The Influence of Gravity Wave Breaking on the General Circulation of the Middle Atmosphere[1]

JAMES R. HOLTON

Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195

(Manuscript received 13 April 1983, in final form 1 July 1983)

ABSTRACT

The zonal mean solstice circulation of the global middle atmosphere is simulated using a semi-spectral numerical model. Radiative heating and cooling is computed by the algorithm of Wehrbein and Leovy. Mechanical dissipation is represented by the gravity wave breaking parameterization of Lindzen. An inertial adjustment parameterization is used to prevent the development of inertially unstable meridional shears near the equator. It is shown that gravity wave drag and diffusion in the mesosphere can account for the observed large departure from radiative equilibrium in both summer and winter.

Experiments incorporating a forced stationary wavenumber 1 disturbance indicate that planetary wave EP flux convergences, although they may modify the mean flow profile significantly, cannot provide the major source of mechanical dissipation in the winter mesosphere.

A simulated sudden warming is accompanied by an equally strong mesospheric cooling. This cooling is caused primarily by the relaxation of the polar mesosphere toward radiative equilibrium when the easterly mean winds in the polar stratosphere induced by the sudden warming reduce the transmission of gravity waves into the mesosphere.

1. Introduction

Since the classic paper of Leovy (1964), it is known that substantial mechanical dissipation is required in the middle atmosphere to compensate for the strong zonal acceleration produced by the Coriolis torque due to the meridional circulation which develops in response to the differential radiative heating of winter and summer hemispheres. Leovy obtained a qualitatively reasonable model of the mean zonal wind and temperature distributions by parameterizing mechanical dissipation in the simplest possible manner as a height-independent Rayleigh friction. Schoeberl and Strobel (1978) and Holton and Wehrbein (1980a; hereafter HWa) found that Leovy's results could be improved by specifying a height dependent Rayleigh friction which in the case of HWa had a time constant >80 days below 50 km decreasing to ~2 days above the mesopause.

All three of the above models incorporated very simple parameterizations of the radiative heating and cooling of the middle atmosphere in which the infrared radiation was treated in terms of a Newtonian cooling to a specified standard temperature distribution. Results from the GFDL "SKY-HI" general circulation model (Fels *et al.*, 1980) indicate, however, that Newtonian cooling is a poor approximation for the radiative drive in the middle atmosphere. When a more accurate radiative algorithm is employed, the mean zonal winds at the solstice period are much stronger than suggested by HWa and Schoeberl and Strobel (1978) or given by observations.

Wehrbein and Leovy (1982) have recently developed a new radiative algorithm for computing the radiative drive in the middle atmosphere in dynamical models. Their scheme employs a Curtis matrix to compute cooling by carbon dioxide.[2] Voigt line shape, vibrational relaxation, line overlap, and temperature dependence of line strength distributions and transmission functions are all incorporated into their matrix elements. Cooling by the 9.6 μm band of ozone is represented by a cool-to-space and cool-to-ground approximation. Absorption of solar radiation by ozone and molecular oxygen is included using the parameterization schemes of Lacis and Hansen (1974) and Strobel (1978), respectively. They find that when this algorithm is employed in the model of HWa the radiative heating (cooling) at the summer (winter) solstice is about three times that given by the HWa model. This increased differential heating drives a mean meridional circulation which is approxi-

[1] Contribution No. 668, Department of Atmospheric Sciences, University of Washington.

[2] Leovy (personal communication, 1983) has indicated that the cooling rates for the 10 μm band of carbon dioxide given in Wehrbein and Leovy (1982) are in error. However, this error should affect the total heating rates by less than 10%, and should have little influence on the present results.

mately twice the strength given by HWa. When the Rayleigh friction profile of HWa is used, the mean zonal winds also are double the amplitudes computed in HWa for both summer and winter solstices. Furthermore, there is little indication of the observed dynamically driven reversal of the meridional temperature gradient near the mesopause.

Thus we conclude that the apparent success of HWa in simulating the observed solstice zonal wind circulation must be attributed to a serious underestimate of the differential radiative forcing. When the infrared radiative cooling is treated more carefully, as in Wehrbein and Leovy (1982), much stronger mechanical dissipation is apparently required to simulate the observed zonal wind profiles. It is still possible to produce a "realistic" zonal wind profile using a Rayleigh friction parameterization, if a damping time as short as ½ day is specified at the mesopause level. Fig. 1 shows the solstice mean zonal wind computed using the HWa model with the Wehrbein and Leovy (1982) radiation code and a height-dependent Rayleigh friction profile which has a damping rate of 1/(80 days) below 50 km, a rapid increase with height from 60 to 80 km, and a rate coefficient of 1/(0.5 days) above 80 km.

Although this strongly enhanced Rayleigh frictional damping in the mesosphere leads to a reasonable zonal wind profile (except that the summer easterlies are too weak) it is difficult to justify this simple form of damping on physical grounds. HWa argued that the Rayleigh friction profile in their model represented a crude parameterization of the wave drag induced by the breakdown of vertically propagating internal gravity waves near the mesopause. However, Hodges (1967) pointed out that breaking vertically propagating internal gravity waves will produce turbulent diffusion. This aspect of the gravity wave–mean flow interaction is not at all represented by a Rayleigh friction parameterization. Lindzen (1981) derived a parameterization for both the diffusion and the wave drag due to breaking waves, and showed that Rayleigh friction is a poor parameterization of the total effect of wave breaking on the large-scale flow. Rayleigh frictional damping with sufficiently short time scales can indeed provide a strong drag force, but only if the mean wind is sufficiently strong. Thus, the simplest form of Rayleigh friction (in which the mean wind is relaxed toward zero) cannot possibly provide the force necessary to balance the Coriolis torque of the mean meridional circulation near the mesopause where the mean zonal flow is very weak. Nor can such a Rayleigh friction parameterization successfully model the observed reversal of the mean winds above the mesopause.

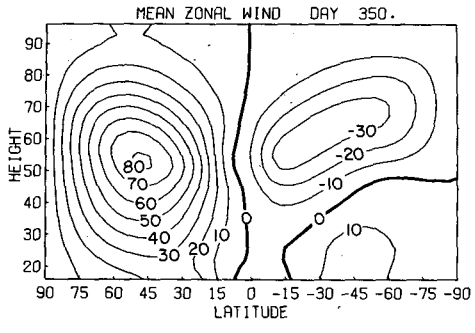

FIG. 1. Mean zonal wind (m s^{-1}) at Northern Hemisphere winter solstice in the zonally symmetric model with the strong Rayleigh friction parameterization described in the text. (A 360-day year has been used throughout this work with the winter solstice at day 350.)

