551.513.1

The cause of the general circulation of the atmosphere

By E. T. EADY
Imperial College of Science and Technology, London

SUMMARY

The aim is to explain the fundamental causes of large-scale atmospheric behaviour and to deduce the observed mean flow pattern directly from the equations of motion, *etc.*, without additional hypotheses. It is inferred from empirical evidence that there exists a poleward and upward transfer of heat and a transfer of angular momentum polewards over most of the earth's surface but equatorwards in high latitudes. The possibility of the main transfer being the result of steady currents, including meridional circulations, is considered and rejected ; it is concluded that large-scale turbulence plays a fundamental role in the mechanism of the atmospheric " engine." The problem therefore is to explain *why* large-scale turbulence develops and to derive its properties with regard to transfer of heat and angular momentum. From earlier work on the stability of baroclinic motion it is concluded that the mean flow is always unstable and must continually generate turbulence. The mechanism by which instability is continually re-established, and turbulent flow consequently maintained, is described and its intimate relation to the observed heat transfer deduced. The transfer of angular momentum is a secondary feature associated with the " constraints," to which the complete set of equations and boundary conditions are equivalent, which modify the form of the overturning process represented by the growth of turbulent disturbances. Difficulties in computing this transfer theoretically arise from the fact that angular momentum is not even approximately conserved during motion. The hypothesis that the vertical component of absolute vorticity is conserved is rejected both for theoretical reasons and because it leads to incorrect results in middle latitudes. It is suggested however that this hypothesis may be valid for approximate calculations near the poles and near the equator because the motion is more nearly barotropic. Transfer behaviour in middle latitudes is deduced from a consideration of the structure of growing disturbances. Surface friction plays an essential part in the circulation in that it limits the amplitude both of individual disturbances and of the zonal currents resulting from angular momentum transfer. The mechanism by which its effect is "geared" to motion above the friction layer is described and it is inferred that *driven* meridional circulations must exist. These meridional circulations are secondary effects so far as heat transfer is concerned. They are consistent with observed precipitation belts and are slow enough to allow maintenance of the observed lapse rate by radiative cooling in the subsiding currents.

1. HEAT TRANSFER ON A LARGE SCALE

The present paper is concerned to explain the observed eddy-transfer of heat and angular momentum between latitudes and incidentally to explain why there is any atmospheric " circulation " at all. The term *circulation* is to some extent misleading in that it suggests that the main heat and momentum balance is maintained by quasi-steady currents whereas according to the views expressed here these currents are secondary features. If this is understood there is no objection to using the conventional term.

It is convenient to commence by examining the facts about heat and angular momentum transfer which have to be explained. Let us consider first heat transport between different zones (bounded by lines of latitude).

(1) The calculations of Simpson (1928) and others indicate that the different zones of the atmosphere are not in radiative equilibrium. Poleward of about latitude 35° there is a net annual loss of heat by radiation ; equatorward of this latitude there is a compensating net annual gain of heat. Although the calculations are necessarily rough there is no reasonable doubt about the general nature of the result. We must infer a net pole-

ward transfer of heat and the only plausible mechanism would appear to be advection by air and/or ocean currents. Since the surface ocean currents are driven by the atmospheric circulation a plausible inference is that at least part and possibly the major part of the heat transfer is by air currents.

(2) Calculations of radiative transfer in the vertical indicate that on the average the free atmosphere below the tropopause is cooling, owing to radiative processes alone, at a rate of something like 1°C per day. Recent computations, not yet published, by Howell at Imperial College, of the change of potential temperature along (3-dimensional) trajectories in the middle troposphere have given independent confirmation of this result. We must infer a net upward transfer of heat and the only plausible mechanism would appear to be advection by air currents. This would include, of course, heat transfer by ordinary convection but this is not necessarily the sole mechanism.

Combining these results it would appear that we have to account for a poleward and upward transfer of heat. The most plausible hypothesis is that this transfer is effected by the motion of relatively potentially warm air polewards and upwards while relatively potentially cold air moves equatorwards and downwards to take its place. Since we have already taken radiation into account we must *logically* regard this process as adiabatic even though in practice the adiabatic and radiative changes are not completely separated. We then notice an interesting feature. For simplicity we shall consider the air to be unsaturated so that in an adiabatic process entropy is conserved. (Condensation is certainly an important atmospheric process but to take it into account merely complicates the argument without affecting the general nature of the result.) Hence we may regard the postulated transfer as a transfer of entropy $\Delta\phi$. If T_1 is the mean temperature at the low-level low-latitude end and T_2 the mean temperature at the high-level high-latitude end then statistical equilibrium requires that the net incoming radiation in low latitudes is proportional to $T_1\Delta\phi$ while the net outgoing radiation in high latitudes is proportional to $T_2\Delta\phi$. Since $T_1 > T_2$ more heat is received than is re-radiated. Thus our supposed mechanism implies that the atmosphere behaves as an engine and does work proportional to $(T_1 - T_2)\ \Delta\phi$. Now by hypothesis statistical equilibrium is maintained so there is no ultimate change in either the potential or internal energy of the atmosphere : the work done must manifest itself as a growth of kinetic energy to balance statistically the loss of kinetic energy (*i.e.*, its conversion into heat which is then radiated away) by frictional forces. In the last analysis there is a complete radiation balance but with the mechanism we have supposed kinetic energy is maintained at a finite level (*i.e.*, there must be a " circulation " of some kind), otherwise there would be no frictional forces. (See Fig. 1)

The inferred mechanism makes it clear *how* an atmospheric circulation is maintained — it is a necessary consequence of the existence of a heat transport in a certain direction. What we have to explain is why such a heat transport exists, *i.e.*, why the atmosphere is not in radiative equilibrium. Before considering the cause of this heat transport we may note that not only is the sink colder than the source, as required by the Second Law of Thermodynamics, but it is also at a higher

