PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Influence of Atmospheric Waves on the General Circulation of the Middle Atmosphere [and Discussion]

D. G. Andrews, F. W. Taylor and M. E. McIntyre

Phil. Trans. R. Soc. Lond. A 1987 **323**, 693-705
doi: 10.1098/rsta.1987.0115

Phil. Trans. R. Soc. Lond. A **323**, 693–705 (1987)
Printed in Great Britain

The influence of atmospheric waves on the general circulation of the middle atmosphere

By D. G. Andrews

Meteorological Office Unit, Hooke Institute, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, U.K.

To a first approximation, the basic features of the globally averaged structure of the middle atmosphere (such as the warm stratopause and cold mesopause) can be understood on radiative grounds alone. However, dynamical processes must be invoked if the observed latitudinally varying structures of the zonal-mean temperature and wind fields are to be explained. Particularly large departures from a hypothetical radiatively determined state occur in the winter stratosphere (especially in the Northern Hemisphere) and in the upper mesosphere at the solstices.

Simple theoretical models indicate that the primary dynamical mechanisms that drive the middle atmosphere away from radiative balance are wave motions, notably large-scale planetary waves and small-scale gravity waves. Much current research is being devoted to understanding the complex transient and irreversible processes by which such waves can influence the zonal-mean state and also lead to the meridional transport of chemical species.

1. Introduction

A fundamental goal of middle atmosphere research is the identification of the major physical and dynamical processes that control the main features of the observed general circulation of the region, together with an appreciation of the ways in which these processes interact with one another. It is difficult to see how such objectives could be achieved from observations alone: models of the atmosphere are essential to the task. One reason for this is that controlled experiments cannot be performed on the large-scale atmosphere: the nearest equivalent is to compare atmospheric states at different times, say, or in different hemispheres. Diagnostic studies of the atmosphere 'as it is' can seldom do more than confirm that certain balances (e.g. the zonal-mean zonal momentum budget) are analogous to those found in models. Terms, like the 'eddy momentum flux', in such balances have limited physical meaning on their own, and do not help us to understand, say, how the momentum budget would change if the eddies were reduced in amplitude for some reason. Even with a numerical atmospheric model including detailed treatments of dynamics, radiation and photochemistry (a proxy atmosphere on which controlled, hypothesis-testing experiments *can* be performed) the interaction between different processes is often so complex that clear chains of cause and effect may remain hard to establish.

In the face of these difficulties, a promising approach is to employ a hierarchy of middle atmosphere models, starting with simple analytical cases and progressing through more complicated 'mechanistic' examples to the highly complex general circulation models (GCMS). The various models in the hierarchy can be compared with each other and with the observed middle atmosphere, by using discriminating diagnostics, where available, to attempt to quantify their success in simulating atmospheric behaviour. A clear grasp of the properties of the simplest models aids the understanding of the more complex models; the latter will, in turn, help in interpreting the behaviour of the large GCMs and, by extension, the atmosphere itself. This approach should give a firm basis for judging whether a GCM is reproducing the observed general circulation in a satisfactory manner (and is doing so for the right reasons) and for correcting any major model deficiencies. One may then be able to decide whether the model is adequate for making predictions of future middle atmosphere behaviour, such as the response to perturbations like the injection of pollutant chemicals.

The term 'general circulation' has no precise meaning in meteorology. In this paper we shall adopt a very simple definition, namely the large-scale zonally averaged climatological state, where 'climatological' refers to time averages over individual calendar months in a multi-year record. Observational statistics on this state are available for the middle atmosphere from a number of sources, such as the proposed CIRA (Barnett & Corney 1985). We shall discuss the interplay of dynamical and radiative processes in the maintenance of this circulation, and especially the role of 'eddy' or 'wave' motions, here defined as departures from the zonally averaged state.

2. A radiative model of the middle atmosphere

A suitable starting point in our hierarchy of models, at least for addressing the question of the relative roles of radiation and dynamics in determining the middle atmosphere circulation, is a model in which dynamics plays no active part at all. Thus a zonally symmetric temperature field $T_r(\phi, z, t)$ (where ϕ is latitude, z is logarithmic-pressure altitude and t is time) is calculated from a radiative balance at each point between heating due to absorption of solar ultraviolet radiation (primarily by ozone) and cooling and heating due to emission and absorption of thermal infrared radiation by various gases (mainly carbon dioxide, ozone and water vapour). The ultraviolet heating depends largely on the ozone distribution and the position of the Sun, whereas the infrared transfer is strongly dependent on temperature.

Further ingredients required for this calculation are suitable radiative conditions at the lower boundary: one can, for example, specify the upward infrared flux from the troposphere and the tropospheric albedo. Time dependence can be included by allowing for the atmosphere's thermal inertia: this is important in the polar night, where radiative relaxation rates can be so small that the atmosphere does not have time to reach equilibrium before the Sun reappears. There are various ways of specifying the ozone distribution: it can for instance be allowed to seek a state of photochemical equilibrium. A somewhat simpler expedient is to use the ozone fields from an observed climatology: the results of the calculation are probably easier to interpret in this case, although the effects of dynamics are then implicitly included in the ozone distribution.