It would be theoretically possible to use a modified Rayleigh friction in which the mean wind was relaxed toward a value other than zero, as suggested by Dunkerton (1982). Such an approach would no doubt provide a better basis for reproducing the observed mean zonal wind profile near the mesopause. However, the wave breaking parameterization suggested by Lindzen (1981) offers a more attractive alternative since it provides an explicit prescription for the turbulent diffusion due to wave breaking as well as the wave drag.

Although there can be little question that gravity wave breaking is an important source for mechanical damping in the mesosphere, and must indeed be the major damping mechanism in the summer mesosphere, there remains the possibility that wave drag due to quasi-stationary planetary waves might be the primary source of mechanical damping in the winter mesosphere. Holton and Wehrbein (1980b; hereafter HWb) found that when a forced planetary wave of zonal wavenumber 1 was included explicitly in their model, the overall zonal mean thermal structure found in HWa was little changed, provided that the height-dependent Rayleigh friction of HWa was retained. However, when Rayleigh friction was omitted, unrealistically strong mean zonal winds developed which blocked planetary wave propagation in the winter hemisphere. HWb concluded that planetary waves do not play an essential role in maintaining the overall zonal mean momentum budget, but rather serve to modify the basic structure established by the balance between the diabatically driven mean meridional circulation and the mechanical damping due to gravity wave breaking. The major role of the planetary waves, at least in their model, was to modify the latitudinal structure of the mean zonal wind by reducing the strength of the westerlies in the subtropics, and to introduce some subseasonal variability in the form of minor stratospheric warmings.

Given the deficiencies of the model used in HWa and HWb, with its oversimplified parameterizations of the diabatic processes and mechanical dissipation, it is not clear whether the conclusions of those papers concerning the primary role of gravity wave breaking

and secondary role of planetary waves are relevant to the real atmosphere. One purpose of this paper is to show that the conclusions of HWa are still valid when more sophisticated models are used for both the diabatic processes and the mechanical dissipation. Thus, the observed structure of the solstice season zonal mean temperature field in the mesosphere can be simulated without inclusion of planetary wave forcing. However, some runs in which planetary wave forcing was included show that waves of sufficient amplitude may significantly modify the solstice season mean zonal flow profile. Such waves may also produce stratospheric sudden warmings. A secondary purpose of this paper is to show that simulated sudden warmings in this model are accompanied by strong mesospheric coolings which are, at least in part, induced by the reduction of gravity wave transmission into the mesosphere which occurs during a stratospheric warming.

2. Model parameters

The model used in the present calculations differs from that of HWa and HWb primarily in the incorporation of the Wehrbein and Leovy (1982) radiative heating algorithm, and the Lindzen (1981) wave-breaking parameterization. But, in addition, an inertial adjustment has been added to prevent the formation of inertially unstable meridional shears in the equatorial region. The heating algorithm has been described in detail by Leovy and Wehrbein and will not be discussed further here. The specification of the wave-breaking parameterization and the inertial adjustment are as follows.

a. Wave drag and diffusion

Although some limited observational data are now becoming available (Vincent and Reid, 1983), it is not yet possible to specify the climatology of the latitude and time dependence of internal gravity wave amplitude and frequency distributions. Following Lindzen (1981) we specify a simple distribution of gravity waves which provides a wave drag field adequate to balance the momentum budget of the mesosphere for the observed mean wind profiles. The actual wave parameters chosen were guided by the experiments of Holton (1982) who used a severely truncated β-plane channel model to show that realistic solstice season mean winds could be modeled with a very simple wave spectrum consisting of three single frequency zonally propagating waves of phase speeds 0, +20 and −20 m s^{-1}, respectively.

Lindzen (1981) argued that the $c = 0$ and $c = +20$ m s^{-1} modes should dominate the gravity wave spectrum as these would be generated by flow over orography and shear instability in the tropospheric jet stream, respectively. The −20 m s^{-1} mode incorporated here was included for reasons discussed below. This choice of wave modes is admittedly somewhat arbitrary. However, the choice is constrained somewhat by the requirement that the deceleration produced by the total wave drag must be about 100 m s^{-1} day^{-1} in both the summer and winter upper mesosphere regions. Only with wave drag of this amplitude is it possible to balance the enormous Coriolis torque of the mean meridional circulation produced by the radiative forcing.

The Holton (1982) model was quasi-geostrophic, while the present model is a primitive equation model in spherical geometry. In such a model great care is required to avoid generation of spurious inertio-gravity oscillations, especially when the model domain extends over many scale heights as in the present case. Thus, it is not surprising that the wave parameters used by Holton (1982) needed to be somewhat modified for the present application. In the Lindzen formulation, the wave drag is strongly peaked near the lowest level at which the wave breaks. This breaking level is itself a function of the Doppler-shifted phase speed and hence changes in time with the mean zonal flow. To avoid excessive noise generation in our primitive equation model it proved necessary to temporally smooth the evolution of the breaking level and to spatially smooth the wave drag force so that the drag decayed exponentially below the breaking level over an atmospheric scale height. Physically such a smearing of the breaking level can be regarded as crudely representing the effects of spatially and temporally varying gravity wave amplitudes.

In the notation of Holton (1982), the Lindzen (1981) parameterization can be expressed for a single-frequency, zonally propagating wave in terms of formulas which give the breaking level z_b, the zonal wave drag F_x, and the turbulent diffusion coefficient D in terms of the zonal mean flow $\bar{u}$, the buoyancy frequency N, the scale height H, and the wave parameters c, k and B. Here c and k are the zonal phase speed and wavenumber, respectively, and B is the vertical perturbation velocity amplitude of the gravity wave at the tropopause. Thus, the breaking level is given by

$$z_b = 3H \ln(|\bar{u} - c|/\tilde{u}), \tag{1}$$

where $\tilde{u} = [BN/(k|u_0 - c|^{1/2})]^{2/3}$ and u_0 is the mean zonal wind speed at the tropopause. The wave drag and diffusion are given respectively by

$$F_x = -A(\bar{u} - c)^2[(\bar{u} - c) - 3Hd\bar{u}/dz], \tag{2}$$

$$D = -(\bar{u} - c)F_x/N^2, \tag{3}$$

where $A = \gamma k/(2NH)$ with γ an "efficiency" factor less than 1, which accounts for the probability that the waves will not be present at all longitudes or all times. Once c and k are specified, Eqs. (1) and (2) show that B and γ cannot be arbitrarily chosen, but

are constrained by the requirements that the waves must break in the mesosphere and produce a large drag. In the present study we have empirically determined A by comparing model runs with several different values and selecting the value which led to the best simulation of the observed mean zonal flow. The wave parameters used in the present model are summarized in Table 1.