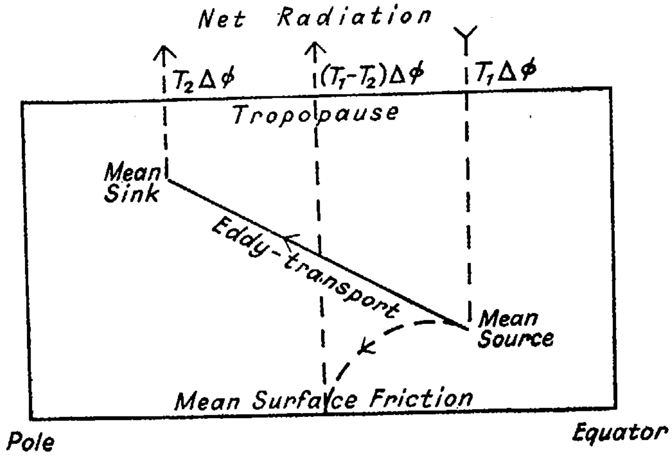

Figure 1. Heat and energy transfer in the atmosphere. The straight dashed lines indicate radiative transfer. The curved dashed line indicates generation of kinetic energy through the working of the atmospheric " engine." Bodily (large-scale eddy) transport of heat is indicated by a full line. The figure represents a section in a meridional plane.

level. Various writers (see *e.g.* V. Bjerknes *et al* 1934, p. 200) have attempted to show that this is a necessary feature for the maintenance of a steady circulation. We shall see that it is also necessary for the maintenance of large-scale turbulence. Thus this feature is almost certainly necessary for the maintenance of any circulation at all and gives no clue to its nature. Nor should we infer that the existence of a circulation is accidental in so far as it depends on the relative levels of source and sink. The genesis of source and sink regions (*i.e.*, regions not in radiative equilibrium) is the *result* of an initial circulation and the mean circulation pattern adjusts itself to maintain statistically the heat and cold sources at levels consistent with its own maintenance.

2. WHY TURBULENT MOTION MUST BE INFERRED

Let us consider whether the atmospheric circulation could be maintained by steady currents. In the northern hemisphere the observed mean currents are somewhat irregular and presumably related to the topography and the distribution of land and sea. In the southern hemisphere there is still a circulation but this appears to be more nearly zonal and moreover the zonal currents are intensified. The inference is that the distribution of land and sea plays no *fundamental* part in maintaining the circulation and that the cause of the circulation must be revealed if we consider an ideal earth with a uniform surface. Then the only *steady* currents allowed by symmetry, in addition to zonal motion, are meridional circulations. The zonal motion does no work and must be a secondary feature. Let us consider the possibility of steady meridional circulations. These could effect the required heat transfer and calculation shows (V. Bjerknes *et al* 1934, p. 732) that the currents required are too small for direct measurement to give conclusive evidence for or against. The hypothesis that such currents exist is attractive, because if the earth did not rotate (while the sun revolved round the earth) there is little doubt that such currents would be generated. However the distribution of vertical motion does not fit with observed precipitation belts and when we take the earth's rotation into account we encounter theoretical difficulties. If we neglect frictional forces large-scale meridional circulations are impossible. This is because the angular momentum of each ring of air is necessarily conserved during its motion. Any motion in a meridional plane would cease when the thermal wind equation is satisfied: the temperature contrast between pole and equator is too small to allow the air to traverse more than a small part of its hypothetical circuit. When the thermal wind equation is satisfied we have a state of zonal motion in equilibrium and moreover (except *possibly* in very narrow belts on the equatorial side of jet streams) in stable equilibrium from the point of view of displacements in a *meridional* plane (Eady 1949, p. 51). Frictionless meridional circulations fail because they are associated with much too large a transfer of angular momentum. Meridional circulations *could* be maintained if the increasing angular momentum difference between poleward and equatorward currents were continually being destroyed by frictional forces, the only plausible mechanism being the transfer of angular momentum by turbulence, *i.e.*, eddy-viscosity. Turbulence is therefore a necessary feature of atmospheric motion whether the heat transfer occurs directly as a result of it or whether the heat transfer is by meridional circulations. Any theory of the atmospheric circulation must be based on a theory of (large-scale) atmospheric turbulence. Moreover this theory must go deeper than most existing theories of turbulence. It is futile to beg the question by introducing hypothetical coefficients of eddy-transfer. We have not only to establish the cause of the turbulence but also to derive from first principles its properties with regard to transfer of heat, angular momentum, *etc.* As we shall see, these properties are not properties of turbulent motion in general but depend on the mechanism by which turbulence is generated and maintained. We shall find that the main heat transfer is direct and necessarily related to the maintenance of turbulence. Meridional circulations do exist and are related to turbulent transfer of angular momentum in the manner described above but they are secondary features different in structure from the hypothetical circulations postulated above and much feebler.