An example of a 'radiatively determined' temperature field, and a comparison with observations, was presented by Fels (1985). A further example, from my colleague Dr K. P. Shine, is shown in figures 1 and 2. The T_r fields in these figures were calculated from a radiation scheme designed for a collaborative middle atmosphere modelling project between the Department of Atmospheric Physics at Oxford University and the Middle Atmosphere group at the U.K. Meteorological Office: full details are given in Shine (1987). Time dependence was included and the ozone field was prescribed from climatological observations: it varies with latitude, height and season.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

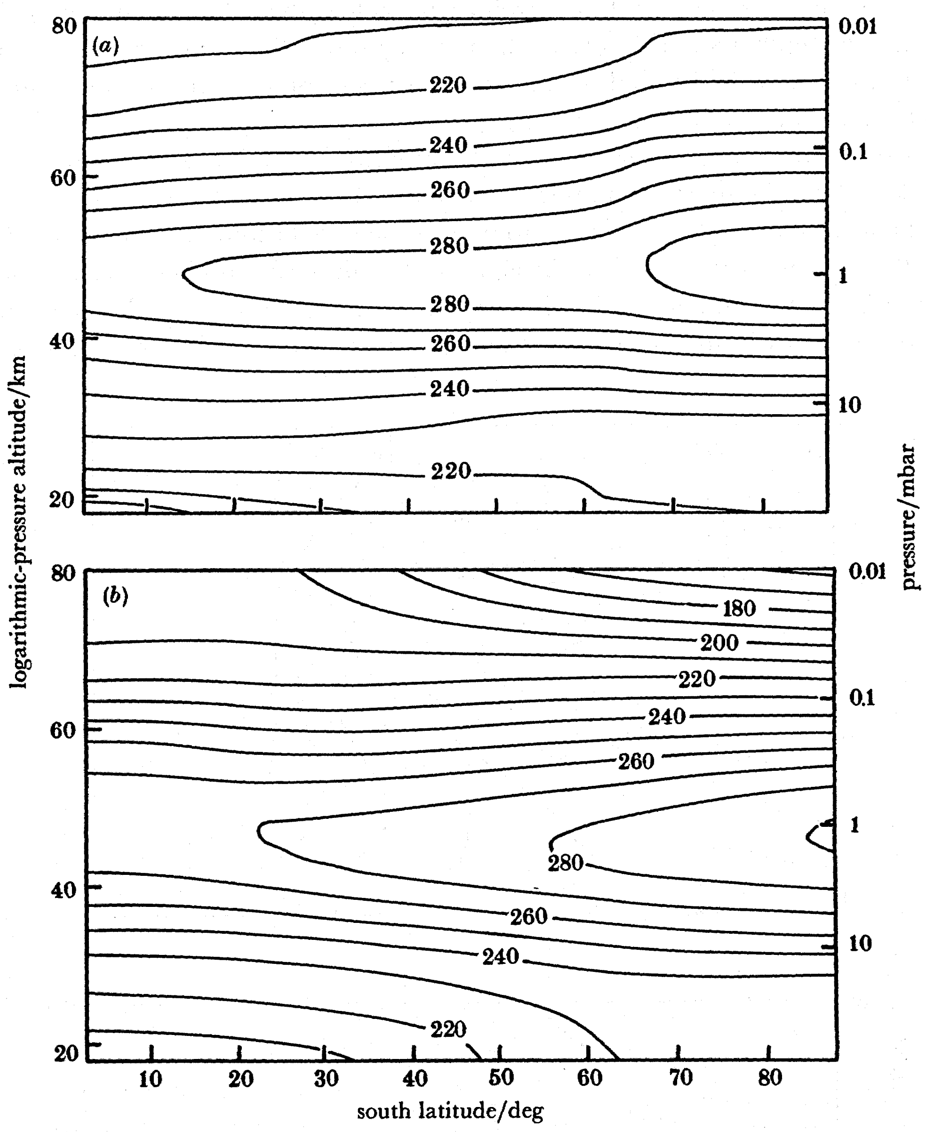

FIGURE 1. (*a*) Radiatively determined temperatures (K) for the Southern Hemisphere middle atmosphere on 21 December (near summer solstice). (*b*) Observed monthly-mean, zonal-mean temperatures (K) for the Southern Hemisphere middle atmosphere in December, interpolated from the climatology of Barnett & Corney (1985). (Redrawn, with some smoothing of contours in (*a*), from Shine 1987.)

Looking first at the summer solstice (December, Southern Hemisphere) we see that the radiatively determined temperature field (figure 1*a*) is broadly similar to the observed climatology (figure 1*b*) throughout much of the middle atmosphere. Both have a well-marked stratopause near 1 mbar† (about 50 km altitude) with a maximum temperature of just above 290 K at the pole (the observed temperature being slightly cooler than T_r there), and temperatures decreasing monotonically above and below. The major difference between the two occurs in the upper mesosphere: whereas the equatorial temperatures agree fairly well there, the latitudinal gradients are opposite in sign. As a result the observed temperature at about

† 1 mbar = 10^2 Pa.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

80 km over the summer pole is 160 K, considerably colder than the corresponding radiatively determined value of 220 K. Such a large discrepancy shows that some process (presumably of dynamical origin) that is essential to the observed upper mesosphere has been omitted from the model calculation leading to figure 1*a*.

Even greater differences occur between the radiatively determined and observed winter solstice fields (June, Southern Hemisphere) depicted in figure 2. In the radiative calculation of figure 2*a*, a clear stratopause extends from low to middle latitudes, to be replaced poleward of about 60° by a core of very cold air throughout the middle atmosphere (whose temperature, in fact, continues to drop until the Sun reappears in spring). The observed climatology (figure 2*b*), although similar in the subtropical stratosphere and lower mesosphere, is very