Lindzen (1981) argued that in the winter hemisphere topographically generated waves with $c = 0$ should dominate, and that in the summer hemisphere waves generated by systems moving with the upper tropospheric flow ($c = 20$ m s^{-1}) should dominate since the $c = 0$ waves would be blocked from reaching the mesosphere in summer by the $\bar{u} = 0$ critical level in the lower stratosphere. Holton (1982) confirmed that indeed a wave "spectrum" consisting of only these two phase speeds does provide fairly adequate wave drag profiles in both summer and winter hemispheres, but that the addition of a third component with $c = -20$ m s^{-1} improved the simulation near the winter mesopause. The Holton (1982) model did not, however, allow for meridional variation in the level of wave breaking. In the spherical model considered here (1)–(3) must be used to determine the breaking level, wave drag and diffusion for every grid point (i.e., at 10° latitude intervals). Since observations are inadequate to specify the latitudinal dependence of wave properties we have assumed that the wave parameters are all independent of latitude with one exception. For Spring season conditions in the Southern Hemisphere the mean wind distribution in the lower stratosphere occasionally blocked the propagation of the $c = +20$ m s^{-1} wave and spuriously large mesospheric easterly accelerations then occurred at those latitudes where the wave propagation was blocked. The source of the blocking proved to be a critical level due to strong westerlies at 16 or 21 km. Thus, to prevent this wave blockage the westerly phase speed gravity wave was modified such that its phase speed at any latitude was given by the greater of +20 m s^{-1} or $\bar{u}$(21 km) + 5 m s^{-1}. For the solstice season circulations which are discussed below no such adjustment was necessary.

The gravity wave with -20 m s^{-1} phase speed was included in the present model for two reasons. First, as found in the Holton (1982) model, inclusion of such a mode provides a mechanism to drive the mean wind above the winter mesosphere to the low speeds actually observed. Second, and more importantly for these calculations, the combination of the +20 and -20 m s^{-1} modes provided a strong drag which prevented the mean zonal winds in the equatorial region from accelerating to unrealistically large values. In the HWa and HWb models the strong Rayleigh friction in the mesosphere kept the equatorial winds under control; however, in the absence of Rayleigh friction it was found that a pair of gravity waves with equal and opposite phase speeds worked equally well. That a pair of gravity waves if equal and opposite phase speeds should tend to keep $\bar{u}$ near zero is a direct consequence of the functional dependence of the wave drag force on $(\bar{u} - c)$ as shown in (2). If $\bar{u}$ is positive (negative) then the wave component with $c < 0$ ($c > 0$) will produce a stronger wave drag than will the other component and the net wave drag force will be negative (positive).

TABLE 1. Water parameters.

Mode	c (m s^{-1})	$\hat{u}$ (m s^{-1})	A (m^{-2} s^{-1})
c_1	0.0	3.0	0.8×10^{-8}
c_2	20.0	2.5	0.8×10^{-8}
c_3	−20.0	2.0	0.8×10^{-8}

The major source of gravity waves in the tropics is probably large-scale convective disturbances. The tendency for much of the convection in the tropics to be geographically fixed but oscillating in time provides at least some physical justification for a wave spectrum with equal amplitude in eastward and westward moving gravity wave modes. However, the choice of the +20 and -20 m s^{-1} modes is dictated in the present model mainly by the desire for computational efficiency.

b. *Inertial adjustment*

One of the interesting aspects of HWa was the occurrence of an equatorial semi-annual oscillation due to the advection of the summer hemisphere easterlies across the equator by the diabatically driven mean meridional circulation. However, it has been pointed out by Dunkerton (1981) that the resulting circulation with strong westerly shears at the equator satisfies the conditions for zonally symmetric inertial instability. Dunkerton showed that the growth rates for such instabilities are largest for small vertical scales. For the rather crude 5 km resolution of the HW model viscosity apparently suppressed the instability. Nevertheless it seems undesirable to permit the existence in the model of a mean flow distribution which likely could not exist in the atmosphere. Thus, in the present model, we have included an inertial adjustment process similar in spirit to the convective adjustment used in general circulation models (e.g., Manabe *et al.*, 1965). In this adjustment scheme the mean wind distribution is tested at grid points starting at the equator and moving into the winter hemisphere. At each point j the flow is tested for inertial instability by evaluating the sign of

$$f[f - (\cos\phi)^{-1}\partial(\bar{u}\cos\phi)/\partial y]. \tag{4}$$

If the sign of (4) is negative the mean wind is adjusted at points j and $j + 1$ to restore the flow to neutral inertial stability while preserving the mean

angular momentum. As Dunkerton (1981) pointed out, this inertial adjustment can under some conditions generate a barotropically unstable profile. Thus we have followed the inertial adjustment with a barotropic adjustment by setting

$$\partial[f - (\cos\phi)^{-1}\partial(\bar{u}\cos\phi)/\partial y]/\partial y$$

equal to zero whenever it is negative by again adjusting the mean wind between two grid points while conserving angular momentum.

3. The solstice circulation in the two-dimensional model

The model of HWa, modified as described in the previous section, was used to simulate the solstice season mean circulation. The wave drag force (2) was included in the zonal mean momentum equation and the vertical eddy diffusion due to gravity wave breaking was included in both the momentum and thermodynamic energy equations. The model has been described in detail in HWa and the reader is referred to that paper for technical details. In the runs reported here the model was initialized to barotropic mean flow conditions with radiative forcing corresponding to the Northern Hemispheric autumnal equinox. Integrations were then carried out for a 90-day period terminating with Northern Hemisphere winter solstice conditions. It was found necessary to gradually invoke the radiative forcing and mechanical damping effects over the course of the first 10 days of integration in order to avoid excessive generation of gravity wave noise. All dynamical fields were settled into slow seasonal variation by the time the integration reached the Northern Hemisphere winter solstice.

Height–latitude sections for various fields are shown in Figs. 2–8. The mean zonal wind distribution (Fig. 2) and temperature distribution (Fig. 3) are quite similar to those of HWa. However, the winter hemisphere westerly jet core is located about a scale height (7 km) lower in the present case than given by HWa and by observations. The height of the westerly jet core is

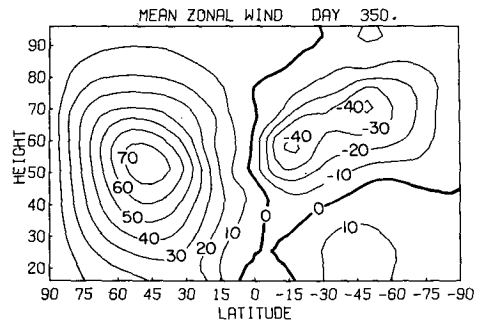

FIG. 2. Mean zonal wind (m s^{-1}) at Northern Hemisphere winter solstice in the zonally symmetric model with the wave drag and diffusion parameterization described in the text.