3. THE CAUSE OF LARGE-SCALE TURBULENCE

We can infer, as a theoretical necessity, that atmospheric motion is turbulent if we can show that any hypothetical laminar motion is unstable. Then any small irregularity must lead to the generation of turbulent disturbances. The statistical effect of these disturbances is an alteration in the mean flow and the new flow pattern is possible only because of the transfer properties of turbulence. As friction damps out the turbulent disturbances we must revert towards laminar motion which, by hypothesis, is unstable so that eventually new turbulence is generated, and so on. The distinction between the breakdown process and subsequent regeneration of an unstable system (associated with frictional dissipation of turbulent energy and radiative compensation of the heat transferred by turbulence) is logically convenient although such a clear-cut distinction is not observed in practice. What happens is that, precisely because the breakdown is irregular, the potentialities for the generation of turbulence are not immediately exhausted and regeneration begins while breakdown is still occurring. The result is that the actual mean motion is itself unstable. However the distinction is, at least in the present instance, quite trivial since the actual mean motion and the hypothetical laminar flow differ, in essence, only quantitatively. Now a state of motion is unstable if we can demonstrate the existence of *any* kind of unstable disturbance. Moreover on a uniform earth the only type of laminar motion we need consider is zonal motion satisfying the thermal wind equation, *i.e.*, steady baroclinic flow. But the writer has shown elsewhere (Eady 1949, pp. 36-46) that this flow is unstable. The disturbances which develop correspond to overturning of the air roughly in planes having half the slope of the broad-scale isentropic surfaces and result in the transfer of potentially warm air polewards and upwards side by side with equatorwards and downwards transfer of potentially cold air. (This type of instability is not to be confused with what is normally referred to as " dynamic instability " which involves overturning in a vertical plane.) If in place of the hypothetical laminar flow we commence with the actual (baroclinic) mean flow we obtain a similar result. (There is no " criterion " for the development of *this* kind of turbulence.)

Since the kind of turbulent motion which theory has shown ought to exist produces heat transfer of the kind inferred from observation we have some grounds for claiming to have explained in principle the fundamental cause of atmospheric motion. Complete quantitative explanation is not yet possible because of the rudimentary state of the theory of turbulence. Nevertheless the intimate relation between heat-transfer and the maintenance of turbulence is clear. We have already found it convenient to distinguish logically between the (adiabatic) generation of turbulence and the regeneration of instability by radiation. Considering the adiabatic part, the overturning process initially manifest as an unstable wave, we find that this occurs with the transformation of potential and internal energy into the kinetic energy of the disturbance. The potential and internal energies of the broad-scale distribution (which, as is well known, are closely related) are reduced precisely *because* potentially warm (and therefore less dense) air moves to higher levels and the mean centre of gravity of the atmosphere is depressed. A simultaneous poleward component of the motion of the warm air is necessary because continued displacement can take place (just as in ordinary convection) only if ascending air remains warmer than its " environment " and this implies motion at a smaller slope than that of the broad-scale isentropic surfaces. In this analysis we have ignored any contribution to turbulent energy from changes in zonal kinetic energy during the adiabatic process because calculation shows they must be quite small (of the order of the reciprocal of the broad-scale Richardson number). These changes are associated with changes in the thermal gradient consequent on the transfer of heat and with meridional displacements of air but since the initial state is restored later by radiation no meridional *circulation* occurs. The intimate relation between turbulence and heat-transfer is merely an expression of the fact that large-scale atmospheric turbulence is thermally driven. A similar relation holds for turbulence resulting from ordinary convection but *not* for turbulence resulting from shearing motion in a boundary layer or generated at a velocity discontinuity maintained mechanically. In these latter cases the intimate relation is between turbulence and momentum-transfer.

4. THE SIGNIFICANCE AND APPLICATION OF THE RESULTS OF PERTURBATION ANALYSIS

In view of the approximations made in the theory of instability of rotating baroclinic fluids and the rudimentary nature of the results one or two remarks may be apposite to avoid misunderstanding. Firstly, this theory was developed initially ignoring frictional forces whereas we have inferred that in the long run the kinetic energy destroyed by friction must exactly balance that generated in the atmospheric " engine." The point is that the theory, as developed so far, applies only to the initial stage of development of turbulent disturbances. In this stage, and only during this stage, the growth of kinetic energy considerably exceeds the frictional loss (as may be verified by rough calculations) and the modifications which would result from including frictional terms are quite small. This result is not really surprising for developments occur most rapidly (and become dominant) precisely in those regions where the excess of energy-generation over frictional loss is greatest and our approximations must be justified in these regions because of lack of uniformity. Of course a complete theory of the whole life-history of disturbances must take friction into account directly but the rudimentary theory of the initial stages is sufficient to prove that turbulence must develop. Secondly, the theory was developed on the assumption of adiabatic motion whereas in the long run radiation must balance dynamical heat transfer. Once again the approximations involved are justified during, and only during, the initial stages when the dynamic energy transfer considerably exceeds that due to differential radiation : they must be justified if we admit non-uniformity. Both in regard to friction and to radiation we find that the *logical* distinction made earlier between the overturning process and restoration of the *status quo* is not far from being an actual distinction so far as the initial development of an individual disturbance is concerned.

Thirdly, the theory was developed with approximations equivalent to the assumption that from *some* points of view the motion could be considered as incompressible. The aim was to eliminate certain elastic forces which would complicate but not affect the general nature of the results (they are important only for disturbances propagated at a speed comparable with that of sound) and to obtain a more tractable continuity equation, the modifications due to which were examined in the detailed analysis but not in the general energy theory. It should be added that the modifications in the general theory are now clear. The " equivalent incompressible " model used is valid for the actual, compressible atmosphere if we let " potential " energy in the model correspond to " potential-plus-internal " energy in the atmosphere. Then the two systems are *dynamically* equivalent (with the exception of a distortion of wave-form for very deep disturbances) and this is sufficient for the uses to which the theory is put. It is unprofitable to use the model for *thermodynamic* discussions because we have postulated a fluid more complicated than a perfect gas — one with a large coefficient of thermal expansion but zero compressibility : for this type of discussion we revert to a consideration of the actual atmosphere.

