Stochastic Parameterization of Gravity Wave Stresses

TIMOTHY J. DUNKERTON

National Center for Atmospheric Research1, Boulder, CO 80307

(Manuscript received 8 January 1982, in final form 6 April 1982)

ABSTRACT

Selective transmission of gravity waves into the upper mesosphere and lower thermosphere leads to the generation of mean flows opposite to those below. This interaction is addressed in the context of a simplified transient, stochastic, conservative wave model. The dominant phase speed in the spectrum appears to determine the magnitude of the upper level flow, while the height of interaction is determined by the forcing amplitude. Observed features of the upper atmosphere are efficiently explained by this model, and the results compare very well to recent steady wave models, despite the differing formulations.

The model also provides a tentative explanation of the semianual oscillation in the tropical upper mesosphere.

1. Introduction

Perhaps the most disheartening aspect of atmospheric modeling is the dependence of large-scale motions on subgrid-scale processes. Those familiar with the upper atmosphere will recognize that this region of the atmosphere is not immune to this difficulty. Small-scale gravity wave oscillations are evident in meteor trails, grenade explosions and noctilucent clouds and are measured by radar techniques. For many years theorists have realized that these small-scale waves are essential for diffusion in the upper atmosphere (Lindzen, 1967; Hodges, 1969) and for deceleration of the mean flow there (Leovy, 1964). The latter seems required by the apparent magnitude of the diabatic circulation (Lindzen, 1981; Holton, 1982).

Because these gravity waves are ubiquitous and are too complex to model in detail, some parameterization of their twin effects, diffusion and deceleration, must be employed. It seems significant that numerical models not specifically incorporating the wave stress have thus far failed to simulate realistic mesospheric mean flows (Lindzen, 1981).

a. Leovy's (1964) Rayleigh friction

A parameterization of gravity wave stresses appeared implicitly in Leovy's (1964) important paper (implicitly, since Leovy's concern lay with eddy effects as a whole). Leovy invoked a “Rayleigh friction” in the mean flow tendency

\[ \vec{u} = -\alpha_M \vec{u} + \cdots \] (1.1)

where \( \alpha_M \) is a damping rate coefficient independent of height. Leovy then solved for the lowest harmonic of the annual cycle driven by a “return to radiative equilibrium” diabatic heating. Using a value of \( \alpha_M \) approximately an order of magnitude greater than the annual cycle frequency allowed Leovy to simulate a realistic mesospheric jet structure of \( \mathcal{O}(100 \text{ m s}^{-1}) \) or less.

An interesting feature of Leovy’s (1964) parameterization is that while Rayleigh friction does not explicitly parameterize any known atmospheric eddy effect, the overall sense of this friction is to bring the mean flow back to zero. This is precisely what is to be expected from the absorption of stationary internal gravity waves in the middle atmosphere. A well-known property of these waves is to drive mean flows in the direction of their phase speed. Of course, the local interaction, in time and space, of a gravity wave with the mean flow is not specifically given by (1.1). Nevertheless, since the stationary waves are probably the most prevalent of all gravity waves, their ability to retard the mean flow in the sense indicated in (1.1) lends considerable credence to this parameterization.

An obvious deficiency in Leovy’s (1964) use of (1.1) is that \( \alpha_M \) was assumed independent of height. Recently Schoebel and Strobel (1978) and Holton and Wehrbein (1980) sought to improve upon this by allowing \( \alpha_M \) to be significant only in the mesosphere. After all, it is only in this region that difficulty simulating mean flows is encountered. Further rationale for this height dependence has come from Lindzen’s (1981) discussion of wave “saturation.”

Lindzen recognized that steady and otherwise conservative gravity wave stresses are independent of height (Eliassen and Palm, 1960) except in a region

0022-4928/82/081711-15$07.75
© 1982 American Meteorological Society
where the wave “breaks.” This region, according to Lindzen, is where linear theory predicts convective instability. Such instabilities arise from wave growth with height due to mean flow changes and the density effect. By assuming that the implied convective instabilities generate just enough mixing to maintain neutrally stable lapse rates within the wave, Lindzen (1981) analytically determined the magnitudes of the diffusion and wave stress divergence in the saturated region. Because the latter becomes nonzero above the wavebreaking level, the steady wave’s ability to drive the mean flow in the direction of the wave phase speed manifests itself only in the saturated region (up to a critical level, if it exists, where the mean flow equals the wave phase speed). Lindzen (1981) provided some observational evidence of wavebreaking in the mesosphere, in support of the notion of a height-dependent $\Delta M$.

b. Parameterizations incorporating a steady wave spectrum

A second deficiency in (1.1) was recently noted independently by Matsuno (1982) and Holton (1982). Because gravity waves of several phase speeds other than zero are present in the atmosphere at various times, it is possible that on occasion the mean flow is being driven to equilibrium values other than zero. This fact led these authors to question the Rayleigh friction idea and replace it with a more realistic parameterization. Matsuno (1982) considered a spectrum of steady waves with five discrete phase speeds and four angles of orientation; in his model the stress divergence was calculated for each wave with the help of a prespecified eddy diffusion profile in a manner analogous to that of Plumb and McEwan (1978) who described a laboratory simulation of the quasi-biennial oscillation. Matsuno’s simulations were able to generate reversals in the mean zonal wind above the mesopause, a dynamical effect in some respects similar to the quasi-biennial oscillation (Holton and Lindzen, 1972; Plumb, 1977).