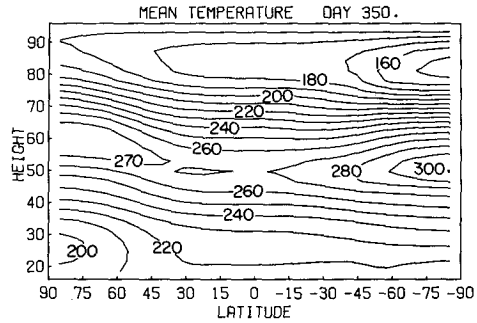

FIG. 3. As in Fig. 2, but for the zonal mean temperature profile (K).

determined in the model by a balance between the radiative forcing, which generates westerly shear with height at all levels, and the wave-drag force which strongly decelerates the mesospheric winds. The wave parameters could be adjusted to move the primary wave drag to a higher elevation in the winter hemisphere, but only at the expense of producing unrealistically strong winds in the jet core. Due to the lack of information on the actual gravity wave climatology in the mesosphere it does not seem worthwhile to attempt to obtain minor improvements in the mean wind profile by excessive "tuning" of the gravity wave parameters. Moreover, it seems unlikely that adjusting the gravity wave drag distribution can resolve all the discrepancies between Fig. 2 and observations. In particular, as can be seen from Figs. 2 and 3, in the lower stratosphere the mean zonal wind is too strong and the polar temperatures are a few degrees too cold. These cold temperatures may be caused by deficiencies in the radiative algorithm in the lower stratosphere, or may be due to the neglect of planetary waves, or both. Ramanathan *et al.* (1983) have recently shown that the winter temperature distribution in the lower stratosphere is apparently quite sensitive to the radiative forcing despite the relatively weak cooling rates at those altitudes. They also found that planetary waves may play an important role in amplifying the thermal response to changes in the radiative forcing.

Although the winter polar lower stratosphere is too cold in the model, the temperature structure in the mesosphere is quite reasonable.

The cold summer mesopause and relatively warm winter mesopause characteristic of observations are nicely simulated. This simulation of the mesopause temperature structure is of course a direct consequence of the strong wave drag which constrains the mean winds near the mesopause and hence, through the thermal wind balance, also constrains the temperature field. Thus, as speculated by Houghton (1978) and others, the reversed meridional temperature gradient at the mesopause must be regarded as a response to dynamical processes which are suffi-

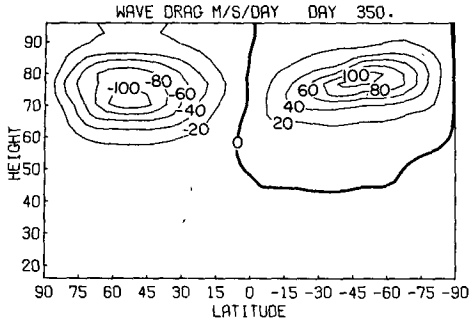

FIG. 4. As in Fig. 2, but for the gravity wave drag (m s^{-1} day^{-1}).

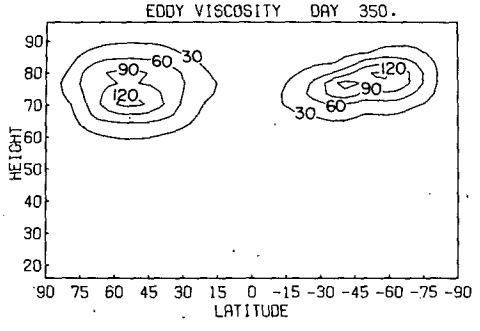

FIG. 5. As in Fig. 2, but for the gravity-wave-induced eddy viscosity coefficient (m s^{-1}).

ciently strong to completely overwhelm the radiative drive.

The dominance of the wave drag and diffusion near the mesopause is illustrated in Figs. 4 and 5 which show the acceleration due to the wave drag and the turbulent diffusion coefficient, respectively. In accord with Lindzen's (1981) suggestion we have set the amplitude (e.g., the parameter ũ in Table 1) of the topographic ($c = 0$) wave greater than that of the other modes so that the wave breaking occurs lower in the winter hemisphere than in the summer hemisphere. The drag and diffusion are smeared out over a greater depth in the winter hemisphere since both the $c = 0$ and $c = -20$ m s^{-1} modes contribute to the drag in the winter and the latter mode breaks about 10 km higher than the former.

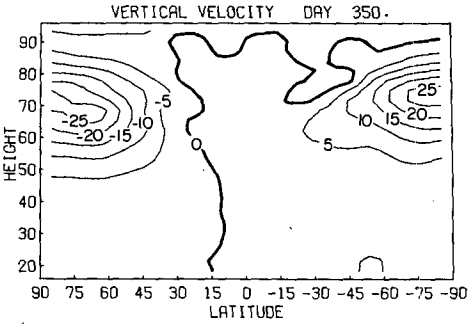

FIG. 7. As in Fig. 2, but for the zonal mean vertical motion (mm s^{-1}).

The amplitudes of the wave drag force maxima in the middle latitudes of both hemispheres are extremely large compared to the actual mean zonal wind accelerations. Such large magnitudes are required, however, to balance the huge Coriolis torques due to the mean meridional wind shown in Fig. 6. In fact, probably the most dramatic difference in the dynamical fields between the present model and the HWa model is the nearly fourfold increase in magnitude of the mean meridional wind. This increase (which is, of course, also present in the vertical velocity field of Fig. 7) is consistent with the fourfold increase in the magnitude of the differential radiative heating in the mesosphere (Fig. 8) compared to that of the Newtonian cooling model of HWa. Since the mean meridional circulation must provide adiabatic heating/cooling to compensate the radiative forcing, it is not surprising that the strength of the meridional circulation is directly proportional to the strength of the radiative drive.

Gravity wave drag and diffusion are both implied by the Lindzen parameterization. But for the mean circulation the wave drag is the essential physical process. Only wave drag can force the mean zonal winds near the mesopause to remain weak in the presence of the very strong differential radiative drive. The role of wave diffusion is a secondary one. It acts to ver-

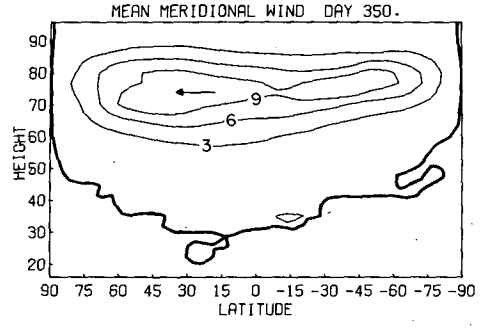

FIG. 6. As in Fig. 2, but for the mean meridional wind (m s^{-1}).

