the energy of the basic flow, even though the equations do not explicitly alter the basic flow. When it is impossible to extract energy from the basic flow and at the same time satisfy the remaining constraints, the basic flow must be stable.

A study of stability as such is not concerned with the manner in which the basic flow is established or maintained, and in many studies the original non-linear equations may be simplified by omitting friction and thermal forcing altogether. When this is done, virtually any field of motion which is independent of time and longitude satisfies the non-linear equations, whether or not it bears any resemblance to a circulation which could be maintained against friction by heating. There is thus a wealth of basic flows whose stability may be investigated, and many studies seek general criteria for instability, rather than testing the stability of special basic flows. On the other hand, removing friction removes a sink of energy from the eddies — sometimes the only sink. As a result, a large class of basic flows may prove to be neutral. When friction is included the eddies must also have a source of energy, and the neutral flows are generally restricted to those critical ones for which the source and sink of energy just balance.

There are a number of mechanisms through which a basic flow may transfer its energy to a disturbance, and the unstable conditions resulting from different mechanisms are generally regarded as different types of instability. The two most important types from the point of view of the general circulation are barotropic instability and baroclinic instability.

The former type occurs when the eddies receive their energy from the kinetic energy of the basic flow. It has been treated in detail by Kuo (1949) and subsequent authors. No turning of the wind direction with height is involved, and the phenomenon is most easily studied by disregarding the vertical dimension altogether, and letting the original non-linear equations describe the flow of a two-dimensional incompressible fluid. A further common simplification is the beta-plane approximation. In giving up kinetic energy the basic flow must still conserve its total angular momentum. Since, for fixed angular momentum, a flow with uniform angular velocity (or uniform linear velocity, in the case of a beta-plane) contains the minimum kinetic energy, a flow cannot be barotropically unstable unless its angular velocity varies from one latitude to another. In this event the coefficients in the linear equations vary with latitude, and the explicit determination of normal modes is often rather difficult. Energy considerations, on the other hand, show that a necessary condition for barotropic instability is a maximum or minimum of absolute vorticity somewhere other than at the Poles (or the extreme latitudes, in the case of a beta-plane). Since, when instability does occur, the growing disturbances feed upon the kinetic energy of the basic flow, they must have a structure suitable for transporting angular momentum, in the mean, toward latitudes of lower angular velocity. They need not transport any sensible heat.

Baroclinic instability occurs when the eddies receive their energy from the available potential energy of the basic flow. There is no need for a direct exchange of kinetic energy with the basic flow; hence the basic flow need not have any horizontal shear, and the phenomenon is most easily investigated by suppressing all variations of the basic velocity with latitude. Baroclinic instability was first studied by Charney (1947), Eady (1949), and Fjørtoft (1950), and in more detail by Kuo (1952). In contrast to the barotropic stability problem, the coefficients in the linear equations become constants when the vertical wind shear is sufficiently simple, in which case the normal modes are easily determined. The various investigators have used different simplifying assumptions, and their results are not in complete agreement. Nevertheless all agree that instability is favoured by a large Coriolis parameter, large vertical wind shear, and low static stability. For vertical wind shears typical of the middle-latitude westerlies, when critical conditions are slightly exceeded, the most rapidly amplifying normal mode generally consists of a chain of six, seven, or eight waves extending around the globe. Since the growing disturbances feed upon the available potential energy of the basic flow, they must transport sensible heat, in the mean, toward latitudes of lower temperature. They need not transport any angular momentum.

The most general basic flow which is independent of time and longitude possesses both horizontal and vertical shear. Such a flow may be unstable barotropically, or baroclinically, or both barotropically and baroclinically, but separate consideration of the conditions for barotropic and baroclinic instability is not sufficient to determine whether the flow is stable or unstable. The linear equations are much more complicated than in the special cases where either horizontal or vertical shear is absent, and in general the normal modes are not readily determined analytically, although they can be found by numerical methods. The problem has recently been treated in great detail by Pedlosky (1964), using the geostrophic approximation. When the equations are further simplified by the two-level approximation, a necessary condition for instability is the existence of both positive and negative gradients of "potential vorticity", which in this case is a linear combination of vorticity and temperature. Pedlosky has considered certain flows which satisfy this condition but do not satisfy the condition for barotropic instability, and has found growing disturbances which gain their energy from the available potential energy of the basic flow by transporting sensible heat toward latitudes of lower temperature, but give up kinetic energy to the basic flow by transporting angular momentum toward latitudes of higher angular velocity. The flows therefore behave in the manner which would be expected if they possessed independently the properties of baroclinic instability and barotropic stability, and they possess an energy cycle which is qualitatively like the one observed in the atmosphere.

The linearized equations have been used for a long time to study the development of cyclones, but their application to the global circulation necessarily came only after the realization that the eddies played a significant role in the circulation. Considerable caution is required in using the equations. In the first place, the flow in the atmosphere almost never consists of a zonally symmetric flow plus small disturbances. The existing flow may be averaged with respect to longitude, and this averaged flow may be regarded as the basic flow, but the departures from this average are seldom if ever small. Use of the linearized equations when the eddies are large omits the non-linear effects of the field of eddies upon itself; these effects would be mainly noticeable as distortions. This omission is equivalent to the hypothesis that large disturbances superposed upon a basic flow behave in the same manner as small disturbances superposed upon the same flow. One might argue, however, that such an assumption is just another simplifying approximation which, like the geostrophic approximation, is not satisfied perfectly. In any event, the assumption need not violate energy principles.

A more serious shortcoming of the linearized equations is the omission of the non-linear effects of the eddies upon the basic flow. These effects would alter the energy of the basic flow, and hence the overall intensity. Because of this omission, the linearized equations say nothing about the variations of the basic flow and, as a consequence, they are incapable of explaining the ultimate intensity of the eddies, since growing or decaying normal modes will continue to grow or decay as long as the basic flow does not vary.

It is possible for a particular solution of the linearized equations to resemble the general circulation of the atmosphere, if a neutral basic flow is chosen. The energetics of the eddies may even be qualitatively correct. But the basic flow persists only because all sources and sinks have been removed.

It would be possible at this point to acknowledge the non-linear effect of the eddies upon the basic flow, and assume that heating is just sufficient to cancel this effect. But the complete system of equations would then no longer be linear, even though the equations governing the eddies would be linear. No system of linear equations can by itself yield a complete explanation of the general circulation; at most it can explain certain features after others have been assumed. For the complete problem, linearization is not "just another approximation".

A further effect not admitted by the linearized equations with a constant basic flow is the instability of a system consisting of this basic flow plus superposed eddies with respect to still further disturbances. Here the instability of a time-variable flow is involved. Such instability appears to be responsible for the occurrence of vacillation in some of the laboratory experiments and non-periodic flow in the atmosphere, in place of the uniformly travelling waves which would appear if a single normal mode were dominant.

There are at least two ways in which the theory of stability can add to our understanding of the global circulation. As early as 1937, V. Bjerknes postulated that the Hadley flow which would prevail if no disturbances were present would be unstable. Eady (1950) re-emphasized this point, and characterized the instability as baroclinic. Simplified analytic solutions for the Hadley circulation such as Charney's (1959), and numerical solutions such as Phillips' (1956), support this assumption. For the real atmosphere the exact Hadley circulation, or the nearest circulation to it which could exist in view of the geographical irregularities, has not been determined, and its stability cannot readily be investigated.

The presence of disturbances, at least in a sufficiently idealized atmosphere, is thus explained; any circulation devoid of them would soon become unstable. This does not mean that the disturbances found in the atmosphere originated as small perturbations on a nearly symmetric circulation, or for that matter that a nearly symmetric circulation ever existed. It does mean that if the disturbances should ever for any reason temporarily disappear or nearly disappear, the remaining symmetric flow would evolve toward the Hadley circulation, which is unstable, whereupon the disturbances would regenerate.

The other approach considers the stability of the existing zonally averaged flow, rather than the flow which would prevail if no disturbances were present. This flow appears to be nearly always baroclinically unstable but generally barotropically stable. Moreover, the most rapidly growing normal modes have dimensions and structure somewhat like the observed eddies in the atmosphere, and they possess a similar energy cycle. Under the hypothesis that the normal modes which are indicated by the linear equations as amplifying most rapidly are the ones which ultimately acquire and retain a large amplitude, the typical size and shape of the eddies, and the observed energy cycle, are qualitatively accounted for in terms of the observed basic flow.

