THE NATURE AND THEORY
OF THE GENERAL CIRCULATION
OF THE ATMOSPHERE

BY

EDWARD N. LORENZ

WORLD METEOROLOGICAL ORGANIZATION
1967

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_vT, 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

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

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

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

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

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

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

$$I = c_{\mathrm{v}}T. \tag{20}$$

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

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

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

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

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 $\mathbf{U}_\mathrm{d}$. 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 $K = \mathbf{U}_\mathrm{r} \cdot \mathbf{U}_\mathrm{r}/2$, and all the quadratic terms in the vorticity equation except those involving $\mathbf{U}_\mathrm{r}$ only must be discarded. The vorticity equation and the thermodynamic equation then assume the form

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

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

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

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

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 $c_\mathrm{p} T$, 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

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

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

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

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 $-\mathbf{\nabla}\psi\cdot\mathbf{\nabla}f\times\mathbf{k}$ 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 $\mathbf{\nabla}T\cdot\mathbf{\nabla}\chi$ 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 (~) 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

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

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

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

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

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.