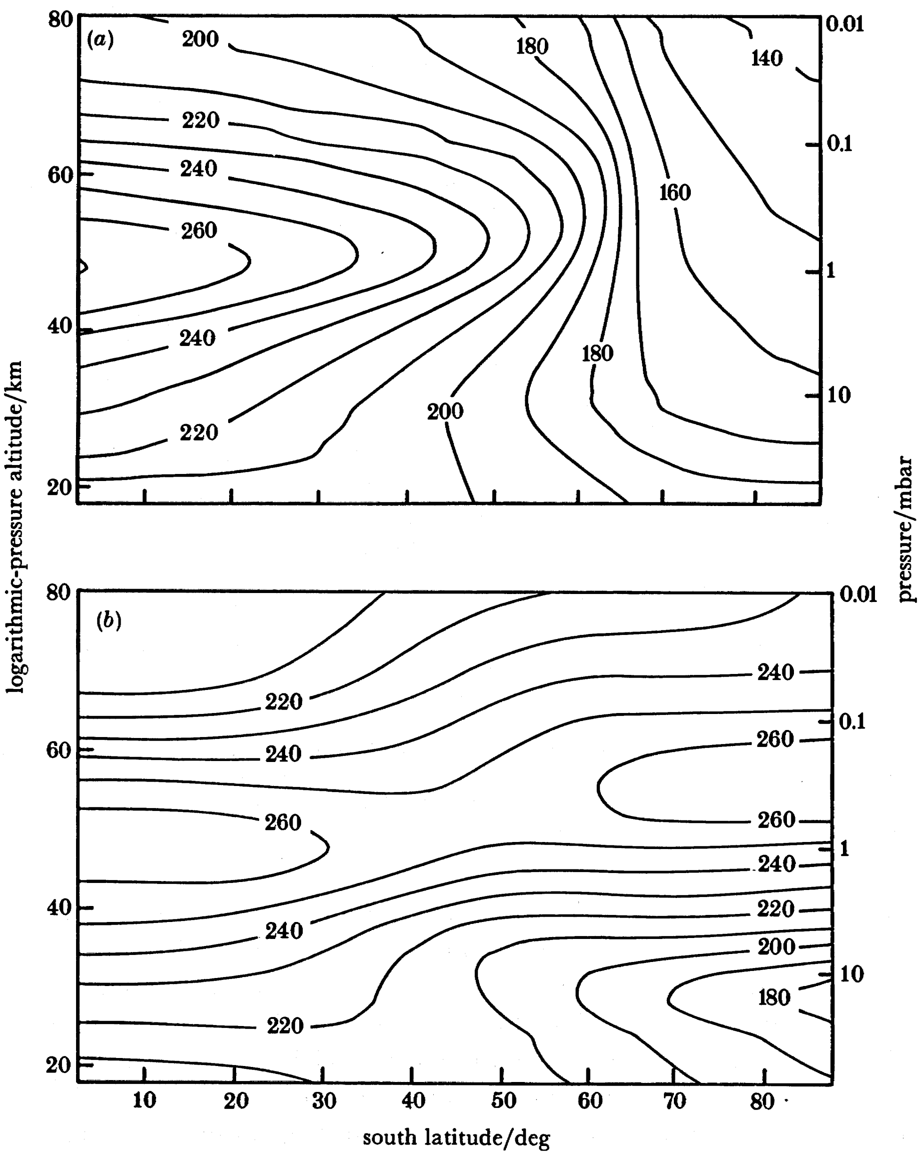

FIGURE 2. (*a*) As figure 1*a*, but for Southern Hemisphere middle atmosphere on 19 June (near winter solstice). (*b*) As figure 1*b*, but for Southern Hemisphere middle atmosphere in June. (Redrawn, with some smoothing of contours in (*a*), from Shine 1987.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

different in the upper mesosphere and the polar regions. The reversed latitudinal temperature gradient observed in the summer upper mesosphere (figure 1 *b*) continues across into the winter hemisphere as well, with warmest temperatures occurring over the winter pole; near 80 km the latter is at about 230 K, some 100 K above the radiative value. A polar stratopause is observed near 0.3 mbar (57 km), and the whole polar middle atmosphere is much warmer than the corresponding radiatively determined field, with minimum temperatures of just below 180 K appearing at about 20 mbar, or 28 km, in the lower stratosphere. (The northern winter lower stratosphere is even warmer.) A dynamical explanation seems to be required for the differences between figure 2 *a*, *b* not only in the upper mesosphere, as in the summer, but also throughout the winter polar stratosphere and mesosphere (with the possible exception of the Antarctic lower stratosphere in winter).

With the temperature fields of figures 1 and 2 and a little dynamics, in the form of hydrostatic and gradient wind balance, one can calculate the corresponding zonal wind fields, using observed winds as a lower boundary condition. These again differ markedly: in particular, the winter solstice 'radiatively determined' westerly winds increase monotonically with height, reaching 300 m s^{-1} at the mesopause near 65° latitude, whereas the winter winds determined from observations peak in westerly jets of about 120 m s^{-1} near 60 km altitude and 40° latitude, and decrease above that level. Gradient wind balance cannot be obeyed for the summer mesospheric T_r field of figure 1 *a*, owing to the strong poleward increase between 60° and 70°. Further details are given by Shine (1987).

We now move on to the next example in our hierarchy of models, and turn to consider the dynamical processes that may help to account for the differences between the radiatively determined state and the observed climatological state of the middle atmosphere.

3. Dynamical processes

The simplest middle atmosphere model with 'active' dynamics, capable of transporting heat and chemical constituents meridionally, is one in which a zonally symmetric circulation with north-south and vertical components is allowed in addition to any radiative processes. Simple scaling arguments, based on quasigeostrophic theory (see, for example, WMO 1986, §6.2.2), suggest that such a circulation is likely to be too weak to account for the observed reversed temperature gradient in the upper mesosphere and the observed warm winter stratosphere. Although there may exist special ageostrophic flows for which such scaling arguments are not valid (see Tung 1986), they have been found to hold quite well in detailed calculations by my colleague Dr D. J. Carruthers, with a zonally symmetric version of the Meteorological Office–Oxford University middle atmosphere GCM incorporating the radiation scheme of Shine (1987). This model cannot reproduce the upper mesospheric latitudinal temperature gradient or the warm winter stratosphere.