Fourthly, we have brushed aside the complications associated with condensation of water vapour. If the atmosphere were and remained everywhere saturated the modifications in the theory would be trivial. We need only replace the entropy of unsaturated air (differing insignificantly from that of dry air) by that of air plus water vapour plus condensed water (cloud), equivalent to replacing dry by wet adiabatics, and the general theory is modified only quantitatively. In practice however we very frequently find ascent of saturated air associated with descent of unsaturated air. It is important to note that this does not necessarily involve serious difficulties in the theory of the *initial* development of disturbances. The phenomenon of saturation is equivalent to a (very considerable) reduction in static stability (and consequently in Richardson number) so that developments tend to be concentrated in saturated regions and to be more rapid but they are of the same general type as those in unsaturated air. Complications arise later because air not originally saturated becomes so, and vice versa. We may expect that the effect on the structure of disturbances is important but there is no reason to infer any change in the fundamental nature of the overturning process. The latter is in many ways closely analogous to the vertical overturning which occurs during ordinary convection and we are aware that the type of convection which involves cumulus formation is fundamentally similar to " dry " convection. Further complications arise because developments in regions containing both saturated and unsaturated air generate velocity discontinuities on which we may expect turbulence of the mechanically driven type to develop. But such developments are clearly secondary and the transfer associated with this (smaller-scale) turbulence must be much less than that associated with the fundamental, thermally driven turbulence. From a thermodynamic point of view the important modification is that the heat of vaporisation to restore initial water vapour content must be added to that required to restore initial temperature when computing radiative restoration of the *status quo* and we must include latent heat with real heat in evaluating transport. Effectively the air at low levels is at a higher temperature than measured by the thermometer (not exceeding its equivalent temperature) and the distinction between real and latent heat appears in the immediate potentialities for development (while the air is unsaturated) rather than in the ultimate result.

Although the *ultimate* objective of our analysis is a complete quantitative explanation of atmospheric behaviour it is clear that owing to the rudimentary state of development of the theory we cannot at present hope to do more than make clear the basic fundamental principles. It is certainly encouraging to note that the dimensions and growth-rate of disturbances predicted by the theory of unstable waves on a uniform baroclinic current agree well enough in order of magnitude with observations and that there is also very frequently a fair amount of agreement as regards motion and structure but we can hardly expect anything like exact agreement nor should we be surprised if from time to time disturbances develop which appear to be of a different kind. As developed so far, theory refers only to an idealised system in which trains of disturbances develop independently in a regular manner. In practice disturbances develop from irregular systems irregularly and in short trains, frequently not exceeding a single wave-length. Whether or not it is theoretically convenient to regard disturbances as interacting they certainly appear to do so and it is only occasionally that we can identify a relatively " pure " wave-disturbance. This result is not surprising because theory *predicts* irregular development except for the *exactly* uniform initial condition postulated to give a tractable theory. It does however raise the question of what we mean by an individual disturbance. In actual irregularly-developing systems it may well be that we can assign no more than a rather hazy individuality to disturbances. If this is so it implies limitations to the present method especially as regards detailed bahaviour. But, in the first place, it is the virtue of the present approach that it makes readily intelligible the basic principles governing atmospheric behaviour and without this understanding we can hardly hope to develop a more refined technique. And, in the second place, the limitations are not necessarily so serious when it comes to evaluating large-scale long-term statistical behaviour : we have the example of statistical physics where a certain amount of haziness regarding the atomic constituents may not preclude the accurate evaluation of statistical properties. The point is that although individual disturbances may differ quite widely from the ideal model the basic driving forces exist, in a more or less distorted form, in the majority of them and must do so if the atmospheric engine is to continue working.

It is incorrect to argue that because atmospheric disturbances are most often " finite," *i.e.*, there does not exist a stage when we can see clearly an individual disturbance growing from small beginnings, we can learn nothing from the theory of small perturbations. The forces responsible for the instability of simple systems do not cease to operate when the disturbances become finite, otherwise a new steady flow would result, contrary to observation. All that is implied by the criticism is what has already been admitted, that the concept of an individual disturbance in practice is necessarily hazy and that we cannot hope to verify our theory from synoptic charts in more than a very rough manner. Independent verification is made more difficult by the fact that the concept of an individual disturbance is often even more hazy than it need be. Precisely because an adequate theoretical background has been lacking the idea has been allowed to grow up that the essential feature of the disturbance is a centre of low pressure on the surface chart. Now even rudimentary theory teaches us that the anticyclones (or ridges) play an equally important part and the term " disurbance " should apply to the complete structure. Moreover disturbances have a three-dimensional structure and in attempting to identify disturbances of a given type and the process which their development represents we cannot confine our attention to surface charts. If we merely follow surface " lows " we not only ignore independent evidence but run the risk of confusing a number of different things (changes of pressure associated with development on a larger scale, changes associated with change of latitude, changes associated with distortion of structure, *etc.*, as well as changes associated with true development). The relative (three-dimensional) motion of warm and cold air is often a better guide to what is happening because it is more directly related to the fundamental process which maintains atmospheric motion.

The dynamical part of the process envisaged consists essentially of a transformation of potential and internal energy into eddy kinetic energy and the type of disturbance which becomes dominant does so because it effects this transformation most rapidly. Even in the model, where development is supposed regular, the efficiency of transformation is not a maximum everywhere (*i.e.*, the motion is not everywhere in the plane corresponding to maximum energy-release) because of the constraints represented by all the equations of motion and boundary conditions. In the more complex (and more realistic) models we find regions where the reverse process is occurring : the principle of selection by maximum growth-rate applies to the *total* energy released. Although the details are not yet clear it is likely that the formation of jet-streams is an example of this reverse process. In the actual atmosphere, where development is irregular, the variety of ways in which the reverse process may occur in limited regions is increased. In particular, because large-scale disturbances are dispersive we may get energy concentrated in regions where it has not been released. That is why it was stated above that we should not be surprised to find disturbances originating in quite a different manner from those associated with the direct overturning process. But from a theoretical point of view all these complications are secondary in that they could not continue to occur without the main process, which alone is self-sufficient.