Real disturbances do not amplify forever, and it might appear that the typical basic flow should be neutral rather than unstable, in order that the disturbances should simply maintain their intensity. However, there are limitations to the assumption that large disturbances behave like small ones. Individual cyclones tend to have a life cycle. By the time that they reach occlusion their shape has changed considerably. The change of shape and the cessation of growth are not predicted by the linearized equations. They may result in part from a change in the basic flow, in which case they could be described by linear equations with prespecified time-variable coefficients. It is likely however that the occlusion of a cyclone is partly due to the non-linear distortive effects of the field of disturbances upon itself, in which case it can occur when the basic flow is still unstable rather than neutral.

The equations for the basic flow

What is mainly lacking in theoretical work based upon linearized equations is an explanation of the observed basic flow and its variations. Growing disturbances, in removing energy from a basic flow, ordinarily render it less unstable; ultimately they render it neutral and the disturbances cease growing, or, as just noted, they may cease growing while the basic flow is still unstable. Thus the disturbances act as a governor, maintaining the basic flow at nearly neutral stability. There is, however, a wide variety of neutral basic flows. The appropriate flow is not simply the Hadley flow, reduced by a constant factor. Stability considerations therefore place a constraint upon the basic flow, but do not determine it.

It is nevertheless possible in principle to determine the basic flow if the properties of the disturbances are known, not by means of the linearized equations, but by the zonally-averaged equations which govern the basic flow. These equations are identical with the ones which govern the Hadley flow, except that the convergences of the eddy-transports of angular momentum and sensible heat appear as additional mechanical and thermal forcing. If these transports are prespecified, the procedure for solving the equations is similar to the procedure for solving for the Hadley circulation, and it should be approximately as difficult.

A solution of these equations would not by itself explain the basic flow, since the prespecified transports actually depend upon the eddies, which in turn are influenced by the basic flow. Consequently the determination of a solution would be a somewhat unrewarding accomplishment. Since it would also be a difficult task, it is not surprising that it has not been carried out. Nevertheless, in combination with a suitable solution of the linearized equations, a solution of the equations governing the basic flow might offer the best attainable approximation to an analytic solution of the complete system.

We have noted that under the hypothesis that the normal modes which grow most rapidly when small in amplitude are the ones which will remain after reaching finite amplitude, the linearized equations may be solved for eddies of an unknown amplitude but a known shape, in terms of a prespecified basic flow. Likewise, we have seen that the zonally averaged equations may in principle be solved for the basic flow, provided that the transports of angular momentum and sensible heat by the eddies are prespecified. The combined system of equations should therefore be solvable for the zonal flow and the eddies simultaneously, provided simply that the amplitude of the eddies is prespecified. The appropriate amplitude may be determined by the condition that the basic flow should be neutral, so that the eddies will undergo no net gain or loss of energy.

In reality the new system of equations does not differ too greatly from the original system of governing equations upon which it is based, and it may be regarded as "another approximation". It has been simplified to the extent of omitting the non-linear effects of the field of eddies upon itself, but it is still a closed non-linear system. The linearized equations governing the eddies still appear among the equations, but they are now no longer linear, since the coefficients, which depend upon the basic flow, are now unknowns of the system. From the point of view of duplicating the circulation, there is nothing to recommend a solution of this system over a numerical solution of the original system from which it was derived. However, the process of solving the system may yield some insight into the general circulation which is not afforded by a typical numerical experiment. The maintenance of the eddies by a basic flow which is baroclinically unstable or neutral but barotropically stable, and the control of the basic flow by the transports of angular momentum and energy accomplished by the eddies, should be clearly revealed. An added feature is that the restriction to a single normal mode will automatically yield a steady-wave solution.

An approximate method of solving a system of equations of this sort was used by Charney (1959) to construct a model of the general circulation. In this study Charney began with a simplified system of equations essentially the same as the one used by Phillips (1956) in his original numerical experiment. This system includes the beta-plane, two-layer, and geostrophic approximations. He first solved analytically for the Hadley circulation — a task made possible by the simplicity of the equations. He found this circulation to be unstable, and he determined the most rapidly amplifying normal mode. He then postulated that the growing disturbance would retain its shape while it modified the zonal flow, and he solved for a new zonal flow in terms of an unknown amplitude of the eddies, including as additional thermal and mechanical forcing the effects of the eddy-transports of angular momentum and energy. Finally he determined the amplitude of the disturbance by requiring that the available potential energy given to the disturbances by the zonal flow and by external heating should balance the kinetic energy removed from the disturbances by the zonal flow and by friction. A balancing energy cycle was thus assured.

The picture of the general circulation obtained by Charney is realistic in certain features. Easterly surface winds appear in low and high latitudes, with westerlies in between. The energy cycle proceeds in the right direction. More significant from the point of view of method is the close resemblance to the results of Phillips, since Charney was attempting to solve by an approximate method nearly the same equations which Phillips had solved numerically.

It would seem possible to obtain an exact solution of the new system by a successive-approximation scheme in which the first few steps duplicate the procedure used by Charney. Following these steps, a second approximation to the field of eddies, again in terms of an unknown amplitude, is obtained by finding the most rapidly growing mode corresponding to the new zonal flow. The transports of angular momentum and energy by the new eddies are used to solve for the next approximation to the basic flow, and again the amplitude of the eddies is determined by requiring the energy cycle to balance. This scheme is then repeated until convergence is obtained. There is of course no assurance that such a scheme will converge at all, but in view of the general resemblance of Charney's approximation to Phillips's solution, which is indicative of the final approximation, it seems possible that the method will converge rather rapidly.

It remains to be seen whether some further extension of this method, capable of representing vacillation or perhaps some more irregular behaviour, can yield a more realistic representation of the general circulation while continuing to offer as much insight into the manner in which the circulation operates.

CHAPTER IX

THE REMAINING PROBLEMS

The large numerical experiments, in which the state of the atmosphere is sometimes represented by many thousands of numbers, constitute our closest approach to a theoretical demonstration that the circulation must assume the form which it does rather than some other conceivable form. The demonstration is by no means complete. Certain quantities which really depend upon the circulation, such as the spatial distribution of absorption of radiation by water vapour, have usually had their observed values preassigned to them. Certain features, such as the presence of tropical hurricanes, have not yet been reproduced. Moreover, since even the most realistic numerical solutions are particular solutions, the possibility of other particular solutions with rather different properties cannot be completely eliminated. Nevertheless, shortcomings of these sorts have characterized virtually all theoretical studies of the general circulation, usually to a much greater extent. One gets the impression that the experiments will ultimately duplicate the atmospheric circulation to any desired degree of accuracy.

As demonstrations that the atmosphere must behave as it does, theoretical studies employing more classical mathematical procedures have not as yet compared with the numerical experiments. One of the closer approximations to such a demonstration is the previously cited study by Charney (1959). Yet Charney found it necessary to introduce the *ad hoc* assumption that the fully developed disturbances, superposed upon the prevailing zonal flow, would possess the same shape as the most rapidly growing incipient disturbances superposed upon the Hadley flow. Altogether he succeeded in duplicating no more than the grossest features of the general circulation. Nevertheless, by following through his complete procedure one can gain a certain insight concerning the reasons why the atmosphere behaves as it does, which one might not gain from an inspection of the millions of tabulated numbers forming the complete output of a numerical experiment.

A deeper physical insight may sometimes be afforded by a simple qualitative description. In this concluding chapter we shall outline as nearly complete a qualitative explanation for some of the main features of the circulation as we feel can at present be formulated. We shall not attempt to make our presentation rigorous, and our arguments will not always be demonstrations that the atmosphere must behave as it does, rather than in some other manner. We shall however attempt to present the correct reasons for the observed behaviour, to the extent that these are known. In addition we shall indicate the areas where suitable qualitative explanations have yet to be offered.

An explanation of the circulation of the atmosphere logically begins with the driving force. There seems to be no question but that this is solar radiation, and that it is the greater intensity of this radiation in lower than in higher latitudes which enables it to produce available potential energy, which it may do either by heating the atmosphere directly, or by heating the underlying surface which in turn transmits energy to the atmosphere.