It follows that departures from zonal symmetry ('eddies' or 'waves') are essential for maintaining the deviations of the observed climatology of figures 1 *b* and 2 *b* from the radiatively determined states of figures 1 *a* and 2 *a*, at least in the upper mesosphere and winter stratosphere and lower mesosphere. (This fact first appears to have been appreciated by Dickinson (1969).) A convenient way of investigating the effect of zonal asymmetries on the zonal-mean state is provided by the 'transformed eulerian-mean formulation' of the mean-flow equations (see, for example, Andrews & McIntyre 1976, 1978; Edmon *et al.* 1980; Dunkerton *et al.* 1981).

Under quasigeostrophic scaling, the influence of the eddies can be summarized by an effective zonal force per unit mass G acting on the zonal-mean flow, with no corresponding heating term being present. The contribution of large-scale eddies to G equals $\rho_0^{-1}\nabla\cdot\boldsymbol{F}$, where $\boldsymbol{F}$ is the Eliassen–Palm flux vector, a mean quadratic function of eddy quantities, $\rho_0(z)$ is the basic density in logarithmic-pressure coordinates (proportional to pressure) and the usual cartesian β-plane geometry is used instead of spherical coordinates. Contributions by small-scale eddies can also be incorporated in G, and will be represented by X here. It can be shown that the relevant small-scale waves, like the quasigeostrophic eddies, introduce no heating term into the zonal-mean thermodynamic equation, although they may bring about weak vertical diffusion of heat. The latter effect will be ignored here (but see the discussion in §5).

In a steady state, the transformed eulerian-mean zonal momentum, thermodynamic and continuity equations for this model atmosphere are

$$-f_0\bar{v}^* = G \equiv \rho_0^{-1}\nabla\cdot\boldsymbol{F}+X, \tag{1}$$

$$S\bar{w}^* = \bar{J}/c_{\mathrm{p}}, \tag{2}$$

$$\frac{\partial\bar{v}^*}{\partial y}+\rho_0^{-1}\frac{\partial(\rho_0\bar{w}^*)}{\partial z} = 0. \tag{3}$$

Here y is northward distance and the vertical coordinate z (linear in logarithmic pressure) is approximately equal to geometric altitude, $(\bar{v}^*, \bar{w}^*)$ are the corresponding northward and 'vertical' velocity components, f_0 is a mean Coriolis parameter, $S(z)$ is a measure of static stability, $\bar{J}$ is the net heating rate per unit mass (here equal to the net radiative heating rate), and c_{p} is the specific heat at constant pressure. For further explanation of these equations see, for example, Dunkerton *et al.* (1981).

Looking first at the zonal momentum equation (1), we see that in this steady model the force G must be balanced by the Coriolis acceleration associated with a northward velocity $\bar{v}^*$. Through the continuity equation (3), $\bar{v}^*$ will generally imply non-zero vertical velocities $\bar{w}^*$, whose thermodynamic effects (the 'adiabatic heating' $S\bar{w}^*$) must be balanced by a nonzero $\bar{J}$, by (2). Thus the presence of a non-zero force G drives a circulation $(\bar{v}^*, \bar{w}^*)$ that in turn forces a non-zero net heating rate. This circulation is sometimes called a 'diabatic circulation' on account of the balance expressed by (2). When $G = 0$, the diabatic circulation vanishes and $\bar{J} = 0$, implying a purely radiative balance, with the temperature equal to its (steady-state) radiatively determined value everywhere.

This very simple model can be extended a little by including time dependence: then $\partial\bar{u}/\partial t$ is added to the left of (1) and $\partial\bar{T}/\partial t$ to the left of (2), where $\bar{u}$ and $\bar{T}$ are the zonal-mean zonal wind and temperature, respectively. The thermal wind equation, expressing proportionality of $\partial\bar{u}/\partial z$ and $\partial\bar{T}/\partial y$, must be incorporated as well. It is convenient also to use a simple parametrization of $\bar{J}$ in terms of $\bar{T}$, such as the newtonian cooling form

$$\bar{J}/c_{\mathrm{p}} = -(\bar{T}-T_{\mathrm{r}})/\tau_{\mathrm{r}}, \tag{4}$$

where $T_{\mathrm{r}}(y,z,t)$ is the radiatively determined temperature and $\tau_{\mathrm{r}}(z)$ is a radiative relaxation time. Scaling arguments applied to the resulting equations show that if τ_{r} is much shorter than seasonal and relevant dynamical timescales (as is usually the case) then the steady-state balance described above holds quite accurately at each point in time (see, for example, WMO

1986, §6.2.2.). Again the diabatic circulation $(\bar{v}^*, \bar{w}^*)$, the net diabatic heating $\bar{J}$ and, by (4), the departures of $\bar{T}$ from T_r are caused by the imposed force G. One can think of the dynamics as 'pushing' against the 'radiative spring' expressed by (4) to give the deviations from radiative balance (Fels 1985). This model makes it particularly clear that one cannot regard $\bar{J}$ as an externally imposed heating that *drives* the diabatic circulation.