5. ANGULAR MOMENTUM TRANSFER BETWEEN LATITUDES

Let us now consider the problem of transfer of angular momentum between latitudes. As in the case of heat transfer we can infer from observations the actual transfer which we have to explain. Seasonal charts of surface isobars and surface winds indicate in each hemisphere belts of easterly winds in low latitudes and belts of westerlies in middle latitudes with smaller belts of easterlies in high latitudes. Although these are only average results there can hardly be any doubt that there is in each belt a mean torque due to friction at the earth's surface acting roughly in the opposite direction to the mean wind. Since friction is destroying westerly angular momentum in the west-wind belts and creating it in the east-wind belts, equivalent to a transfer of angular momentum from the west-wind to the east-wind belts since in the long run the total angular momentum of the atmosphere is unchanged, statistical equilibrium can be maintained only if some other process is continually transferring angular momentum in the opposite direction (see Fig. 2). This other process might be either meridional overturning or turbulent transfer associated with the large-scale turbulent motion whose existence has already been inferred. Now we have seen that meridional overturning cannot take place in the absence of turbulent transfer of angular momentum precisely because frictionless flow is associated with such a large angular momentum transfer. Hence the fundamental cause of angular momentum transfer and the consequent zonal currents must be turbulent transfer. What we have to show is that large-scale turbulence necessarily involves transfer of angular momentum. Having determined the turbulent transfer we shall then have to re-examine the question of meridional circulation.

The existence of turbulent motion allows us to infer directly, as a necessary consequence, *one* transfer property, that associated with the supply of energy to the turbulent disturbances, because the transfer must be in such a direction as to reduce the energy (kinetic or potential-plus-internal) of the mean motion. If atmospheric

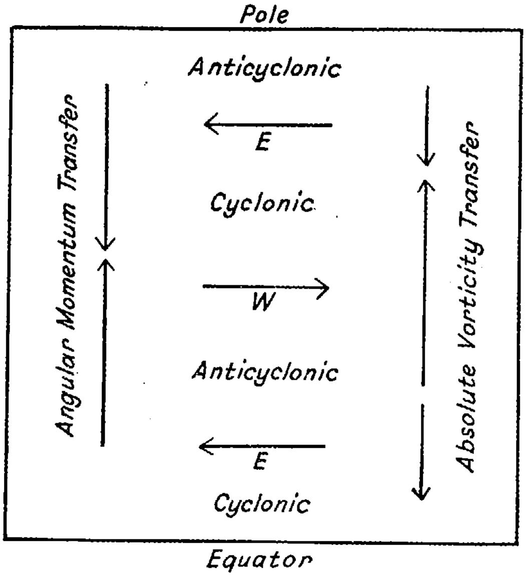

Figure 2. Angular momentum and absolute vorticity transfer between latitudes inferred from the observed zonal surface wind distribution. The figure represents a horizontal plane.

motion were mechanically driven this transfer property would refer to angular momentum and the turbulence would be such as to reduce zonal kinetic energy towards the minimum. It is easily demonstrated that this minimum corresponds to equality of mean angular velocity of the various zones. In order to maintain turbulence of this type we should have to have some other process acting which continually regenerated a state of inequality of mean angular velocity. It is true that such a process exists because a state of inequality is actually observed. (This is not quite the same thing as the inequality of angular velocity at the earth's surface, already referred to, because of the variation of wind with height but computations using upper air charts show that *mean* angular velocity is also unequally distributed.) But the only process which appears capable of accounting for this state of inequality is the *thermally* driven turbulent motion itself. In other words the inequality of mean angular velocity is not the source of turbulent energy but a result of turbulence. We have in fact an example of a *reverse* process (in a slightly more general sense than that in which the term was previously used) necessitated by the constraints imposed by the equations and boundary conditions of motion. The possibility of such a state of affairs is readily confirmed by computing the maximum growth-rate of hypothetical mechanically-driven disturbances and comparing them with thermally-driven disturbances, or, what is nearly the same thing, by comparing the available energies (*i.e.*, excess above the minimum produced by " mixing ") of kinetic and potential-plus-internal type. We find that the latter type of energy source very considerably exceeds the former, consistent with our supposition that it is the latter type which is the driver of atmospheric motion.

Our inference has been that the transfer of angular momentum is not *directly* connected with the maintenance of turbulence but is a secondary feature and we shall see that there is no general principle, as in the case of heat transfer, for determining its direction. In a sense the transfer of angular momentum is accidental since it is a by-product resulting from a particular set of constraints acting on developing disturbances. The mere existence of constraints is insufficient to determine the transfer or even its direction : we have to discover how these constraints act, *e.g.*, by considering their effect on individual disturbances. The analysis is made difficult by the fact that angular momentum, unlike entropy, is not even approximately conserved during motion. Within certain limits we can regard air particles as taking their entropy with them (*i.e.*, when the adiabatic approximation is valid) but it is only in very special cases that we can regard them, even very approximately, as taking their angular momentum with them : usually the torque associated with the pressure field is so large that the motion is nothing like that corresponding to conservation of angular momentum. We can speak of the transfer of angular momentum by disturbances because (and only because) the *total* angular momentum of the atmosphere is, in the long run, conserved but this is quite a different matter : the term " transfer " refers to the change produced in the mean flow pattern and does not necessarily, or even usually, imply simple bodily transport. The advantage of using conservative quantities when discussing transfer problems is so considerable that it is worth while examining the possibility of using some other quantity than angular momentum in our discussion. For example, if the vertical component of absolute vorticity were conserved during motion this would imply conservation of the total for the whole atmosphere and we could reformulate our problem as the determination of transfer of vorticity. Now in the very special case of purely meridional circulation angular momentum is conserved while in the very special case of purely horizontal barotropic flow vorticity is conserved. But in general, including the type of motion with which we are now concerned, *neither* quantity is conserved.