It follows that the atmosphere must possess a circulation. For in the absence of motion, each latitude would assume a state of thermal equilibrium, losing as much heat as it gained, and in order to lose more heat than the higher latitudes, the lower latitudes would have to be warmer. But a cross-latitude temperature contrast would be incompatible with a state of no motion, for if hydrostatic equilibrium prevailed, there would be cross-latitude pressure gradients, and horizontal motions would develop, while if hydrostatic equilibrium did not prevail, vertical motions would develop.

The required circulation must transport energy from low to high latitudes, and thereby bring about a weaker poleward temperature gradient than would otherwise prevail, but it cannot destroy the gradient altogether. This follows because friction is continually dissipating the kinetic energy of the circulation, and new kinetic energy must be produced at the expense of available potential energy, whence new available potential energy must be generated by heating of the warmer regions and cooling of the colder ones. If every latitude remained in thermal equilibrium in spite of the circulation, there would be no local heating or cooling, while if the circulation destroyed the temperature contrast altogether, there would be no warmer and colder regions to be heated and cooled. If the circulation transported energy equatorward and thereby maintained a stronger temperature contrast than would otherwise prevail, the tropical regions would be cooled and the polar regions would be heated by radiation, and available potential energy would be destroyed. Likewise, if the circulation transported enough energy poleward to reverse the temperature gradient, the now warmer polar regions would be cooled and the now cooler tropical regions would be heated by radiation, and available potential energy would again be destroyed.

The arguments which we have presented are far from rigorous. First of all, available potential energy may be generated without any horizontal heating contrast if the atmosphere is statically unstable. Without going into detail, we shall simply note that radiation evidently tends to produce a stable stratification throughout much of the atmosphere; where this is not the case, small-scale convective motions tend to develop and stabilize the stratification. Even assuming a statically stable atmosphere, however, we have not considered the vertical structure of the atmosphere in sufficient detail. In the lower stratosphere, for example, the equatorial latitudes are the coldest, and heating destroys available potential energy. Our general conclusions can apply only to some sort of vertical average, but neither the outgoing radiation nor the generation of available potential energy depends simply upon vertically averaged temperatures. Finally, we have omitted the possibility that the necessary generation of available potential energy could result from cross-longitude temperature and heating contrasts, which might perhaps arise from the cross-longitude contrast between land and ocean. Nevertheless, we feel that our explanations, while not rigorous, are essentially correct.

Accompanying the poleward temperature gradient there must be a poleward pressure gradient at higher levels or an equatorward pressure gradient at lower levels, or both. Absence of these gradients would require too small a vertical pressure gradient at high latitudes or too large a vertical pressure gradient at low latitudes to be in hydrostatic equilibrium. There would then have to be net downward acceleration in high latitudes or net upward acceleration in low latitudes. The full explanation for hydrostatic equilibrium is rather complex, involving an explanation of the typical scale of motion, but at this point we are concerned only with the equilibrium of the average state over an extensive region of the atmosphere. We must nevertheless turn to quantitative considerations. A net upward acceleration over an extensive region in low latitudes requires that the air leaving the region be moving upward more rapidly than the air entering the region. In the present instance, the velocities which would be required at the boundary of the region in the absence of the stipulated pressure gradients would be greater than any which could be maintained against friction, even if solar heating produced available potential energy with the maximum possible efficiency.

At the Earth's surface as a whole neither easterly nor westerly winds can predominate. This follows because there can be no net long-term transfer of angular momentum from the Earth to the atmosphere, and hence no net surface torque. Either easterly winds must prevail at some latitudes and westerlies at others, or else there can be no systematic distribution of surface easterlies and westerlies at all.

It is natural to conclude that general westerly winds must predominate aloft. Absence of these winds would require too large a poleward pressure gradient aloft, or too large an equatorward pressure gradient at the surface, or both, to be in geostrophic equilibrium. Hence there would have to be net poleward acceleration at high levels or net equatorward acceleration at low levels. Like the explanation for hydrostatic equilibrium, the full explanation for geostrophic equilibrium is rather complex, but again we are at present concerned only with the equilibrium of an average state over an extensive region. However, in this case the possibility of non-geostrophic flow cannot be eliminated by energy considerations alone.

A highly non-geostrophic circulation, in which the pressure decreases equatorward at low levels and poleward at high levels, while there are no systematic easterly or westerly winds at any level, is apparently possible if there is a downward transport of northward momentum across all levels, attaining its maximum value in middle levels. The downward transport could be accomplished by "mesoscale" systems having horizontal dimensions of perhaps a hundred kilometres, in which the poleward-moving air sinks and the equatorward-moving air rises. We shall presently consider geostrophic equilibrium in greater detail; in the meantime we shall merely assume that no mechanism exists for maintaining the mesoscale systems which would bring about the needed downward transport of northward momentum. In that event quasi-geostrophic westerly winds must prevail at upper levels.

The preceding are the features of the circulation which are most readily deduced from basic principles, even if not in a completely rigorous fashion. They include a poleward transport of energy, which is needed to hold the poleward temperature gradient below its thermal-equilibrium value, and thereby enable the heating to produce available potential energy. No conclusions have been drawn concerning the poleward transport of absolute angular momentum; therefore no explanation for the latitudes occupied by the surface easterlies and westerlies has been offered.

One circulation scheme compatible with the properties so far deduced is the zonally symmetric circulation of Hadley, possibly with the modifications introduced by Thomson and Ferrel. In such a circulation there must be general poleward flow at upper levels and equatorward flow at lower levels in order to bring about the necessary poleward transport of energy, since sensible heat plus potential energy increases with elevation in a stably stratified atmosphere. The mesoscale eddies which in the more general case might produce a downward transport of northward momentum are certainly absent, so that upper-level westerlies must be present. Thus the direct meridional cell also brings about a poleward transport of angular momentum, and the surface winds are easterly in low latitudes and westerly in higher latitudes.

Although the Hadley circulation is consistent with the governing physical laws, it does not occur, because it is baroclinically unstable. Baroclinic instability is favoured by a large Coriolis parameter, large vertical wind shear with its accompanying large horizontal temperature gradient, and low vertical stability. In the Hadley circulation these conditions are met in middle and high latitudes. Possibly they are not met in the tropics, where the Coriolis parameter and the poleward temperature gradient are smaller, but the separate latitudes do not act independently of one another, and the Hadley circulation as an entity is unstable. A circulation containing longitude-dependent eddies therefore occurs in its stead. Quantitative considerations are required to determine whether any particular Hadley circulation is baroclinically unstable. That the Hadley circulation belonging to the Earth's atmosphere should meet the conditions for instability must be considered accidental. It is conceivable that the atmosphere of a more slowly rotating planet could possess a stable Hadley circulation.

The eddies must lose kinetic energy through frictional dissipation. Under the assumption that the warmer portions of the eddies will radiate more heat to space than the colder portions, the eddies must lose available potential energy by heating. They must therefore gain either available potential energy or kinetic energy from the zonally averaged flow. In the former case they must bring about a cross-latitude transport of sensible heat, and in the latter case a cross-latitude transport of angular momentum. In either event the zonally averaged flow must then differ from the Hadley circulation which would prevail if the eddies were absent.

The principal remaining properties of the circulation, including the occurrence of hydrostatic and geostrophic equilibrium, and the distribution of the surface easterly and westerly winds, depend upon the scale and structure of the eddies. We consider first the qualitative properties of quasi-hydrostatic equilibrium, quasi-horizontal motion, quasi-geostrophic equilibrium, and quasi-non-divergent motion. Small scales of motion are of course not geostrophic, and the very smallest scales are not even hydrostatic. We are interested in explaining why most of the energy of the atmosphere is in the hydrostatic and geostrophic modes.

It is sometimes stated that the flow must be quasi-horizontal and quasi-hydrostatic because the effective vertical depth of the atmosphere is extremely small compared to its horizontal extent. This argument does not appear sound. There is nothing physically impossible about a thin layer of fluid in which the principal motions or perhaps the only ones are vertically propagating sound waves, which are decidedly not hydrostatic. Likewise, even under quasi-hydrostatic conditions, the motions in such a layer could be confined to gravity waves, which are non-geostrophic. The relative unimportance of such motions in the atmosphere is due to the absence of the processes which would be needed to produce and maintain them.

The adjustment of an initially unbalanced flow toward geostrophic equilibrium has been studied by Rossby (1938b) and in greater detail by Obukhov (1949). However, the adjustment considered by these authors is local, occurring at the expense of geostrophic equilibrium elsewhere. The circulation as a whole becomes neither more nor less geostrophic during the process.