Having seen how a mean zonal force per unit mass G associated with eddies can maintain departures of the observed $\bar{T}$ from the radiatively determined T_r, it is of interest for us to enquire whether G can be estimated from observations. One possible approach is to start with an estimate of the diabatic circulation, following the method of Murgatroyd & Singleton (1961) and many subsequent authors (a recent example being that of Solomon *et al.* 1986). Given observed $\bar{T}$ and ozone fields one calculates $\bar{J}$ as described in §2 and then estimates the $(\bar{v}^*, \bar{w}^*)$ field that satisfies the time-dependent, spherical, primitive-equation versions of (2) and (3) and suitable boundary conditions; no eddy terms are included in (2). Some minor adjustments to the velocities may be required to ensure global mass conservation. The general picture at solstice is of rising motion in the summer upper stratosphere and mesosphere, cross-equatorial flow in the mesosphere and descent in the winter middle atmosphere. In the lower stratosphere this merges with a 'Hadley' circulation of equatorial ascent and extratropical descent in both hemispheres (figure 3).

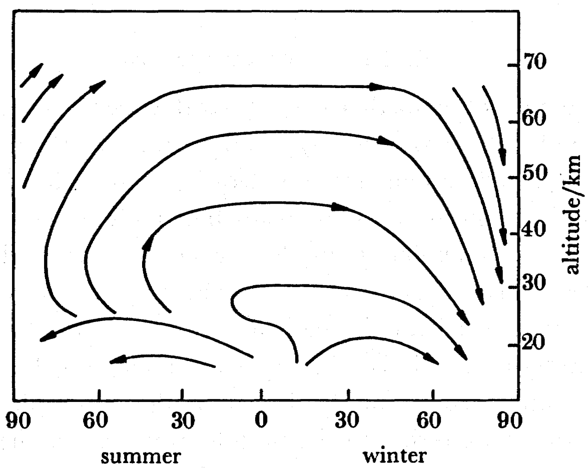

FIGURE 3. Schematic streamlines of the solstice diabatic circulation in the middle atmosphere. (Adapted from Dunkerton 1978.)

Given the diabatic circulation and estimates of $\bar{u}$ and its derivatives, one may use the time-dependent, spherical, primitive-equation version of the zonal-mean momentum equation (1) to estimate G. A preliminary calculation by Dr K. P. Shine indicates that values of tens of metres per second per day (generally opposite in sign to $\bar{u}$) are required in the mesosphere and a few metres per second per day (generally westward) are needed in the winter stratosphere.

An additional step is to subtract from G the part associated with $\nabla \cdot \boldsymbol{F}$ due to large-scale waves resolved by the data, leaving the contribution due to the unresolved waves. Calculations of the latter quantity by essentially this method were performed by Hamilton (1983) for the stratosphere and by Smith & Lyjak (1985) for the stratosphere and lower mesosphere.

We now investigate the physical basis of the force G: for this purpose we recall some theoretical results. It was shown by Eliassen & Palm (1961) that $\nabla \cdot \boldsymbol{F}$ is identically zero for steady, conservative, linear waves on a purely zonal basic shear flow. This 'Eliassen–Palm theorem' has been extended in various ways, to explore how $\nabla \cdot \boldsymbol{F}$ differs from zero under more general circumstances (see, for example, Andrews & McIntyre 1976, 1978; Boyd 1976; Edmon *et al.* 1980; Andrews 1983, 1987; Killworth & McIntyre 1985). The importance of the theorem in the present context is that it focuses attention on those physical properties of the large-scale waves that can lead to a non-zero $\nabla \cdot \boldsymbol{F}$, and hence G, namely wave transience and non-conservative wave processes such as thermal or frictional damping. Similar considerations hold for the contribution X due to small-scale waves. Other things being equal, we can expect such effects to grow in magnitude as the wave amplitude increases.

We next consider which types of wave might be responsible for the observed departures from a radiatively determined state or, equivalently, for the force G. Thinking first of the winter stratosphere, a notable feature of this region, particularly in the Northern Hemisphere, is the presence of large-amplitude planetary waves, in which the flow can become highly zonally asymmetric. In the upper mesosphere, such planetary waves appear to be less significant; however, this is a region where much small-scale gravity wave activity is revealed by ground-based measurements with radars and lidars. Global-scale atmospheric tides are expected to achieve large amplitudes here, too. These waves are discussed in more detail in the following sections.

4. Planetary waves

It has long been known that the northern winter stratospheric flow is generally characterized (especially in the second half of winter) by strong, quasi-stationary departures from zonal symmetry, usually interpreted as planetary, or Rossby, waves. Modern satellite observations give us detailed, global information on these disturbances. The theory of linear Rossby waves is well developed, and seems to explain many features of the observed waves. In particular the classic work of Charney & Drazin (1961) shows that only the largest-scale stationary linear Rossby waves can propagate vertically in realistic westerly basic winds and that no such waves can penetrate into easterly basic winds. This appears to explain the fact that only large-scale quasi-stationary disturbances (presumably forced in the troposphere) are observed in the northern hemisphere winter westerlies in the middle atmosphere and that no planetary-scale disturbances are observed in the summer easterlies. Linear theory also seems to account quite well for aspects of transient planetary-wave events, including some features of major and minor stratospheric sudden warmings.

In view of the transient nature of some of the observed planetary waves and their inevitable damping by thermal dissipation, one would expect them to be associated with non-zero values of $\nabla \cdot \boldsymbol{F}$, even when they are of small amplitude. But many planetary waves can in fact attain large amplitudes, and may be poorly represented by linear theory. Such events include cases, named 'breaking planetary waves' by McIntyre & Palmer (1983, 1984), in which air masses can be pulled rapidly and presumably irreversibly into long, thin streamers. (These can be seen fairly clearly in the isentropic maps of Ertel's potential vorticity, a quasi-conservative dynamical tracer, and isobaric maps of minor constituents that have been derived from satellite data.) Particularly large values of $\nabla \cdot \boldsymbol{F}$ are expected during these periods of transient, finite-amplitude distortion of the flow.