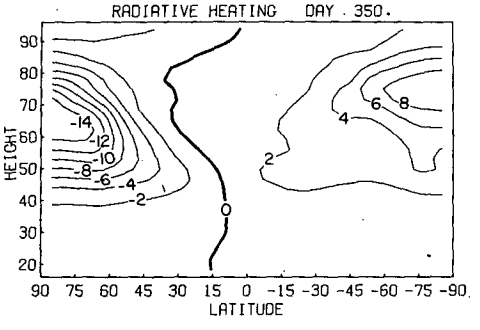

FIG. 8. As in Fig. 2, but for the radiative heating rate (K day^{-1}).

tically smooth the wind profile forced by the wave drag. Schoeberl and Strobel (1983) have argued that the downward heat flux produced by breaking gravity waves is an essential feature of the heat budget of the mesosphere. Although the resultant gravity-wave-induced cooling may indeed be important in the heat budget it can only effect the mean zonal flow through modifying the mean meridional circulation, and thus must be regarded as secondary to the wave drag effect.

In summary, the strong constraint placed on the $\bar{u}$ amplitude by the cubic dependence of F on $(\bar{u} - c)$ assures that the $\bar{u}$ distribution will be qualitatively like that of the observed profile despite the opposing tendencies of the radiative drive. The cross-equatorial mass flow near the mesopause, which depends sensitively on the differential radiative forcing is, according to this model, strong enough to advect an air parcel from the high latitude of the summer hemisphere to the high-latitude winter hemisphere in less than a month. Although such meridional drifts seem rather strong they are consistent with radar observations reported by Nastrom *et al.* (1982) who found that an equatorward drift of 10–15 m s^{-1} at 90 km was typical at 60°N in June.

4. The influence of forced stationary planetary waves on the solstice mean circulation

Gravity wave drag and diffusion appear to be the only mechanical dissipation effects which could maintain the momentum balance in the summer mesosphere where planetary wave activity is very weak. However, it might be argued that in the winter season quasi-stationary planetary waves could provide the required wave drag since such waves are known to penetrate well into the mesosphere (Barnett, 1980). To examine this possibility, the model of HWb, modified to include gravity wave drag and diffusion, was used for an experiment analogous to that described in the previous section, but with the inclusion of a forced stationary planetary wave. In the presence of zonally asymmetric waves the gravity wave breaking level, wave drag and diffusion should all be calculated as functions of longitude as well as latitude. However, with the present severely truncated model and the simple spectrum assumed for the gravity waves such a level of detail seemed unjustified. In this model the wave drag was assumed to act only on the zonal mean flow while the eddy diffusion due to the gravity waves was allowed to act on both the mean flow and the planetary waves. This simplification should be qualitatively valid if indeed the primary role of wave breaking is to provide a strong drag force on the mean zonal flow.

In the HWb model a perturbation in the geopotential topography at the lower boundary (16 km) was specified as a zonal wavenumber 1 component with Fourier coefficient, i.e.,

$$\Psi_s(y) = \begin{cases} gh_B \exp(-i|\theta|) \sin^2(3\theta - \pi/2), & |\theta| > \pi/6 \\ 0, & |\theta| \leq \pi/6 \end{cases},$$

where h_B is the amplitude of the height perturbation and θ is latitude. In the experiments described below, the wave forcing was set to zero initially, and approached its final steady value exponentially with a 10-day e-folding time to minimize noise generated by the initial adjustment process.

The model was initialized with barotropic conditions at the Northern Hemisphere autumn equinox and integrated forward to the winter solstice for several values of steady-state wave forcing amplitude. In nearly all these experiments the wave drag and diffusion parameterizations (2) and (3) were employed. In one case, however, an integration was attempted with the small-scale wave drag and diffusion omitted. The results dramatically illustrated the necessity for gravity wave drag in the mesosphere. In the absence of the parameterized wave drag, the radiative drive quickly accelerated the mean zonal winds in the autumn mesosphere to speeds sufficiently strong to prevent the upward propagation of stationary planetary waves. Hence, the forced planetary waves were reflected and could not provide a wave drag sufficient to prevent further acceleration of the mean zonal winds to very unrealistic values as the solstice season approached. Thus, this model confirms the findings of HWb that planetary wave drag alone cannot provide the deceleration necessary to balance the torque of the mean meridional circulation, even in the winter season.

With the wave drag and diffusion parameterization included it was found that for a zonal wavenumber 1 forcing of 250 m amplitude at the lower boundary, the Eliassen-Palm (EP) flux divergence of the planetary waves was too small in the upper stratosphere and mesosphere to have any significant effect on the solstice circulation. The mean zonal flow and temperature distributions were virtually identical to those shown in Figs. 2 and 3, respectively. Several previous studies (e.g., Schoeberl and Strobel, 1980; Holton and Wehrbein, 1981) have shown that strong planetary wave mean-flow interaction only occurs when the planetary wave forcing exceeds a certain threshold value. In the present model when the wave forcing was raised to 275 m amplitude there was a strong planetary wave–mean flow interaction, resulting in an erosion of the low-latitude flank of the westerly jet, as is clear from Fig. 9. The resulting mean wind profile has the jet core in the winter hemisphere far poleward of that in the zonally symmetric model shown in Fig. 2. The mean zonal wind pattern of Fig. 9 is similar to that shown as Fig. 8 in HWb. As pointed out in HWb the zonal mean temperature distribution is changed considerably less than the wind profile because the major wind changes occur

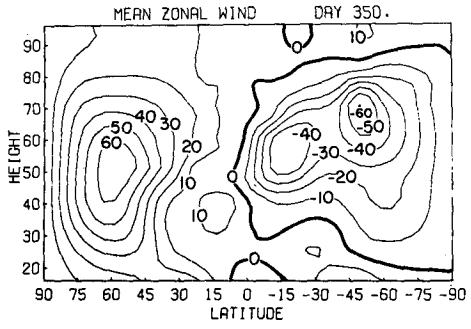

FIG. 9. Mean zonal wind (m s^{-1}) at Northern Hemisphere winter solstice in the presence of a forced planetary wavenumber 1 disturbance.

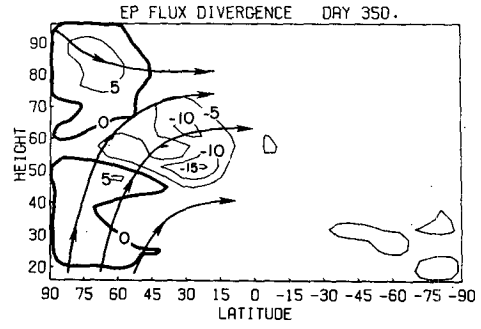

FIG. 11. The EP flux divergence due to the forced planetary waves for the mean wind profile of Fig. 9. Arrows show some integral curves of the EP flux vector.

in the subtropics where small temperature changes can support large changes in the thermal wind.