The transfer of any conservative quantity Q between two zones is obviously proportional to : $\int\int \rho V_y \cdot \Delta Q \, dx \, dz$ where ρ is the density, V_y the poleward velocity, x measures distance along a line of latitude, z distance vertically and ΔQ is the anomaly of Q (difference from mean value) per unit mass, the integration being over the complete circle of latitude and all heights. In other words the transfer is proportional to the correlation between ΔQ and V_y. Less obviously, the transfer of angular momentum, which is not conservative, is given by a formula of exactly the same type (Jeffreys 1926) *i.e.*, it is proportional to $\int\int \rho V_y \cdot r V_x \, dx \, dz$ where V_x is the velocity along and r the radius of the circle of latitude. (Another way of stating the result is that this formula measures the Reynolds stress.) The transfer is proportional to the correlation between V_x and V_y so that, knowing the observed transfer, what we have to explain is *why* V_x and V_y are correlated in a particular way. It is readily deduced from the observed zonal flow that we have to explain a positive correlation between V_x and V_y over most of the earth's surface but a negative correlation in high latitudes. It is interesting to note that the reality of this correlation has been shown by direct computation over a particular period, (Widger 1949).

6. HYPOTHESES OF CONSERVATION OF (A) VORTICITY, (B) ANGULAR MOMENTUM DURING TRANSFER

The form of the transfer formula does not distinguish between conservative and non-conservative quantities so it may not be immediately obvious why it is so much easier to deal with quantities of the former type. The fact is that when dealing with conservative quantities we can usually determine at least the direction of transfer at once. As an example consider hypothetical disturbances of barotropic flow in which the motion is supposed to remain horizontal. Then we know that the vertical component of absolute vorticity is a conservative quantity. If the mean flow pattern is similar to that observed on seasonal mean surface charts there is everywhere a strong gradient of mean absolute vorticity directed polewards. A particle which *has been* displaced polewards retains its absolute vorticity so that in the northern hemisphere it *must* have a negative vorticity anomaly. If the displacement is associated with a self-generating turbulent disturbance (initially an unstable wave) motion will continue in the same direction so that there is a negative correlation between poleward velocity and vorticity. Similarly, equatorward displacement is associated with positive vorticity anomaly and the turbulent motion as a whole must be associated with an equatorward transfer of absolute vorticity in the northern hemisphere. Taking into account the change in the sign of vorticity in the southern hemisphere we have the result that this kind of turbulence must give everywhere a southward transfer of absolute vorticity. Now from the observed flow pattern we have inferred the transfer of angular momentum needed to maintain it against surface friction and since, by hypothesis, total vorticity is conserved we may easily re-express our results in terms of the vorticity transfer needed to maintain the observed mean flow. We find that over most of the earth's surface (between very roughly 10° and 60° lat. in each hemisphere) the transfer must be northward, corresponding to the maintenance of the two major regions of vorticity anomaly, cyclonic anomaly centred near 60° lat. and anticyclonic anomaly centred near 30° lat. (see Fig. 2). This is directly contrary to the hypothesis that vorticity is conserved during motion. On the other hand the existence of minor centres of vorticity anomaly, anticyclonic in the polar regions, cyclonic near the equator, implies a southward transfer of absolute vorticity both in very high and in very low latitudes. In these extreme regions the hypothesis that absolute vorticity is approximately conserved *might* be valid. As a further example consider hypothetical disturbances in which flow is purely meridional so that angular momentum is a conservative quantity. Then since the mean flow shows a strong gradient of angular momentum which is everywhere directed equatorwards an argument similar to that used above shows that turbulent motion would be associated with a poleward transfer of angular momentum. This result is not *in itself* inconsistent with observation over most of the earth's surface but it is inconsistent with observation in high latitudes. Quite apart from other considerations relating to the possible mechanism of turbulence maintenance it is evident that neither the hypothesis of vorticity conservation nor the hypothesis of angular momentum conservation can account for the observed facts.

In passing we may note that with *conservative* quantities the transfer is always in such a direction as to destroy the gradient in the mean flow pattern (which, consequently, must be maintained by other means). Since in the dynamical theory we have treated entropy as a conservative quantity the result must apply to this particular case and it is easily verified that this is so. The apparent discrepancy, that entropy is transferred upwards against the gradient (for statically stable atmospheres) is easily removed. For when the static stability is positive there is always a horizontal component of the motion. Transfer takes place in a direction which makes a smaller angle with the horizontal than the isentropic surfaces and it is easily seen that its effect is to reduce the entropy gradient *in the direction of " mixing."* We may also note that this direction is different from that in which angular momentum is transferred. For angular momentum both the source and sink are on the earth's surface. Hence even if we can discover a process which transfers angular momentum in the free atmosphere we shall have to discover another process which brings the transferred momentum down to the earth's surface : in the free atmosphere the net transfer of angular momentum must be zero. We shall find that this downward transfer is effected by subsidiary quasi-steady meridional circulations which are in fact set up automatically. The possibility of such a solution (the subsidiary nature of the meridional circulation being a necessary condition) derives from the fact that meridional circulations transfer angular momentum vertically even more effectively than they transfer it horizontally.

7. THE CAUSE OF THE OBSERVED TRANSFER OF ANGULAR MOMENTUM

Since angular momentum is not conserved and we cannot determine the correlation between V_x and V_y directly from general considerations it appears that the only way in which this correlation may be determined theoretically is by computing the velocity field associated with individual disturbances. The computation is made difficult by the fact that, as might be expected from the consideration that the transfer is a *result* of the operation of constraints, we can only use models which reproduce fairly accurately the constraints (in particular the boundary conditions) actually obtaining in the atmosphere. (When we were studying entropy transfer this difficulty did not arise : the constraints could be varied within wide limits without affecting the direction of transfer.) It will be convenient to consider the effect of constraints as a combination of two separate effects.