Observational estimates of the zonal force per unit mass associated with $\nabla \cdot \boldsymbol{F}$ due to large-scale planetary waves in the lower winter stratosphere (see, for example, Hamilton 1982; Geller *et al.* 1983) tend to be consistent in order of magnitude (a few metres per second per day) and sign (generally negative, i.e. westward) with Dr Shine's diagnostic estimate of G for that region, mentioned in §3. (A discrepancy in sign appears in high latitudes; however, as shown by Robinson (1986), the quasigeostrophic approximation to $\nabla \cdot \boldsymbol{F}$ used by Hamilton and Geller *et al.* is likely to be unreliable there.)

The foregoing discussion makes it plausible that the planetary waves observed in the northern winter stratosphere may indeed be the main agents responsible for driving the observed climatology there away from a radiatively determined state. This is qualitatively consistent with the fact that the Southern Hemisphere winter, which contains weaker planetary-wave activity than the northern winter, is closer to the radiatively determined state in the lower and middle stratosphere. It is also consistent with certain GCM experiments (see, for example, Mahlman & Umscheid 1984; Fels 1985) that are found to have excessively cold temperatures and excessively strong westerly winds in the winter stratosphere, and at the same time are known to be deficient in stratospheric planetary-wave activity (see also Boville 1985).

There is clearly scope in this area for further experiments with GCMs to test hypotheses concerning the effects of planetary waves on the general circulation of the middle atmosphere; the theoretical framework described above has already helped in the design and interpretation of such experiments. A related need is for detailed numerical and observational studies to explore the 'planetary-wave breaking' phenomenon, to establish how irreversible it really is and to assess the relative roles of the radiative and mixing processes associated with it. The phenomenon is also clearly relevant to the meridional transport of chemical species such as ozone (Leovy *et al.* 1985), thus perhaps influencing the observed temperature and wind climatology in quite subtle ways, by affecting the solar heating rates; these questions deserve further study, too.

5. Gravity waves and tides

We now consider possible causes of the large departures of observed climatologies from radiative balance in the upper mesosphere. As mentioned in §3, this is a region of relatively weak planetary-wave activity but strong gravity-wave activity. The gravity waves have horizontal wavelengths of up to 100–200 km, vertical wavelengths of 5–15 km and periods of minutes to hours. They are believed to be generated mostly in the troposphere or lower stratosphere, whence they propagate upwards into the rarefied mesosphere. As they do so their velocity and temperature amplitudes increase roughly exponentially with height, and isentropic surfaces become more and more distorted. Eventually, at great enough altitudes, the isentropes overturn, rather like breaking water waves on a shelving beach, leading to turbulence, mixing and dissipation.

This gravity-wave breaking phenomenon is complex and nonlinear, and is far from being well understood. However, it is likely to inhibit the upward growth in amplitude of the gravity waves, thus altering the mean upward transfer of horizontal momentum that is associated with them and hence exerting a drag on the mean flow. This drag contributes to the zonal force per unit mass G; the turbulence produced by the breaking process is likely to diffuse momentum in the vertical, thus providing another contribution to G.

Some simple parameterizations of the effects of breaking gravity waves on the large-scale flow, based on extensions of linear gravity-wave theory, have been proposed by Lindzen (1981) and Matsuno (1982) and tested in idealized middle-atmosphere models by Holton (1982, 1983), Miyahara (1984) and others. These studies suggest that the gravity-wave drag and diffusion terms are of roughly the right order of magnitude (tens of metres per second per day) and sign in the upper mesosphere to provide the force per unit mass on the zonal flow needed to account for the strong departures from radiative balance evidenced by the reversed temperature gradient near the mesopause.

Much more work on the fundamental dynamics of breaking gravity waves will be needed to fill out this picture in detail; this will presumably include high-resolution numerical simulations and perhaps laboratory experiments. Among other things, a clear understanding is required of the vertical mixing of heat and chemical tracers, as well as of momentum, by breaking gravity waves. There is some evidence that this process may be responsible for the vertical transport of chemicals in the upper mesosphere (Garcia & Solomon 1985), although some authors believe that the vertical diffusivity for heat and long-lived tracers may be much less than that for momentum (Chao & Schoeberl 1984; Fritts & Dunkerton 1985). There is also a need for further quantitative measurements of gravity waves in the mesosphere. Some pioneering observations in Australia by Vincent & Reid (1983) have provided estimates of the vertical momentum flux convergence associated with breaking gravity waves. A large, worldwide dataset for this type of quantity would be most valuable.

Evidence also exists that drag due to orographically induced gravity waves may be important in the troposphere and lower stratosphere: inclusion of simple parameterizations of this effect appears to bring significant improvements to some numerical weather forecasts and climatological GCM simulations (Palmer *et al.* 1986). In the winter polar stratosphere it is possible that gravity waves, in addition to the planetary waves discussed in §4, may contribute to the climatological forcing that drives $\overline{T}$ away from T_r. More indirectly, gravity waves may damp or force the planetary waves, thus influencing the values of $\nabla \cdot \boldsymbol{F}$ associated with the latter (Holton 1984; Miyahara 1985).

A further class of atmospheric waves that may contribute significantly to the forcing of the climatological mean state away from radiative balance are the atmospheric tides. The linear theory of such waves is well known, but their nonlinear effects, which are expected to be important in the rarefied mesosphere, are less clearly understood. However, it is likely that these disturbances too will 'break' in the mesosphere, and contribute to G there (Lindzen 1968, 1981; Miyahara 1984). Once again, more study of these nonlinear processes is required.