The mean meridional circulation in this case is also similar to that of the zonally symmetric case (Fig. 6). However, here the momentum budget is maintained not merely by a balance between the Coriolis torque and the wave drag shown in Fig. 10; the planetary-wave-induced mean wind deceleration is also significant. This deceleration is measured by the EP flux divergence (Andrews and McIntyre, 1976) shown in Fig. 11. The planetary waves clearly provide most of the drag responsible for decelerating the weak westerlies on the equatorial side of the jet where there is a strong EP flux convergence (equatorward potential vorticity flux) which balances the Coriolis torque of the mean meridional circulation. But, in the polar region, where there is a significant EP flux divergence (poleward potential vorticity flux), the planetary waves actually help maintain the westerly jet.

5. Stratospheric warmings and mesospheric coolings

In the past decade a number of mechanistic numerical models have confirmed the general validity of Matsuno's (1971) theory of the sudden stratospheric warmings. According to this theory, the zonal mean wind deceleration and accompanying warming which occur in the polar stratospheric during a sudden warming event, arise from the zonal force due to the potential vorticity flux (EP flux divergence) of forced, vertically propagating, planetary waves. A number of aspects of the dynamics of sudden warmings have been reviewed by Schoeberl (1978), Holton (1980) and McIntyre (1982). McIntyre's paper is particularly notable for its analysis of the zonal mean potential vorticity distribution which allows planetary wave activity to be ducted into the polar stratosphere and which, therefore, is favorable for the occurrence of sudden warmings. Recent studies (e.g., Palmer, 1981; Dunkerton *et al.*, 1981; Butchart *et al.*, 1982; Palmer and Hsu, 1983) have served to further elucidate the role of EP flux divergence patterns and their relationship to the mean potential vorticity distribution in the generation of sudden warmings.

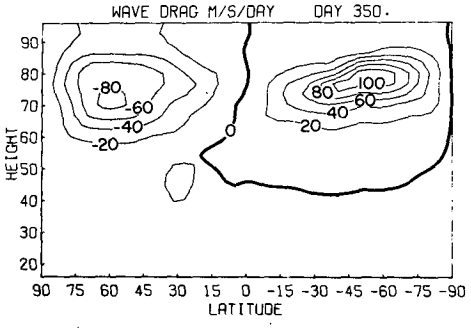

FIG. 10. The gravity wave drag (m s^{-1} day^{-1}) corresponding to the mean flow profile of Fig. 9.

The first warming simulation by Matsuno (1971) suggested that compensating mesospheric coolings would accompany stratospheric warmings. Some observational evidence of such coolings has appeared in the literature (Quiroz, 1969; Labitzke, 1972; Hirota and Barnett, 1977), but there has been very little theoretical analysis of them. An exception is the study of Matsuno and Nakamura (1979) who showed that a compensating cooling was required above the planetary wave critical line in a simple model for stratospheric warmings. In their model the cooling was entirely due to adiabatic ascent of a forced mean meridional circulation induced by the strong equatorward potential vorticity flux at the critical level. This mechanism has been demonstrated to produce strong mesospheric coolings in conjunction with very high altitude sudden warmings in a numerical simulation by Schoeberl and Strobel (1980). This type of secondary meridional circulation probably plays some part in mesospheric coolings observed during actual warming events. However, it must be remembered that the climatological winter mesosphere is much warmer than radiative equilibrium, and that this thermal structure is maintained by subsidence warming associated with the mean meridional circulation driven by the gravity wave drag and diffusion.

As Lindzen (1981) noted, the formation of a zero mean zonal wind critical line in the stratosphere in the course of a stratospheric warming would reduce the gravity wave transmission into the mesosphere. In particular, the orographically generated stationary gravity waves would be absorbed as they approached the zero mean zonal wind level. Thus, the wave breaking level in the mesosphere would be raised substantially. The resulting reduction in the gravity wave drag and diffusion would imply a reduced amplitude for the gravity-wave-induced mean meridional circulation. Hence, the polar mesosphere should radiatively cool until a new equilibrium could be established between adiabatic warming and radiative cooling. This mechanism potentially could account for substantial cooling over the entire depth of the mesosphere as reported in some rocket observations during stratospheric warmings (Labitzke, 1972). The Matsuno–Nakamura mechanism, on the other hand, could only cool the region within a Rossby height of the critical level (~15 km for the relevant meridional scale). Thus, the latter mechanism could hardly produce significant cooling in the mesosphere once the critical line descended well into the stratosphere.

In order to examine the potential of the reduced gravity wave transmission to indirectly generate mesospheric coolings, the single planetary wave–mean flow interaction model used for the experiments of the previous section was employed in a "pulse" warming experiment similar to an experiment described by Palmer and Hsu (1983). In this experiment the dynamical fields corresponding to the winter solstice (day 350) described in the previous section were used as the initial conditions for a 30-day integration during which the lower boundary forcing was specified as

$$h_B = (275\text{m})[1 + \sin(t_d\pi/30)],$$

where t_d is the time in days after the solstice.

In the example reported by Palmer and Hsu (1983) this type of pulse forcing generated a major sudden warming in which, as shown in Fig. 15 of their paper, the mean zonal flow in the region poleward of 60°N became easterly over the entire middle atmosphere above ~30 km. The model used in their study was essentially that used in HWb; thus, Newtonian cooling and Rayleigh friction parameterizations were employed. In the present model the evolution of the sudden warming in the stratosphere is quite similar to their results, but as shown in Fig. 12 the mean wind evolution in the mesosphere is greatly changed. The profile 15 days following the pulse initiation (Fig. 12a) shows weak easterlies in the mesosphere similar to those in the Palmer-Hsu experiment, but only 5 days later (Fig. 12b) the mesospheric winds have accelerated to much stronger westerly values than in the initial conditions, while in the stratosphere strong easterlies

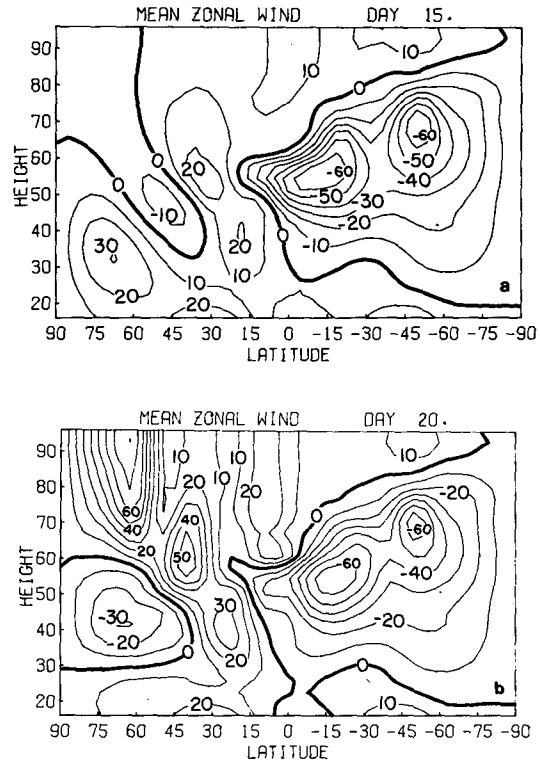

FIG. 12. Mean zonal wind profiles (m s^{-1}) for the sudden warming experiment described in the text: (a) day 15, (b) day 20. Times shown correspond to number of days following the initiation of the pulse forcing.