The problem may be clarified through a consideration of scale theory, in the manner suggested by Charney (1948). Assuming that the motions have horizontal velocities, horizontal and vertical dimensions, and time scales typical of the principal motions in the atmosphere, and applying order-of-magnitude considerations to the various terms in the governing equations, Charney finds that the motions must be nearly hydrostatic, horizontal, geostrophic, and non-divergent. The problem of explaining these properties thereby becomes equivalent to the problem of explaining the typical observed scales of atmospheric motions. It is therefore only slightly less involved than the whole problem of the general circulation, which requires an explanation of the shapes of the systems of motion as well as their dimensions and amplitudes.

It is thus obvious that hydrostatic and geostrophic equilibrium cannot be explained, any more than can the whole circulation, without considering the thermal forcing. The principal component of the forcing, the Equator-to-Pole gradient of heating, is of large scale and infinite period, and the motion which it directly forces, namely the Hadley circulation, is likewise of large scale and infinite period. We have already noted that the Hadley circulation is nearly hydrostatic and geostrophic. It is sometimes considered non-geostrophic because the transports of energy and momentum are accomplished entirely by the small non-geostrophic meridional motions. Nevertheless the zonal flow, which contains most of the kinetic energy, is approximately geostrophic.

In the real atmosphere the seasonal variations of heating provide another large-scale long-period component of the forcing. The diurnal variations provide a further component which has a large scale but a short period. The latter component forces the well-known thermal atmospheric tides which are decidedly non-geostrophic, but which contain only a minor amount of the total energy, and which in general do not appear to interact too strongly with the remaining motions.

Somewhat smaller-scale motions are directly forced in the real atmosphere as a result of the heating contrast between continents and oceans, or between smaller geographical features, but most of the smaller-scale motions, either in the real or the idealized atmosphere, result from the non-linear interactions of larger-scale motions. Among the smaller-scale motions are those resulting from instability.

Instability, regardless of how readily it may be investigated by linearized equations, is a non-linear process. It is a special case of the non-linear interaction of two superposed fields of motion to produce a third, when the amplitude of the first of the interacting fields is much larger than that of the second. Because of the disparity in amplitudes, the amplitude of the first field remains nearly constant, and may be treated as a constant in the equations governing the second and third fields.

Typical studies of the baroclinic instability of zonally symmetric motion similar to that occurring in the atmosphere indicate that the most rapidly growing disturbances have the proper space and time scales to be nearly hydrostatic and geostrophic, according to scale theory. If we postulate that fully developed eddies have the same scales as the most rapidly growing small-amplitude eddies superposed upon the same basic flow, we obtain a fairly acceptable explanation for hydrostatic and geostrophic equilibrium. Here a word of warning is needed. The studies indicating that hydrostatic and geostrophic modes of motion amplify most rapidly are based for the most part upon equations which contain the hydrostatic and geostrophic approximations, and which are therefore incapable of revealing the possible growth of non-hydrostatic or non-geostrophic modes. At the present time it can only be assumed that further non-linear processes, whether interactions of two modes of eddy motion which are each demanded by the instability of the zonally averaged flow, or instabilities of a zonal flow plus a single mode with respect to still further modes, will lead only to new motions also having the proper space and time scales, or else to motions of such small amplitude that they are not important components of the total circulation.

The remaining properties of the circulation depend upon the shapes of the eddies as well as their sizes. Thus a poleward transport of angular momentum is generally produced by a pattern of troughs and ridges which are displaced eastward with increasing latitude, while a poleward transport of sensible heat is produced by troughs and ridges which are displaced westward with increasing elevation. We regard the problem of explaining the pattern of the transport of angular momentum by the eddies as the most important problem in general-circulation theory among those for which we now lack a fairly adequate qualitative explanation. The convergence of the angular-momentum transport exerts a controlling influence upon the latitudes chosen by the surface easterlies and westerlies. It also tends to disrupt the geostrophic equilibrium aloft, and thereby leads to the formation of direct meridional cells in low latitudes and indirect cells in middle latitudes, which tend to restore the equilibrium. Thus the cells in the tropics are much stronger than they would be in the Hadley circulation. These direct cells transport large amounts of water equatorward, which then produce the heavy precipitation in the tropics. They also transport sensible heat plus potential energy away from the Equator.

The earliest attempts to explain the eddy transport of angular momentum were based upon an analogy with classical turbulence theory. The eddies were assumed to transport angular momentum toward latitudes of lower angular velocity, so that the stronger westerlies would effectively drag the weaker westerlies ahead. However, there is no physical basis for applying classical turbulence theory to eddies of cyclone scale; indeed the assumption that all eddies transport angular momentum in accordance with diffusion concepts is equivalent to the assumption that all zonally symmetric flows other than solid rotation are barotropically unstable. Moreover, the theory would yield incorrect results, since throughout most of the tropics and subtropics angular momentum is transported toward latitudes of stronger westerlies.

Studies of baroclinic stability of flow on the beta-plane, such as the one by Pedlosky (1964), yield an angular-momentum transport of the proper sign at low latitudes, but they also indicate a countergradient transport on the poleward side of the westerly-wind maximum, where it is not actually observed. Possibly the latter result is only a shortcoming of the beta-plane, since the numerical experiments performed on the beta-plane, such as that of Phillips (1956), yield similar results. However, studies based upon perturbation theory cannot apply to finite-amplitude disturbances, unless further hypotheses are introduced.

The postulate that finite-amplitude eddies superposed upon a zonally averaged flow should have the same shape as the most rapidly amplifying incipient disturbances superposed upon the same flow would lead to simpler and more regularly shaped eddies than those actually found. Classical turbulence theory, on the other hand, would lead to more irregular eddies than those observed. In either event the eddies would be assumed to possess some equilibrium form determined by the zonally averaged flow.

We feel that there are good reasons for believing that the properties of the eddies cannot be represented in terms of the current zonal flow. Let us assume that there does exist some equilibrium configuration which the eddies would ultimately attain if the zonal flow did not vary. The time required for eddies of cyclone scale to reach approximate equilibrium might be one or two days. But during this time the zonal flow will vary, largely as a result of the transport of angular momentum and energy by the eddies. The new zonal flow will demand a new equilibrium configuration, and in approaching this new configuration the eddies will further alter the zonal flow, etc., and no equilibrium will ever be reached. We may note that these considerations do not preclude the possibility that the effects of small-scale eddies can be fairly well represented in terms of the larger-scale flow, since small-scale eddies may attain an equilibrium configuration in the course of an hour or less, while the larger-scale flow should remain reasonably constant for considerably longer.

Despite these observations, the postulate that finite-amplitude eddies possess the same general shape as rapidly amplifying infinitesimal eddies leads to a number of correct conclusions. Sensible heat is transported toward colder latitudes, so that the eddies gain available potential energy from the zonal flow. Angular momentum is transported predominantly toward latitudes of higher angular velocity, so that the eddies give up kinetic energy to the zonal flow. We should therefore note that any explanation based upon the theory of baroclinic stability has departed considerably from the type of qualitative explanation which we presented earlier in this chapter. The mathematical work required to find the form of the most rapidly amplifying disturbance, when the zonal flow varies both horizontally and vertically, is extremely involved. The investigator who has solved the problem may still gain little physical insight as to why the eddies prove to have one particular shape rather than another.

If the eddies are to transport angular momentum into the latitudes of maximum westerlies, as they apparently do on the beta-plane, the trough and ridge lines must assume somewhat the same shape as the westerly wind profile itself, with their maximum eastward positions coinciding with the maximum westerlies. The troughs and ridges therefore act somewhat in the manner of elastic bands, which are stretched out by the zonal flow, but are prevented by the elastic restoring force from being pulled too greatly out of shape. Yet we can offer no simple explanation as to why the troughs and ridges should behave in this manner. Their typical shape on the sphere presents an equally perplexing problem.

Simple qualitative arguments could perhaps be offered to explain some of the principal remaining features, such as the equatorward temperature gradient in the stratosphere and the reversed stratospheric energy cycle. By this point, however, our explanation of the general circulation has become so incomplete that we shall not attempt to extend it any further.