6. Conclusions

This paper has shown how even the first steps in a hierarchy of models of the middle atmosphere, including a detailed treatment of radiation but simple dynamics, can yield qualitative insights into the interplay of the physical and dynamical phenomena that control the observed zonal-mean general circulation of the region. In particular, certain wave motions (which are thought to be mainly generated in the troposphere) have been identified as prime agents for driving the middle atmospheric state away from a purely radiatively determined structure. Furthermore, results from the theory of wave, mean-flow interaction have shown how specific properties of the waves are important in this process.

The next stages in the hierarchy involve more quantitative modelling of the relevant waves,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

including their complex behaviour at finite amplitude. Full general circulation models should in principle be able to simulate the large-scale aspects of planetary waves quite well; in practice, however, as mentioned in §4, they frequently seem to underestimate the corresponding mean zonal force G, leading among other things to a northern winter stratosphere that is too 'radiative'. The reasons for this are not entirely clear, but may be partly associated with an incorrect treatment by the model of the dissipation of small-scale potential vorticity structures produced by the 'breaking' of the planetary waves. Experiments with high-resolution idealized models, as well as with GCMS, should aid in the understanding of these issues. As for the drag and diffusion associated with gravity waves, detailed studies aimed at testing and improving parameterizations for global models are clearly needed in addition to the fundamental studies mentioned in §5. Recent theoretical results on the interaction of finite-amplitude waves with mean flows may perhaps assist in these investigations.

Other stages in the model hierarchy will need to focus on the complex feedbacks that may arise in the more detailed GCMS and the real middle atmosphere. For example, one should not think of the forcing G as totally independent of the mean flow on which the waves are superimposed. A question that has so far received little attention in any but the simplest models is the possible sensitivity of the planetary wave structure, and hence $\nabla \cdot \boldsymbol{F}$, to the zonal mean-flow configuration. Fels (1985) presents some interesting speculations as to how this might couple with radiative effects to give strong feedbacks. A different class of feedbacks, associated for example with the transport, photochemistry and radiative effects of ozone could also be of considerable importance to the climate of the middle atmosphere. These, too, will need to be studied with a range of models of varying sophistication.

Progress on all these fronts will be essential to improving our understanding of GCMS and to helping us correct their deficiencies, thus leading to improved simulations of the current general circulation. This in turn should improve our understanding of the maintenance of the present structure of the middle atmosphere. It should also give us greater confidence in the use of complex models for experiments to study the future climatic response to perturbations, whether natural or manmade.

Several of the basic ideas presented here were originally explained to me by S. B. Fels; they have been further developed in the course of discussions with K. P. Shine and D. J. Carruthers, to whom I am indebted for permission to quote unpublished results. I am also grateful to K. P. Shine and M. E. McIntyre for helpful comments on the first draft of this paper.

References

Andrews, D. G. 1983 *J. atmos. Sci.* **40**, 1877–1883.
Andrews, D. G. 1987 *Q. Jl R. met. Soc.* **113**, 323–338.
Andrews, D. G. & McIntyre, M. E. 1976 *J. atmos. Sci.* **33**, 2031–2048.
Andrews, D. G. & McIntyre, M. E. 1978 *J. atmos. Sci.* **35**, 175–185.
Barnett, J. J. & Corney, M. 1985 In *Middle atmosphere program. Handbook for MAP*, vol. 16, pp. 47–85. Urbana, Illinois: SCOSTEP Secretariat.
Boville, B. A. 1985 *J. atmos. Sci.* **42**, 904–916.
Boyd, J. P. 1976 *J. atmos. Sci.* **33**, 2285–2291.
Chao, W. C. & Schoeberl, M. R. 1984 *J. atmos. Sci.* **41**, 1893–1898.
Charney, J. G. & Drazin, P. G. 1961 *J. geophys. Res.* **66**, 83–109.
Dickinson, R. E. 1969 *J. atmos. Sci.* **26**, 73–81.
Dunkerton, T. J. 1978 *J. atmos. Sci.* **35**, 2325–2333.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Dunkerton, T. J., Hsu, C.-P. F. & McIntyre, M. E. 1981 *J. atmos. Sci.* **38**, 819–843.
Edmon, H. J. Jr, Hoskins, B. J. & McIntyre, M. E. 1980 *J. atmos. Sci.* **37**, 2600–2616 (also corrigendum 1981 *J. atmos. Sci.* **38**, 1115).
Eliassen, A. & Palm, E. 1961 *Geofys. Publr* **22**(3), 1–23.
Fels, S. B. 1985 *Adv. Geophys.* **28** A, 277–300.
Fritts, D. C. & Dunkerton, T. J. 1985 *J. atmos. Sci.* **42**, 549–556.
Garcia, R. R. & Solomon, S. 1985 *J. geophys. Res.* **90**, 3850–3868.
Geller, M. A., Wu, M.-F. & Gelman, M. E. 1983 *J. atmos. Sci.* **40**, 1334–1352.
Hamilton, K. 1982 *J. atmos. Sci.* **39**, 2737–2749.
Hamilton, K. 1983 *Mon. Weath. Rev.* **111**, 1434–1441.
Holton, J. R. 1982 *J. atmos. Sci.* **39**, 791–799.
Holton, J. R. 1983 *J. atmos. Sci.* **40**, 2497–2507.
Holton, J. R. 1984 *J. atmos. Sci.* **41**, 3427–3430.
Killworth, P. D. & McIntyre, M. E. 1985 *J. Fluid Mech.* **161**, 449–492.
Leovy, C. B., Sun, C.-R., Hitchman, M. H., Remsberg, E. E., Russell, J. M. III, Gordley, L. L., Gille, J. C. & Lyjak, L. V. 1985 *J. atmos. Sci.* **42**, 230–244.
Lindzen, R. S. 1968 *Proc. R. Soc. Lond.* A **303**, 299–316.
Lindzen, R. S. 1981 *J. geophys. Res.* **86**, 9707–9714.
McIntyre, M. E. & Palmer, T. N. 1983 *Nature, Lond.* **305**, 593–600.
McIntyre, M. E. & Palmer, T. N. 1984 *J. atmos. terrest. Phys.* **46**, 825–849.
Mahlman, J. D. & Umscheid, L. J. 1984 In *Dynamics of the middle atmosphere* (ed. J. R. Holton & T. Matsuno), pp. 501–525. Tokyo: Terrapub.
Matsuno, T. 1982 *J. met. Soc. Japan* **60**, 215–226.
Miyahara, S. 1984 In *Dynamics of the middle atmosphere* (ed. J. R. Holton & T. Matsuno), pp. 271–287. Tokyo: Terrapub.
Miyahara, S. 1985 *J. atmos. Sci.* **42**, 100–107.
Murgatroyd, R. J. & Singleton, F. 1961 *Q. Jl R. met. Soc.* **87**, 125–135.
Palmer, T. N., Shutts, G. J. & Swinbank, R. 1986 *Q. Jl R. met. Soc.* **112**, 1001–1039.
Robinson, W. A. 1986 *J. atmos. Sci.* **43**, 1017–1023.
Shine, K. P. 1987 *Q. Jl R. met. Soc.* **113**, 605–635.
Smith, A. K. & Lyjak, L. V. 1985 *J. geophys. Res.* **90**, 2233–2241.
Solomon, S., Kiehl, J. T., Garcia, R. R. & Grose, W. L. 1986 *J. atmos. Sci.* **43**, 1603–1617.
Tung, K. K. 1986 *J. atmos. Sci.* **43**, 2600–2618.
Vincent, R. A. & Reid, I. M. 1983 *J. atmos. Sci.* **40**, 1321–1333.
WMO 1986 World Meteorological Organization: Global Ozone Research and Monitoring Project, report no. 16. Atmospheric Ozone 1985.