have developed. This evolution of the mean flow can be understood in terms of the evolution of the distributions of the EP flux divergence and the gravity wave drag shown in Figs. 13 and 14, respectively. During the early warming period shown in Figs. 13a and 14a the EP flux was strongly convergent throughout the polar middle atmosphere and the gravity wave drag was weak. But at the peak of the warming shown in Figs. 13b and 14b, the EP flux pattern has regions of both divergence and convergence and the gravity wave drag actually vanishes near 60°N. Thus, there is no drag force to prevent the Coriolis torque from strongly accelerating the mean zonal flow, and as shown in Fig. 12b strong westerlies develop in the mesosphere.

The polar temperature changes associated with this mean flow evolution are shown in Fig. 15. An examination of the thermal budget indicated that over the entire period of the pulse warming the mesospheric cooling was due almost entirely to radiative processes. The residual zonal mean vertical velocity (as defined by Edmon *et al.*, 1980) remained quite small during the whole period—even when the EP flux convergence was very strong in the mesosphere. Thus we conclude that indeed the mesospheric cooling associated with stratospheric warmings is primarily generated by a relaxation toward

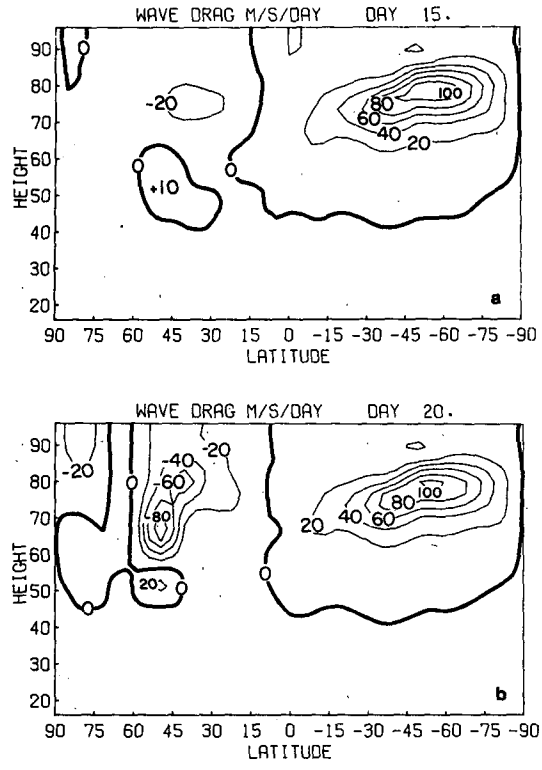

FIG. 13. Gravity wave drag (m s^{-1} day^{-1}) for the mean wind profiles of Fig. 12.

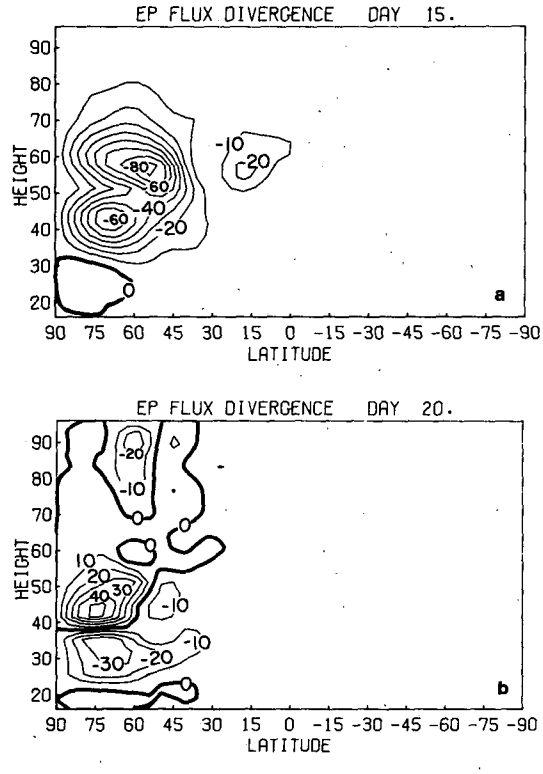

FIG. 14. EP flux divergence for the mean wind profiles of Fig. 12.

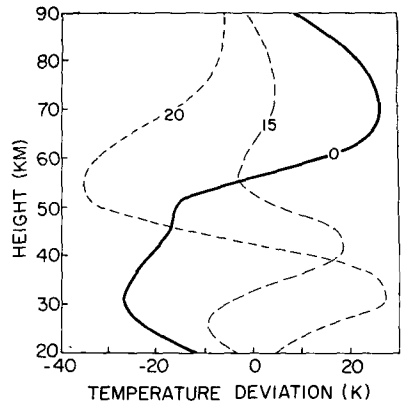

FIG. 15. Polar temperature changes (K) for the pulse warming experiment described in the text. Time labels on the curves correspond to days following the pulse initiation.

radiative equilibrium which can occur with the reduction of gravity wave transmission during the stratospheric warmings.

Vanishing of the gravity wave drag and diffusion in the vicinity of 60°N during the stratospheric warming is partly an artifact of the simplified gravity wave spectrum used in the model. For the mean zonal wind conditions in the stratosphere at the warming peak, all three gravity waves are blocked from penetrating into the mesosphere. In a more realistic model some gravity wave transmission would undoubtedly occur, but the wave drag in the mesosphere would still likely be quite weak so that the conclusions drawn from the present model should be qualitatively valid.

6. Concluding remarks

The numerical experiments described above have provided additional evidence that mechanical dissipation by the breaking of vertically propagating gravity waves is an essential component of the general circulation of the mesosphere. The parameterization scheme suggested by Lindzen (1981) provides a sound physical basis for incorporating wave drag and diffusion processes into models of the large-scale circulation. The cubic dependence of the wave drag on the Doppler-shifted phase speed provides a strong constraint which assures that the mean wind at the breaking level cannot deviate excessively from the phase speed range of the assumed gravity wave spectrum. Thus, the mean flow distribution is not excessively sensitive to poorly known parameters such as the wavenumber and amplitude spectra.

In the experiments described here a highly simplified three-component wave spectrum has been used. In reality little is yet known about the temporal and spatial distribution of gravity waves in the middle atmosphere. Although the model used here can provide some guidance as to the types of waves required

to satisfy the momentum balance, modeling alone cannot confirm the existence of such waves.