In closing this monograph, we must express an intuitive conviction that a complete qualitative explanation of the principal features of the general circulation will eventually be found. It seems possible, for example, that there should be some system of equations and ordered inequalities, with as many unknowns as relations, which will yield the rigorous result that the poleward eddy-transport of angular momentum across middle latitudes is greater than zero. It seems further possible that this rigorous qualitative result might then be converted into a comprehensible qualitative argument. To our knowledge, however, the desired system of relations has yet to be found.

REFERENCES

The numbers in parentheses following each reference indicate the pages in this work where the particular reference is cited. When a reference is cited more than once in the same subsection of a chapter, only one page, usually the first one, is listed.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Computational Phys.*, **1**, 119–143. (132)

Benton, G. S. and Estoque, M. A., 1954: Water-vapor transfer over the North American continent. *J. Meteor.*, **11**, 462–477. (80)

Bergeron, T., 1928: Über die dreidimensional verknüpfende Wetteranalyse. *Geofys. Publ.*, Vol. **5**, No. 6, 111 pp. (59, 75)

Bigelow, F. H., 1900: *Report on the international cloud observations.* Report of the Chief of the Weather Bureau, 1898–99, Vol. II. Washington, 787 pp. (71)

Bigelow, F. H., 1902: Studies of the statics and kinematics of the atmosphere in the United States. *Mon. Weather Rev.*, **30**, 13–19, 80–87, 117–125, 163–171, 250–258, 304–311, 347–354. (71)

Bjerknes, J., 1919: On the structure of moving cyclones. *Geofys. Publ.*, Vol. **1**, No. 2, 8 pp. (74)

Bjerknes, J., 1948: Practical application of H. Jeffreys' theory of the general circulation. Réunion d'Oslo, Assoc. de Météor., UGGI. Programme et Résumé des Mémoires, 13–14. (78)

Bjerknes, V., 1921: On the dynamics of the circular vortex with applications to the atmosphere and atmospheric vortex and wave motions. *Geofys. Publ.*, Vol. **2**, No. 4, 88 pp. (75, 124)

Bjerknes, V., 1937: Application of line integral theorems to the hydrodynamics of terrestrial and cosmic vortices. *Astrophys. Norv.*, Vol. **2**, No. 6, 263–339. (75, 142)

Bjerknes, V., Bjerknes, J., Bergeron, T. and Solberg, H., 1933: *Physikalische Hydrodynamik.* Berlin, J. Springer, 797 pp. Also (1934) *Hydrodynamique Physique.* Paris, Presses Univ. de France, 864 pp. (74, 75, 139)

Brillouin, M., 1900: *Mémoires originaux sur la circulation générale de l'atmosphère.* Paris, Carré et Naud, 163 pp. (68)

Brunt, D., 1926: Energy in the Earth's atmosphere. *Phil. Mag.*, **7**, 523–532. (100)

Buch, H., 1954: *Hemispheric wind conditions during the year 1950.* Final Report, Part 2, Contract AF 19(122)–153, Dept. of Meteorology, Mass. Inst. of Technology. (32, 80)

Buchan, A., 1889: Report on atmospheric circulation. *Report on the scientific results of the exploring voyage of the H.M.S. Challenger, 1873–76.* Physics and Chemistry, Vol. 2. H.M. Stationery Off. (25)

Budyko, M. I., 1956: *Teplovoi balans zemnoi poverkhnosti.* Gidrometeorologicheskoe Izdatel'stvo, Leningrad, 255 pp. English trans. : Stepanova, N. A., 1958 : *The heat balance of the Earth's surface.* Washington, U.S. Dept. of Commerce, Off. Tech. Serv., 259 pp. (49)

Budyko, M. I., 1963: *Atlas teplovogo balansa zemnogo shara (Atlas of the heat balance of the Earth).* Glavnaia Geofys. Observ., Moscow, 69 pp. (49)

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–162. (131, 140)

Charney, J. G., 1948: On the scale of atmospheric motions. *Geofys. Publ.*, Vol. **17**, No. 2, 17 pp. (148)

Charney, J. G., 1959: On the theory of the general circulation of the atmosphere. *The atmosphere and the sea in motion*, B. Bolin, Ed. New York, Rockefeller Inst. Press, 178–193. (142, 143, 145)

Charney, J. G., Fjørtoft, R. and von Neumann, J., 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254. (131)

Craig, R. A., 1945: A solution of the nonlinear vorticity equation for atmospheric motion. *J. Meteor.*, **2**, 175–178. (138)

Crutcher, H. L., 1959: *Upper wind statistics charts of the northern hemisphere.* Off. of Chief of Naval Operations, Washington. (33, 88, 112)

Crutcher, H. L., 1961: *Meridional cross-sections. Upper winds over the northern hemisphere.* Tech. Pap. No. 41, U.S. Weather Bureau, Washington. (33)

Dalton, J., 1793: *Meteorological observations and essays.* Manchester, Harrison and Crosfield. (60)

Davies, T. V., 1959: On the forced motion due to heating of a deep rotating liquid in an annulus. *J. Fluid Mech.*, **5**, 593–621. (125)

Davis, P. A., 1963: An analysis of the atmospheric heat budget. *J. Atmos. Sci.*, **20**, 5–22. (93)

Defant, A., 1921: Die Zirkulation der Atmosphäre in den gemässigten Breiten der Erde. *Geograf. Ann.*, **3**, 209–266. (72)

Douglas, C. K. M., 1931: A problem of the general circulation. *Q. J. Roy. Meteor. Soc.*, **57**, 423–431. (73)

Dove, H. W., 1835: Über den Einfluss der Drehung der Erde auf die Strömungen ihrer Atmosphäre. *Poggendorff's Ann. Phys. und Chem.*, **36**, 321–351. (61)

Dove, H. W., 1837: *Meteorologische Untersuchungen.* Berlin, Sandersche Buchhandlung, 344 pp. (61)

Dutton, J. A. and Johnson, D. R., 1967: The theory of available potential energy and a variational approach to atmospheric energetics. *Advances in Geophysics*, **12** (in press). (107, 111)

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, Vol. **1**, No. 3, 35–52. (140)

Eady, E. T., 1950: The cause of the general circulation of the atmosphere. *Centenary Proc. Roy. Meteor. Soc.*, 156–172. (141)

Eady, E. T., 1954: The maintenance of the mean zonal surface currents. *Proc. Toronto Meteor. Conf. 1953.* London, Roy. Meteor. Soc., 124–128. (138)

Eliassen, A., 1952: Slow thermally or frictionally controlled meridional circulation in a circular vortex. *Astrophys. Norv.*, Vol. **5**, No. 2, 19–60. (93)

Exner, F. M., 1925: *Dynamische Meteorologie.* Vienna, J. Springer, 421 pp. (74)

Faller, A. J., 1956: A demonstration of fronts and frontal waves in atmospheric models. *J. Meteor.*, **13**, 1–4. (117)

Ferrel, W., 1856: An essay on the winds and the currents of the ocean. *Nashville J. Medicine and Surgery*, 11, 287–301. Reprinted (1882) in *Popular essays on the movements of the atmosphere.* Prof. Pap. Signal Serv. No. 12, Washington, 7–19. (64)

Ferrel, W., 1859: The motions of fluids and solids relative to the Earth's surface. *Math. Monthly*, **1**, 140–147, 210–216, 300–307, 366–372, 397–406. (66, 136)

Ferrel, W., 1889: *A popular treatise on the winds.* New York, Wiley, 505 pp. (67)

Fjørtoft, R., 1951: Stability properties of large-scale disturbances. *Compendium of meteorology*, T. F. Malone, Ed. Boston, Amer. Meteor. Soc., 454–463. (140)

Fultz, D., 1951: Experimental analogies to atmospheric motions. *Compendium of meteorology*, T. F. Malone, Ed. Boston, Amer. Meteor. Soc., 1235–1248. (115)

Fultz, D., 1952: On the possibility of experimental models of the polar-front wave. *J. Meteor.*, **9**, 379–384. (117)

Fultz, D., Kaiser, J., Fain, M., Kaylor, R. E. and Weil, J., 1964: *Experimental investigations of the spectrum of thermal convective motions in a rotating annulus.* Article 2 B, Final Report, Contract AF 19(604)–8361, Dept. of Geophysical Sciences, Univ. of Chicago. (120)