Discussion

F. W. Taylor (*Department of Atmospheric Physics, Clarendon Laboratory, University of Oxford, U.K.*). Dr Andrews showed a diagram of the diabatic or residual circulation at solstice. Could he describe how this varies with season during the rest of the year?

D. G. Andrews. The diabatic circulation at the equinoxes appears to be rather symmetric, with rising motion in low latitudes throughout the stratosphere and mesosphere, and descent in high latitudes in both hemispheres. A smooth transition between the two régimes occurs at other times of year.

M. E. McIntyre (*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, U.K.*). Along with many others I find this view of the general circulation of the middle atmosphere, in which one thinks of tropospherically driven eddies as holding the mean circulation away from its radiatively determined temperature T_r, very useful and insightful. I would add four comments.

1. In its simplest form, the picture depends on regarding T_r as given, and so leaves out of account the possible feedback of the circulation onto T_r itself. The interesting negative feedback mechanism noted by Farman *et al.* (1985) between T_r, diabatic descent in the Antarctic

summer stratosphere, and ozone photochemistry, illustrates one possibility that might be important. This gives added weight to Dr Andrews's concluding remarks.

2. It is perhaps worth re-emphasizing that for the purpose of wave, mean-flow interaction theory the idea of wave 'transience' is defined, in its most general form, in lagrangian terms. 'Transience' in this context means time dependence of material displacement amplitudes. Time dependence of, say, the geopotential amplitude of a Rossby wave, is not directly relevant. That is why it makes sense, for instance, to define the idea of wave 'breaking', for gravity as well as for Rossby waves, in terms of the behaviour of certain material contours; the distinction between breaking and non-breaking waves then becomes directly relevant to wave, mean-flow interaction theory. The relevant material contours are those that would undulate reversibly, under the influence of the waves' restoring mechanism, in those circumstances for which linearized, non-dissipative wave theory is a self-consistent approximation to nonlinear reality. Further discussion can be found in a recent paper by McIntyre & Palmer (1985). A breaking wave is one that causes some or all of the relevant material contours to deform irreversibly, implying that material displacements do not return to zero and that the effects of wave 'transience' on the mean state can become relatively permanent.

3. As remarked earlier in connection with Dr O'Neill's paper, no one fully understands the nature of the tropospheric forcing function, which is imagined to be the dynamical source of eddy motion in the middle atmosphere. There are, to be sure, obvious topographic sources of gravity and planetary waves; but there is also reason to believe that internal generation by nonlinear fluid-dynamical processes is involved (somewhat reminiscent of Lighthill's theory of aerodynamical sound generation), as well as overtly thermal sources, all of them presumably related to geography. I think this poses one of the most subtle challenges to dynamical researchers trying to understand the vertical coupling between troposphere and middle atmosphere.

4. The not inconsiderable success of quasi-two-dimensional chemical models prompts one to wonder whether we have yet got the best mileage out of them. For instance, as the Antarctic ozone-hole problem has reminded us, averaging around latitude circles is unlikely to be the optimal way of validating such a model against observational data. It would presumably be better to interpret the models' predictions, and further develop the models themselves, in terms of concepts such as 'inside the main vortex', 'outside the main vortex', etc., where 'main vortex' is understood in the lagrangian sense already alluded to in my comments on Professor Bowhill's contribution. Some of the relevant ideas have been discussed most recently by Butchart & Remsberg (1986).

References

Butchart, N. & Remsberg, E. E. 1986 *J. atmos. Sci.* **43**, 1319–1339.
Farman, J. C., Murgatroyd, R. J., Silnickas, A. M. & Thrush, B. A. 1985 *Q. Jl R. Met. Soc.* **111**, 1013–1025.
McIntyre, M. E. & Palmer, T. N. 1985 *Pure appl. Geophys.* **123**, 964–975.