Until recently there was little possibility of determining a gravity wave climatology for the mesosphere. However, the recent development of the MST radar (Balsley and Gage, 1980) provides the potential for the design of an observational program which might provide much of the required information. Ideally, it would be desirable to know the latitudinal and longitudinal distribution of gravity wave phase speeds, amplitudes and horizontal wavelengths for all seasons. In the absence of a satellite-based observational technique such a global climatology is an unrealistic goal. However, measurements from a few locations at polar, middle and low latitudes would provide a far better basis for parameterizing wave drag and diffusion than is provided by the current sparse rocket data. Such observations should have high priority in middle atmosphere observational studies in the next several years.

Acknowledgments. I wish to thank Drs. Tim Palmer and Michael McIntyre for helpful communications. This work was supported by the National Science Foundation's Atmospheric Research Section, under NSF Grant ATM79-24687, and the National Aeronautics and Space Administration, under NASA Grant NAG-2-66.

REFERENCES

Andrews, D. G., and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: The generalized Eliassen-Palm relation and the zonal mean acceleration. *J. Atmos. Sci.*, **33**, 2031–2048.

Balsley, B. B., and K. S. Gage, 1980: The MST radar technique: Potential for middle atmosphere studies. *Pure Appl. Geophys.*, **118**, 452–493.

Barnett, J. J., 1980: Satellite measurements of middle atmosphere temperature structure. *Phil. Trans. Roy. Soc. London*, **A296**, 41–57.

Butchart, N., S. A. Clough, T. N. Palmer and P. J. Trevelyan, 1982: Simulations of an observed stratospheric warming with quasi-geostrophic refractive index as a model diagnostic. *Quart. J. Roy. Meteor. Soc.*, **108**, 475–502.

Dunkerton, T. J., 1981: On the inertial stability of the equatorial middle atmosphere. *J. Atmos. Sci.*, **38**, 2354–2364.

——, 1982: Stochastic parameterization of gravity wave stresses. *J. Atmos. Sci.*, **39**, 1711–1725.

——, C.-P. Hsu and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38**, 819–843.

Edmon, H. J., B. J. Hoskins and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616.

Fels, S. B., J. D. Mahlman, M. D. Schwarzkopf and R. W. Sinclair, 1980: Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response. *J. Atmos. Sci.*, **37**, 2265–2297.

Hirota, I., and J. J. Barnett, 1977: Planetary waves in the winter mesosphere—Preliminary analysis of the Nimbus 6 PMR results. *Quart. J. Roy. Meteor. Soc.*, **103**, 487–498.

Hodges, R. R., 1967: Generation of turbulence in the upper atmosphere by internal gravity waves. *J. Geophys. Res.*, **72**, 3455–3458.

Holton, J. R., 1980: The dynamics of sudden stratospheric warmings. *Annual Review of Earth and Planetary Science*, Vol. 8, Annual Reviews, 169–190.

——, 1982: The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *J. Atmos. Sci.*, **39**, 791–799.

——, and W. M. Wehrbein, 1980a: A numerical model of the zonal mean circulation of the middle atmosphere. *Pure Appl. Geophys.*, **118**, 284–306.

——, and ——, 1980b: The role of forced planetary waves in the annual cycle of the zonal mean circulation of the middle atmosphere. *J. Atmos. Sci.*, **37**, 1968–1983.

——, and ——, 1981: A further study of the annual cycle of the zonal mean circulation in the middle atmosphere. *J. Atmos. Sci.*, **38**, 1504–1509.

Houghton, J. T., 1978: The stratosphere and mesosphere. *Quart. J. Roy. Meteor. Soc.*, **104**, 1–29.

Labitzke, K., 1972: Temperature changes in the mesosphere and stratosphere connected with circulation changes in winter. *J. Atmos. Sci.*, **29**, 756–766.

Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.

Leovy, C. B., 1964: Simple models of thermally driven mesospheric circulation. *J. Atmos. Sci.*, **21**, 327–341.

Lindzen, R. S., 1981: Turbulence and stress due to gravity wave and tidal breakdown. *J. Geophys. Res.*, **86**, 9707–9714.

McIntyre, M. E., 1982: How well do we understand stratospheric warmings? *J. Meteor. Soc. Japan*, **60**, 37–65.

Manabe, S., J. Smagorinsky and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.*, **93**, 769–798.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479–1494.

——, and K. Nakamura, 1979: The Eulerian and Lagrangian-mean meridional circulations in the stratosphere at the time of a sudden warming. *J. Atmos. Sci.*, **36**, 640–654.

Nastrom, G. D., B. B. Balsley and D. A. Carter, 1982: Mean meridional winds in the mid- and high-latitude summer mesosphere. *Geophys. Res. Lett.*, **9**, 139–142.

Palmer, T. N., 1981: Diagnostic study of wavenumber-2 stratospheric sudden warming in the transformed Eulerian mean formalism. *J. Atmos. Sci.*, **38**, 844–855.

——, and C.-P. F. Hsu, 1983: Stratospheric sudden coolings and the role of nonlinear wave interactions in preconditioning the circumpolar flow. *J. Atmos. Sci.*, **40**, 909–928.

Quiroz, R. S., 1969: The warming of the upper stratosphere in February 1966 and the associated structure of the mesosphere. *Mon. Wea. Rev.*, **97**, 541–552.

Ramanathan, V., E. J. Pitcher, R. C. Malone and M. L. Blackmon, 1983: The response of a spectral general circulation model to refinements in radiative processes. *J. Atmos. Sci.*, **40**, 605–630.

Schoeberl, M. R., 1978: Stratospheric warmings: Observations and theory. *Rev. Geophys. Space Phys.*, **16**, 521–538.

——, and D. F. Strobel, 1978: The zonally averaged circulation of the middle atmosphere. *J. Atmos. Sci.*, **35**, 577–591.

——, and ——, 1980: Numerical simulation of sudden stratospheric warmings. *J. Atmos. Sci.*, **37**, 214–236.

——, —— and J. P. Apruzese, 1983: A numerical model of gravity wave breaking and stress in the mesosphere. *J. Geophys. Res.*, **88**, 5249–5259.

Strobel, D. F., 1978: Parameterization of the atmospheric heating rate from 15 to 120 km due to O_2 and O_3 absorption of solar radiation. *J. Geophys. Res.*, **83**, 6225–6230.

Vincent, R. A., and I. M. Reid, 1983: HF Doppler measurements of mesospheric gravity wave momentum fluxes. *J. Atmos. Sci.*, **40**, 1321–1333.

Wehrbein, W. M., and C. B. Leovy, 1982: An accurate radiative heating and cooling algorithm for use in a dynamical model of the middle atmosphere. *J. Atmos. Sci.*, **39**, 1532–1544.