Fultz, D., Long, R. R., Owens, G. V., Bohan, W., Kaylor, R. and Weil, J., 1959: *Studies of thermal convection in a rotating cylinder with some implications for large-scale atmospheric motions.* Meteor. Monographs, Amer.Meteor. Soc., 104 pp. (115, 120, 124)

Gavrilin, B. L., 1965: Numerical experiments on the general circulation of the atmosphere. *Izvestia, Atmos. and Ocean Phys.*, **1**, 1229–1259. (134)

Gilman, P. A., 1965: The mean meridional circulation of the southern hemisphere inferred from momentum and mass balance. *Tellus*, **17**, 277–284. (95)

Goldie, N., Moore, J. G. and Austin, E. E., 1958: *Upper-air temperature over the world.* Geophys. Mem. No. 101, Meteor. Off. London, H. M. Stationery Off., 228 pp. (39)

Hadley, G., 1735: Concerning the cause of the general trade-winds. *Phil. Trans.*, **29**, 58–62. (1, 60, 117, 136)

Halley, E., 1686: An historical account of the trade-winds and monsoons observable in the seas between and near the tropicks with an attempt to assign the physical cause of said winds. *Phil. Trans.*, **26**, 153–168. (1, 97, 126)

Hann, J., 1901: *Lehrbuch der Meteorologie.* Leipzig, Chr. Herm. Tauchnitz, 805 pp. (69)

Haurwitz, B., 1940: The motion of atmospheric disturbances on the spherical Earth. *J. Marine Res.*, **3**, 254–267. (138)

Heastie, H. and Stephenson, P. M., 1960: *Upper winds over the world. Parts I and II.* Geophys. Mem. No. 103, Meteor. Off. London, H. M. Stationery Off., 217 pp. (33)

Helmholtz, H. v., 1868: Über discontinuirliche Flüssigkeitsbewegungen. *Monatsber. Kön. Akad. Wiss.*, Berlin, 215–228. English trans.: Abbe, C., 1893 : *The mechanics of the Earth's atmosphere.* Washington, Smithsonian Inst., 58–66. (70)

Helmholtz, H. v., 1888: Über atmosphärische Bewegungen. *Sitz.-Ber. Akad. Wiss. Berlin*, 647–663. English trans.: Abbe, C., 1893: *The mechanics of the Earth's atmosphere.* Washington, Smithsonian Inst., 78–93. (70)

Helmholtz, H. v., 1889: Über atmosphärische Bewegungen, II. *Sitz.-Ber. Akad. Wiss. Berlin*, 761–780. English trans.: Abbe, C., 1893 : *The mechanics of the Earth's atmosphere.* Washington, Smithsonian Inst., 94–111. (70)

Hide, R., 1953: Some experiments on thermal convection in a rotating liquid. *Q. J. Roy. Meteor. Soc.*, **79**, 161. (119)

Hildebrandsson, H. H., and Teisserenc de Bort, L., 1900. *Les bases de la météorologie dynamique.* Vol. 2. Paris, Gauthier-Villars, 345 pp. (71)

Holopainen, E. O., 1963: On the dissipation of kinetic energy in the atmosphere. *Tellus*, **15**, 26–32. (101)

Holopainen, E. O., 1965: On the role of mean meridional circulations in the energy balance of the atmosphere. *Tellus*, **17**, 285–294. (90)

Holopainen, E. O., 1966: *Some dynamic applications of upper-wind statistics.* Finnish Meteor. Off. Contrib., No. 62, 22 pp. (38, 52, 88, 94, 112)

Hutchings, J. W. and Thompson, W. J., 1962: The torque exerted on the atmosphere by the Southern Alps. *New Zealand J. Geol. Geophys.*, **5**, 18–28. (52)

Jeffreys, H., 1926: On the dynamics of geostrophic winds. *Q. J. Roy. Meteor. Soc.*, **52**, 85–104. (73, 78)

Kant, I., 1756 (?): *Anmerkungen zur Erläuterung der Theorie der Winde.* Reference quoted from Hann (1901), p. 466. (60)

Kibel, I. A., 1940: Prilozhenie k meteorologii uravnenii mekhaniki baroklinnoi zhidkosti. (Meteorological applications of the equations of mechanics to a baroclinic fluid.) *Izv. Akad. Nauk SSSR*, Ser. Geogr. i Geofiz., No. 5, 627–637. (118)

Kropatscheck, F., 1935: Die Mechanik der grossen Zirkulation der Atmosphäre. *Beitr. Phys. Fr. Atmos.*, **22**, 272–298 (136)

Kung, E. C., 1966: Large-scale balance of kinetic energy in the atmosphere. *Mon. Weather Rev.*, **94**, 627–640. (101)

Kuo, H.-L., 1949: Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteor.*, **6**, 105–122. (140)

Kuo, H.-L., 1951: A note on the kinetic energy balance of the zonal wind systems. *Tellus*, **3**, 205–207. (109)

Kuo, H.-L., 1952: Three-dimensional disturbances in a baroclinic zonal current. *J. Meteor.*, **9**, 260–278. (140)

Kuo, H.-L., 1956: Forced and free meridional circulations in the atmosphere. *J. Meteor.*, **13**, 561–568. (93)

Kuo, H.-L., 1957: Further studies of thermally driven motions in a rotating fluid. *J. Meteor.*, **14**, 553–558. (125)

Kuo, H.-L., 1959: Finite-amplitude three-dimensional harmonic waves on the spherical Earth. *J. Meteor.*, **16**, 524–534. (138)

Leith, C. E., 1965: Numerical simulation of the Earth's atmosphere. *Methods in computational physics.* Vol. 4. New York, Academic Press, 1–28. (132, 133)

Lettau, H., 1954: A study of the mass, momentum, and energy budget of the atmosphere. *Archiv. Meteor. Geophys. Bioklima.*, A, **7**, 133–157. (101)

London, J., 1957. *A study of the atmospheric heat balance.* Final report, contract AF 19(122)–165, Dept. of Meteorology and Oceanography, New York Univ. (92)

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167. (102, 109)

Lorenz, E. N., 1956: A proposed explanation for the existence of two régimes of flow in a rotating symmetrically-heated cylindrical vessel. *Fluid models in geophysics.* Proc. First Sympos. on the Use of Models in Geophys. Fluid Dyn., R. R., Long, Ed. Washington, U.S. Govt. Printing Off., 73–80. (124)

Lorenz, E. N., 1960: Generation of available potential energy and the intensity of the general circulation. *Dynamics of Climate.* R. L. Pfeffer, Ed. New York, Pergamon Press, 86–92. (102)

Lorenz, E. N., 1962: Simplified dynamic equations applied to the rotating-basin experiments. *J. Atmos. Sci.*, **19**, 39–51. (125, 138)

Lorenz, E. N., 1963: The mechanics of vacillation. *J. Atmos. Sci.*, **20**, 448–464. (138)

Manabe, S., Smagorinsky, J., and Strickler R. F., 1965: Simulated climatology of a general circulation model with a hydrological cycle. *Mon. Weather Rev.*, **93**, 769–798. (39, 128, 134)

Margules, M., 1903: Über die Energie der Stürme. *Jahrb. Zentralanst. Meteor., Vienna*, 1–26. English trans. : Abbe, C., 1910 : *The mechanics of the Earth's atmosphere*, 3rd Coll. Washington, Smithsonian Inst., 533–595. (99, 102)

Maury, M. F., 1855: *The Physical geography of the sea.* New York, Harper, 287 pp. (63)

Mintz, Y., 1951: The geostrophic poleward flux of angular momentum in the month of January 1949. *Tellus*, **3**, 195–200. (79)

Mintz, Y., 1954: The observed zonal circulation of the atmosphere. *Bull. Amer. Meteor. Soc.*, **35**, 208–214. (33, 46)

Mintz, Y., 1955: *Final computation of the mean geostrophic flux of angular momentum and of sensible heat in the winter and summer of 1949.* Article 5, Final Report, Contract AF 19(122)–48, Dept. of Meteorology, Univ. of California, Los Angeles. (88, 92, 94)

Mintz, Y., 1964: Very long-term global integration of the primitive equations of atmospheric motion. *WMO-IUGG Symposium on Research and Development Aspects of Long-Range Forecasting.* World Meteor. Org., Tech. Note No. 66, 141–155. (133)

Mintz, Y. and Lang, J., 1955: *A model of the mean meridional circulation.* Article 6, Final Report, AF 19(122)–48, Dept. of Meteorology, Univ. of California, Los Angeles. (94)