The first effect, most noticeable in middle and low middle latitudes, is related to the curvature of the earth's surface in so far as this implies a variation of the Coriolis parameter with latitude. A " correction " to the most elementary theory which is most important for the larger (travelling long-wave) disturbances (Eady 1949, pp. 46-49) shows that besides the advective wave-velocity, corresponding to motion with the mean current in which the disturbance develops, and the purely imaginary wave-velocity corresponding to development there is an additional real wave velocity directed eastwards proportional to the square of the wavelength and the rate of change with latitude of the Coriolis parameter. This feature is similar to that exhibited by barotropic waves and has the same cause, namely the need to balance the divergence or convergence associated with north-south motion because of the variation of the Coriolis parameter. However the additional wave-velocity is smaller than for barotropic disturbances (especially for very long waves) because of the setting up of a kind of rocking motion (meridional motion alternating in direction along the wave). For present purposes we need not consider the details : the important point is that in their generally eastward progress disturbances are more retarded in low than in high latitudes. Now for a disturbance to maintain a constant shape it would have to move faster in low than in high latitudes because of the greater distance to be covered. Hence we should expect to find the axes of the ridges and troughs trailing away in a general NE.-SW. direction in the northern hemisphere and there is synoptic evidence that, on the average, this does happen. The argument may be made more precise. We find that theoretically maximum growth-rate corresponds to N.-S. troughs and ridges. Maximum *amplitude* must correspond to a later stage and therefore to disturbances in which the axes trail as described. Now it is easily verified, by substituting in the transfer formula, that disturbances the axes of whose troughs and ridges trail in this manner produce a poleward transfer of angular momentum (the geostrophic approximation may be applied). In the southern hemisphere we obtain the same final result. Hence this effect accounts for the observed direction of transfer of angular momentum over most of the earth's surface.

The second effect concerns behaviour in the extreme regions near the poles and near the equator. Both at the poles and at the equator the mean baroclinity is zero and a reasonably accurate model of the atmosphere would have maximum baroclinity (minimum Richardson number) in middle latitudes with decreased baroclinity towards the extremes. Analysis of such a model shows that disturbances should develop with maximum amplitude in middle latitudes and exponential decrease of amplitude towards the extremes. In the final stage we may expect to find that mixing extends into the extreme regions but rather as a process forced, through the operation of the constraints on motion, from outside. The energy transformations which drive the motion occur mainly in middle latitudes and in the extreme regions we may expect to find a quasi-barotropic driven turbulence. Now we know that in barotropic conditions (with horizontal motion) the vertical component of absolute vorticity is conserved and we have already considered the implications of this fact so far as transfer is concerned. We found that observed behaviour in the extreme regions was not inconsistent with the hypothesis that vorticity is conserved and the above argument suggests that there are theoretical reasons why such an hypothesis *ought* to give approximately correct results.

8. SURFACE FRICTION AND MERIDIONAL CIRCULATIONS

Since the amplitude of disturbances, both according to theory (Eady 1949, p. 48) and according to observation, increases on the average with height up to the tropopause, above which there is a rapid decrease corresponding to the greater static stability and smaller mean thermal gradient (larger Richardson number), most of the transfer of angular momentum (or vorticity) occurs in the upper troposphere. The sources and sinks are however, as we have noted, at ground level and to complete our analysis we have to show how the angular momentum transferred at high levels is brought down to ground level. A complete account would involve a discussion of the mechanism of small-scale turbulence in the friction (boundary) layer : for present purposes it will suffice to describe the main features. Although there are complications associated with small-scale convection this small-scale turbulence is fundamentally of the mechanically driven type, *i.e.*, the turbulent energy derives from the kinetic energy of the lowest layers. Associated with the development and maintenance of the turbulence there is, of course, a vertical transfer of momentum in such a direction as to destroy the kinetic energy of the boundary layer and were it not for the flow across the isobars towards low pressure the motion would eventually be destroyed. (In spite of complications similar to and perhaps more serious than those appearing on a large scale we can understand the maintenance of small-scale turbulence in terms of the continual growth of unstable disturbances, initially unstable waves of the Tollmien-Schlichting type or, especially over rough surfaces, of the Helmholtz-Rayleigh type developing on the associated velocity discontinuities.) The flow across the isobars in the boundary layer takes place automatically because the mean wind is less than geostrophic (as a *result* of turbulent mixing) so that so long as the mean flow is maintained the whole process will continue. Fundamentally the maintenance of the mean flow is by transfer of angular momentum between latitudes but we still have to show how the *whole column* of air contributes because we know (from the relation of isobars and wind in the free atmosphere) that small-scale turbulent transfer of angular momentum decreases considerably above the boundary layer. Now it follows immediately from the continuity equation that (at least in the long run) there must be a flow across the isobars above the boundary layer towards high pressure equal to the flow below towards low pressure. In other words the effect of surface friction is to generate meridional circulations. Logically we may consider the effect of such circulations apart from friction. Then we have conservation of angular momentum and it is easily seen that in the boundary layer the north-south flow is in such a direction that the transfer of angular momentum associated with meridional circulation compensates the loss associated with friction. Correspondingly, in the free atmosphere there is transfer of angular momentum in the opposite direction which in fact compensates the transfer of angular momentum due to large-scale turbulence. The net effect is that momentum brought in at high levels is transferred by the meridional circulation to the boundary layer for destruction by surface friction. In the long run the compensation is exact at each level because there is automatic adjustment of the mean motion until this is so (see Fig. 3).

The above argument simplifies the problem in so far as we have considered the effect of surface friction on the mean flow instead of the detailed large-scale turbulent flow but the more exact treatment yields only a quantitative difference in the final result. The greater the amplitude of large-scale turbulence the greater the effective coefficient of skin friction applicable to the mean flow : the direction of the cross isobaric flow is unaltered. (Not only does theory require this result but what actual computations have been made appear to verify it.) Hence in the zones of westerlies the lower cross-isobar current is directed polewards while in the zones of easterlies the lower cross-isobar current is directed equatorwards : in the free atmosphere the directions of the cross-isobar currents are reversed. Applying the continuity equation we find that there must be a general upward motion in the equatorial regions and in the low pressure regions near 60° latitude with a general

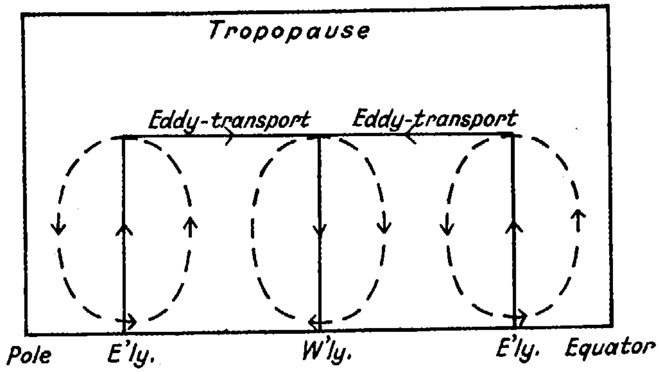

Figure 3. Transfer of angular momentum in the atmosphere. The dashed lines indicate frictionally driven meridional circulations which produce vertical transfer of angular momentum shown by full lines with arrows. The horizontal lines with arrows indicate transfer of angular momentum by large-scale turbulence.

downward motion in the region of the subtropical anticyclones and in the polar regions. In each hemisphere there are three meridional cells, two " direct " in low and high latitudes and one " reverse " in middle latitudes. It will be noted that *these* meridional circulations, in contrast to the hypothetical single direct-circulation considered earlier, are consistent with the observed zonal distribution of precipitation. Now both the direct and the reverse circulations are *driven*, they correspond to processes which absorb energy from outside, *i.e.*, from turbulent overturning. Only a fraction of the energy released by turbulent developments is utilised in this way and it is clear that if our theory is correct the rate of meridional overturning must be appreciably less than that of the hypothetical meridional circulation considered earlier, which was assumed to transfer heat at the same rate as that now attributed to turbulent transfer. Rough calculations based on an empirically determined coefficient of surface friction verify that this is so, the rate of circulation obtained being of the order of one-tenth that of the hypothetical circulation. The vertical velocities involved are now of a size with which observed deviations from radiation balance can cope, *e.g.*, the downward motion in the region of the subtropical anticyclones is so slow that the subsiding air is cooled by radiation fast enough to account for the observed lapse rate.

9. Further considerations

Our account of the mechanism of the " general circulation " of the atmosphere has been concerned only with the troposphere and the very lowest levels of the stratosphere. It is possible to consider events in this region independently of the rest of the atmosphere because the increase in (large-scale) Richardson number in the stratosphere reduces the amount of turbulent transfer and the tropopause acts in many ways like a lid. In the lowest layers of the stratosphere there is driven turbulent mixing as a result of the constraints expressed by continuity in the pressure field but at rather higher levels we should expect relative (though not complete) quiescence. At still higher levels the Richardson number probably again decreases and we should expect developments similar to, but not directly related to, those in the troposphere. An account of behaviour in these regions must await further empirical evidence. It seems unlikely that there is sufficient energy, either potential or kinetic, at very high levels for any *direct* major influence on tropospheric behaviour to be possible. It would however be erroneous to suppose that variations at very high levels (*i.e.*, well above the tropopause) can have no influence on weather. For example, an increase in stratospheric temperature, however slowly developed, would eventually lead to a lower tropopause : the boundary conditions applying to tropospheric developments would be altered, leading to disturbances of smaller size, a reduced rate of heat transfer and consequent modification in zonal characteristics. Most important of all there would result a change in the patterns associated with topography and the distribution of land and sea. Changes of phase (rather than changes in amplitude) of these patterns could be associated with very considerable changes in the climate of particular regions. A discussion of the manner in which these patterns are produced on a non-uniform earth is outside the scope of the present paper. It will suffice to point out that we can account for many features as the result of modifications in the zonal flow due to barriers (*e.g.*, mountain ranges like the Rockies and the ranges of E. Asia) and unevenness of frictional resistance (*e.g.*, owing to distribution of land and sea).

Finally, it should be pointed out that while the basic principles we have discussed are generally valid the application to zonal transfer is valid only for average behaviour (strictly that shown by annual mean charts, more roughly that shown by mean seasonal charts). Variations from mean behaviour are not only possible but are to be expected precisely because turbulent transfer is necessarily irregular.

The problem of the cause of the general circulation of the atmosphere may be regarded as the central problem of meteorology and a large part of meteorological literature bears directly or indirectly on it. Even to give a list of the principal works directly concerned with the problem would require greater acquaintance with the literature than the writer possesses. The following selection has been made therefore because these works are referred to in the text and not because they are, or are considered by the writer to be, the most important works. In addition to the references made earlier it should perhaps be emphasised that the *idea* of large-scale turbulence is not new — see for example Defant (1921). Also it may be noted that the principal features of the perturbation theory, on the results of which the present theory of the general circulation is based, are consistent with two independent investigations (Charney 1947, Fjortoft 1950).

REFERENCES

Bjerknes, V. and J., Solberg, H. and Bergeron, T.	1934	*Hydrodynamique Physique*, Paris.
Charney, J. G.	1947	*J. Met.*, Lancaster, Pa., **4**, p. 135.
Defant, A.	1921	*Geogr. Ann.*, Stockholm, **3**, p. 209.
Eady, E. T.	1949	*Tellus*, Stockholm, **1**, No. 3, p. 33.
Fjortoft, R.	1950	*Geofys. Publ.*, Oslo, **17**, No. 6.
Jeffreys, H.	1926	*Quart. J. R. Met. Soc.*, London, **52**, p. 97.
Simpson, G. C.	1928	*Mem. Roy. Met. Soc.*, London, **3**, No. 21.
Widger, W. K.	1949	*J. Met.*, Lancaster, Pa., **6**, p. 291.