Murphy, J. J., 1856: *On the circulation of the atmosphere.* Belfast Nat. Hist. and Phil. Soc. (oral presentation). (65)

Namias, J., 1950: The index cycle and its role in the general circulation. *J. Meteor.*, **7**, 130–139. (47)

Obasi, G. O. P., 1963: *Atmospheric momentum and energy calculations for the southern hemisphere during the IGY.* Sci. Report No. 6, Contract AF 19(604)–6108, Dept. of Meteorology, Mass. Inst. of Technology. Also (1963): Poleward flux of atmospheric angular momentum in the southern hemisphere. *J. Atmos. Sci.*, **20**, 516–528. (32, 81)

Oberbeck, A., 1888: Über die Bewegungserscheinungen der Atmosphäre. *Sitz.-Ber. Akad. Wiss. Berlin*, 383–395, 1129–1138. English trans.: Abbe, C., 1893: *The mechanics of the Earth's atmosphere.* Washington, Smithsonian Inst., 177–197. (67, 136)

Obukhov, A. M., 1949: K voprosu o geostroficheskom vetre (The problem of geostrophic winds). *Izv. Akad. Nauk SSSR*, Ser. Geogr. i Geofyz., Vol. **13**, No. 4, 281–306. (148)

Oort, A. H., 1964 a: On estimates of the atmospheric energy cycle. *Mon. Weather Rev.*, **92**, 483–493. (101, 111)

Oort, A. H., 1964 b: On the energetics of the mean and eddy circulations in the lower stratosphere. *Tellus*, **16**, 309–327. (112)

Palmén, E. and Newton, C. W., 1967: Atmospheric circulation systems. New York, Academic Press (in press). (39, 90, 92)

Palmén, E. and Vuorela, L., 1963: On the mean meridional circulations in the northern hemisphere during the winter season. *Q. J. Roy. Meteor. Soc.*, **89**, 131–138. (45, 53, 92)

Pedlosky, J., 1964: The stability of currents in the atmosphere and the ocean: Parts I and II. *J. Atmos. Sci.*, **21**, 201–219, 342–353. (141, 150)

Peixoto, J. P., 1958: *Hemispheric humidity conditions during the year 1950.* Sci. Report No. 3, Contract AF 19(604)–2242, Dept. of Meteorology, Mass. Inst. of Technology. (85)

Peixoto, J. P., 1960: *Hemispheric temperature conditions during the year 1950.* Sci. Report No. 4, Contract AF 19(604)–6108, Dept. of Meteorology, Mass. Inst. of Technology. (39, 85)

Peixoto, J. P. and Crisi, A. R., 1965: *Hemispheric humidity conditions during the IGY.* Sci. Report No. 6, Contract AF 19(628)–2408, Dept. of Meteorology, Mass. Inst. of Technology. (39, 85)

Peng, L., 1963, 1965: *Stratospheric wind temperature and isobaric height conditions during the IGY period, Parts II and III.* Sci. Reports Nos. 10, 15, Contract AT (30–1) 2241, Dept. of Meteorology, Mass. Inst. of Technology. (39)

Phillips, N. A., 1951: A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Meteor.*, **8**, 381–394. (131)

Phillips, N. A., 1956: The general circulation of the atmosphere: a numerical experiment. *Q. J. Roy. Meteor. Soc.*, **82**, 123–164. (131, 142, 143, 150)

Phillips, N. A., 1959: An example of non-linear computational instability. *The atmosphere and the sea in motion.* B. Bolin, Ed. New York, Rockefeller Inst. Press, 501–504. (132)

Phillips, N. A., 1960: Numerical weather prediction. *Advances in computers.* New York, Academic Press, 43–90. (129)

Priestley, C. H. B., 1949: Heat transport and zonal stress between latitudes. *Q. J. Roy. Meteor. Soc.*, **75**, 28–40. (78)

Priestley, C. H. B., 1951: A survey of the stress between the ocean and atmosphere. *Australian J. Sci. Res.*, A, **4**, 315–328. (52)

Priestley, C. H. B. and Troup, A. J., 1964: Strong winds in the global transport of momentum. *J. Atmos. Sci.*, **21**, 459–460. (88)

Reed, R. J., 1965: The present status of the 26-month oscillation. *Bull. Amer. Meteor. Soc.*, **46**, 374–387. (10, 46)

Richardson, L. F., 1922: *Weather prediction by numerical process.* Cambridge, University Press, 236 pp. (131)

Riehl, H. and Fultz, D., 1958: The general circulation in a steady rotating-dishpan experiment. *Q. J. Roy. Meteor. Soc.*, **84**, 389–417. (119)

Rossby, C.-G., 1938a: On the maintenance of the westerlies south of the polar front. Fluid mechanics applied to the study of atmospheric circulations. *Pap. Phys. Oceanog. Meteor.*, Vol. **7**, No. 1, 9–17. (77)

Rossby, C.-G., 1938b: On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Marine Res.*, **1**, 239–263. (148)

Rossby, C.-G., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semipermanent centers of action. *J. Marine Res.*, **2**, 38–55. (21, 23, 46, 119, 137)

Rossby, C.-G., 1941: The scientific basis of modern meteorology. *Climate and Man.* Yearbook of Agriculture. Washington, U.S. Govt. Printing Off., 599–655. Also (1945) in *Handbook of meteorology*, F. A. Berry, E. Bollay and N. R. Beers, Eds. New York, McGraw-Hill, 501–529. (78)

Rossby, C.-G., 1947: On the distribution of angular velocity in gaseous envelopes under the influence of large-scale horizontal mixing processes. *Bull. Amer. Meteor. Soc.*, **28**, 53–68. (78)

Rossby, C.-G., 1957: Aktuella meteorologiska problem. *Svensk Naturvetenskap 1956*, 15–80. English translation, 1959: Current problems in meteorology. *The atmosphere and the sea in motion*, B. Bolin, Ed. New York, Rockefeller Inst. Press, 9–50. (97)

Saltzman, B., 1957: Equations governing the energetics of the larger scales of atmospheric turbulence in the domain of wave number. *J. Meteor.*, **14**, 513–523. (112)

Saltzman, B. and Fleisher, A., 1961: Further statistics on the modes of release of available potential energy. *J. Geophys. Res.*, **66**, 2271–2273. (112)

Saltzman, B., Gottuso, R. M. and Fleisher, A., 1961: The meridional eddy transport of kinetic energy at 500 mb. *Tellus*, **13**, 293–295. (90)

Saltzman, B. and Teweles, S., 1964: Further statistics on the exchange of kinetic energy between harmonic components of the atmospheric flow. *Tellus*, **16**, 432–435. (112)

Sellers, W. D., 1966: *Physical Climatology.* Chicago, Univ. of Chicago Press, 272 pp. (49, 54)

Smagorinsky, J., 1963: General circulation experiments with the primitive equations. I. The basic experiment. *Mon. Weather Rev.*, **91**, 99–164. (133)

Smagorinsky, J., Manabe, S. and Holloway, J. L., 1965: Numerical results from a nine-level general circulation model of the atmosphere. *Mon. Weather Rev.*, **93**, 727–768. (130, 133)

Starr, V. P., 1948: An essay on the general circulation of the Earth's atmosphere. *J. Meteor.*, **5**, 39–43. (78)

Starr, V. P. and Dickinson, R. E., 1964: Large-scale vertical eddies in the atmosphere and the energy of the mean zonal flow. *Geof. Pura e Appl.*, **55**, 133–136. (92)

Starr, V. P. and Long, R. R., 1953: The flux of angular momentum in rotating model experiments. *Geophys. Res. Pap.*, No. 24, 103–113. (117)

Starr, V. P. and White, R. M., 1951: A hemispherical study of the atmospheric angular-momentum balance. *Q. J. Roy. Meteor. Soc.*, **77**, 215–225. (80)

Starr, V. P. and White, R. M., 1954: Balance requirements of the general circulation. *Geophys. Res. Pap.*, No. 35, 57 pp. (31, 80)

Sverdrup, H. U., 1917: *Der nordatlantische Passat.* Veröff. Geophys. Inst. Univ. Leipzig, Vol. **2**, No. 1, 96 pp. (100)

Thompson, P. D., 1959: Numerical weather analysis and prediction. New York, Macmillan, 170 pp. (129)

Thomson, J., 1857: *Grand currents of atmospheric circulation.* British Assoc. Meeting, Dublin. (65, 115, 136)

Thomson, J., 1892: On the grand currents of atmospheric circulation. *Phil. Trans. Roy. Soc.*, A, **183**, 653–684. (66)

Tracy, C., 1843: On the rotary action of storms. *Amer. J. Sci. and Arts*, **45**, 65–72. (64)

Tucker, G. B., 1959: Mean meridional circulations in the atmosphere. *Q. J. Roy. Meteor. Soc.*, **85**, 209–224. (45)

Van Mieghem, J., 1956: The energy available in the atmosphere for conversion into kinetic energy. *Beitr. Phys. Fr. Atmos.*, **29**, 129–142. (104)

Vettin, F., 1857: Über den aufsteigenden Luftstrom, die Entstehung des Hagels und der Wirbel-Stürme. *Ann. Phys. und Chem.* (2), Leipzig, 102, 246–255. (115)

Vuorela, L. A. and Tuominen, I., 1964: On the mean zonal and meridional circulations and the flux of moisture in the northern hemisphere during the summer season. *Pure and Appl. Geophys.*, **57**, 167–180. (45)

Waldo, F., 1893: *Modern meteorology.* London, Walter Scott, and New York, Scribner's, 460 pp. (69)

White, R. M., 1949: The role of mountains in the angular-momentum balance of the atmosphere. *J. Meteor.*, **6**, 353–355. (52)

White, R. M., 1951: The meridional flux of sensible heat over the northern hemisphere. *Tellus*, **3**, 82–88. (80)

White, R. M., 1954: The counter-gradient flux of sensible heat in the lower stratosphere. *Tellus*, **6**, 177–179. (88, 112)

Widger, W. K., 1949: A study of the flow of angular momentum in the atmosphere. *J. Meteor.*, **6**, 291–299. (79)

Wiin-Nielsen, A., 1962: On transformation of kinetic energy between the vertical shear flow and the vertical mean flow in the atmosphere. *Mon. Weather Rev.*, **90**, 311–323. (112)

Wiin-Nielsen, A., Brown, J. A. and Drake, M., 1964: Further studies of energy exchange between the zonal flow and the eddies. *Tellus*, **16**, 168–180. (112)

Willett, H. C., 1949: Long-period fluctuations of the general circulation of the atmosphere. *J. Meteor.*, **6**, 34–50. (47)

Winston, J. S., 1967: Global distribution of cloudiness and radiation for seasons as measured from weather satellites. *World Survey of Climatology*, Vol. 3, Ch. 7, Amsterdam, Elsevier Publishing Co. (in press). (55)

Yeh, T.-C. and Chu, P.-C., 1958: *Some fundamental problems of the general circulation of the atmosphere* (in Chinese with English summary). Inst. Geophys. Meteor., Academia Sinica. (52)

LIST OF SYMBOLS

Symbols containing subscripts are not separately listed except where the subscript alters the meaning of the symbol. A subscript o denotes a value at the Earth's surface. Subscripts z and e attached to the symbols A, K, G, C, D denote zonal and eddy quantities. Subscripts λ, φ, z denote the components of a vector.

A	Available potential energy
a	Mean radius of Earth, 6.37×10^3 km, (or radius of container in laboratory experiment)
b	Inner radius of annulus in laboratory experiment
C	Rate of conversion of available potential energy into kinetic energy by reversible adiabatic processes
c	Specific heat of liquid water, 41.85×10^6 cm^2 sec^{-2} deg^{-1}
C_A	Rate of conversion of zonal available potential energy into eddy available potential energy
c_D	Surface drag coefficient
C_K	Rate of conversion of zonal kinetic energy into eddy kinetic energy
c_p	Specific heat of air at constant pressure, 9.96×10^6 cm^2 sec^{-2} deg^{-1}
c_v	Specific heat of air at constant volume, 7.09×10^6 cm^2 sec^{-2} deg^{-1}
D	Rate of dissipation of kinetic energy by friction
E	Rate of upward turbulent transfer of water vapour, per unit horizontal area
E_0	Rate of evaporation from Earth's surface, per unit area
$\mathbf{F}$	Frictional force per unit mass
f	Coriolis parameter
G	Rate of generation of available potential energy by heating
$\mathbf{g}$	Acceleration of gravity
g	Mean magnitude of $\mathbf{g}$, 981 cm sec^{-2}
H	Rate of production of internal energy by heating
h	Depth of fluid in laboratory experiment
I	Internal energy per unit mass
$\mathbf{i}$	Unit vector directed eastward
$\mathbf{j}$	Unit vector directed northward
K	Kinetic energy per unit mass
$\mathbf{k}$	Unit vector directed upward
k	Average number of arithmetic operations needed to compute a single time derivative in a numerical experiment
L	Latent heat of condensation
M	Absolute angular momentum per unit mass about Earth's axis
m	Number of dependent variables in a numerical experiment
N	Efficiency factor, $1 - p^{-\kappa} P^{\kappa}$
n	Total number of time steps in a numerical experiment
P	Average pressure on an isentropic surface

p	Pressure
P_0	Rate of precipitation, per unit area
P_{00}	Standard pressure, 1000 mb
Q	Rate of heating per unit mass
q	Specific humidity
q_s	Specific humidity of saturated air at given pressure and temperature
R	Gas constant for air, $c_p - c_v$, 2.87×10^6 cm^2 sec^{-2} deg^{-1}
$\mathbf{r}$	Position, with respect to Earth's centre
r	Magnitude of $\mathbf{r}$, distance from Earth's centre
R_o	Rossby number
R_{OT}	Thermal Rossby number
R_w	Gas constant for water vapour, 4.62×10^6 cm^2 sec^{-2} deg^{-1}
S	Area of Earth's surface
s	Specific entropy
T	Absolute temperature
t	Time
T_a	Taylor number
T_v	Virtual temperature
$\mathbf{U}$	Horizontal wind velocity, horizontal projection of $\mathbf{V}$
u	Eastward component of $\mathbf{V}$
$\mathbf{U}_d$	Divergent irrotational part of $\mathbf{U}$
$\mathbf{U}_g$	Geostrophic wind velocity
u_g	Eastward component of geostrophic wind
$\mathbf{U}_r$	Rotational non-divergent part of $\mathbf{U}$
$\mathbf{V}$	Wind velocity
v	Northward component of $\mathbf{V}$
w	Upward component of $\mathbf{V}$
X	Arbitrary dependent variable
x	Eastward distance on beta-plane
y	Northward distance on beta-plane
z	Elevation, measured upward
α	Specific volume
β	Derivative of Coriolis parameter with respect to northward distance, df/dy
Γ	Vertical lapse rate of temperature, $-\partial T/\partial^2$
γ	Ratio of specific heats of air, c_p/c_v, 1.405
Γ_d	Dry-adiabatic lapse rate, g/c_p, 9.8°/km
δ	Horizontal divergence, $\nabla\cdot\mathbf{U}$
ε	Coefficient of thermal expansion, $d\ (1)n\alpha/dT$
ζ	Vorticity, $\nabla\cdot\mathbf{U}\times\mathbf{k}$
η	Efficiency of atmospheric energy cycle
θ	Potential temperature
κ	R/c_p, 0.288
λ	Longitude, measured eastward
μ	Coefficient of turbulent viscosity
ν	Kinematic viscosity
ρ	Density
σ	Static stability factor, $-(T/\theta)\partial\theta/\partial p$

τ	Frictional stress per unit horizontal area
Φ	Potential energy per unit mass, geopotential
φ	Latitude, measured northward
χ	Velocity potential for divergent velocity
Ψ	Stream function for meridional circulation
ψ	Stream function for rotational velocity
$\mathbf{\Omega}$	Angular velocity of Earth (or angular velocity of container in laboratory experiment)
Ω	Magnitude of $\mathbf{\Omega}$, 7.292×10^{-5} sec^{-1}
ω	Individual pressure change, dp/dt
$\{( )\}$ . . .	Integral over entire mass of atmosphere
$\widetilde{( )}$	Horizontal average
$( )''$ . . .	Departure from horizontal average
$[( )]$ . . .	Longitudinal average
$( )^*$ . . .	Departure from longitudinal average
$\overline{( )}$	Time average
$( )'$. . . .	Departure from time average