On the other hand, Holton (1982) considered a spectrum of discrete waves of phase speed $c = 0 \pm 20 \text{ m s}^{-1}$ (but with a single zonal wavenumber and angle of orientation). For each wave Holton determined the wavebreaking level, then inserted the resulting saturated wave eddy diffusion and wave stress divergence into the mean flow acceleration, in the context of a one-dimensional model essentially that of Holton and Mass (1976). (Matsuno’s model was similar in this latter respect.) Holton’s simulations also led to reversals in the mean flow above the mesopause, again demonstrating the possible importance of the nonzero phase speed waves.

Holton (1982) recognized that a combination of his approach and Matsuno’s (1982) would be more realistic. Using Holton’s method one could explicitly calculate the eddy diffusion due to a primary wave which breaks at the lowest level. Other waves supposedly breaking at higher levels would then be subject to viscous damping as discussed by Matsuno (1982).

c. A stochastic transient wave approach

The models of Matsuno (1982) and Holton (1982) have given new impetus to mesospheric modeling. It remains to be seen how well these concepts can be employed in more complex multi-dimensional models.

The purpose of this paper is first to discuss, from a slightly different and complementary perspective, the parameterization of gravity wave stresses in the mesosphere and lower thermosphere. Despite their differences, the models of Matsuno (1982) and Holton (1982) have a common feature in that both employ steady wave forcings. As the result, model simulations exhibit an annual cycle but no other variability of a more complex nature. In reality mesospheric gravity waves are likely to be as variable as the tropospheric processes leading to their generation. How this variability affects mesospheric gravity wave, mean-flow interaction is certainly a difficult question. For example, the spatial variability might very well have an impact on the vertical propagation of planetary waves. In this paper, attention will be focused on the temporal variability of gravity wave forcings, in the context of a one-dimensional model like those of Matsuno (1982) and Holton (1982).

The viewpoint adopted in this paper is that gravity wave, mean-flow interactions are fundamentally small-scale, and may be viewed as discrete “events” in time involving a temporary forcing at some lower boundary. The interaction of the gravity wave with the mean flow can be addressed in a straightforward manner by making a separation of scales in time and height. While detailed discussion of this model is reserved for later pages, a novel feature is that, for a given gravity wave event, wave parameters are assigned stochastically using a pseudo-random number generator. Over many events, the wave parameters are constrained to follow realistic probability distributions.

The second, and equally important, purpose of this paper is one of clarification. The theory of transient gravity wave, mean-flow interaction has progressed to the point where it is now possible to make very simple estimates of the net mean flow changes brought about by vertically-propagating, saturated waves. Two wave parameters turn out to be of particular importance in making such estimates: the phase speed $c$ and the time-integrated momentum flux input at the lower boundary $\Delta M$. With the stochastic transient wave model it then becomes possible
to explore, in a computationally simple way, how variations in the spectra of $c$ and $\Delta M$ affect the evolution of mesospheric mean flows.

The stochastic transient wave model also reveals the importance of the frequency of gravity wave events. Although this model can only simulate the mean wind evolution in some prescribed manner, one observes violent accelerations (of order $50 \text{ m s}^{-1} \text{ day}^{-1}$ or more) to and away from radiative equilibrium, which depend in part upon the length of quiet time between individual gravity wave events. While these large accelerations might be viewed as unrealistic, the actual amounts of momentum involved (weighted by density) are quite small.

In order to avoid being bogged down in a mass of detail describing wave transience for each gravity wave event, the stochastic model incorporates two very helpful approximations. These are:

1) During each transient wave event, the mean flow is allowed to change impulsively on account of the wave only, and not because of any external forcing, such as the Coriolis effect;

2) The net mean flow change thereby brought about by each saturated wave event is approximated by its "nonsaturated" counterpart, i.e., as if saturation had not occurred.

Item 2 might appear to undermine the entire concept of saturated wave stresses so crucial to Lindzen's (1981) theory, and Holton's (1982) use thereof, but as noted elsewhere (Dunkerton, 1982) and further discussed below, for many such wave events the saturated and "nonsaturated" mean flow changes brought about by a transient wave event are actually quite similar, a similarity due to an effective critical layer absorption in the "nonsaturated" case (Bretherton, 1966).

These approximations are helpful because they lead to a very simple expression for the net mean flow change due to a transient wave event, viz.,

$$\bar{u} = c \text{ in the interval } (z_f, z_c),$$

where

$$\int_{z_f}^{z_c} \rho_0 (c - \bar{u}) dz = \Delta M.$$  \hspace{1cm} (1.2)

Here, $\rho_0$ is the density $\rho_0 \exp -z/H$, $z_c$ is the lowest critical level above the forcing, and

$$\Delta M = \int_0^{t_f} B_0(t) dt,$$  \hspace{1cm} (1.3)

where $B_0$ is the momentum flux input at the level of forcing. Equation (1.3) defines $z_f$ to be used in (1.2).

This level is actually the asymptotic height of the trailing shock in the "nonsaturated" solution (Dunkerton, 1982). As shown in Fig. 1, the net result of the gravity wave event is to set $\bar{u}$ equal to $c$ in the interval ($z_f, z_c$). If no critical level exists initially, we take $z_c = \infty$ in the absence of any other dissipation. These relations, of course, assume an angle of orientation $\theta$ parallel to the mean flow, but one could generalize by making the replacement $c \rightarrow c \sec \theta$ (Matsuno, 1982).

In the next section we further discuss the accuracy of these model equations. As will be seen in later sections, the excellent agreement between the various gravity wave models of Matsuno (1982), Holton (1982), and the present paper is due to the effective critical layer absorption which is dominant in all models. In Section 3, a midlatitude model is formulated and results are given in Section 4. In Section 5 the concept of zonal mean transmissivity is briefly discussed. Finally, Section 6 presents a tentative model of the mesopause semiannual zonal wind oscillation (Hirota, 1978) using a tropical version of this model.

In the Appendix, the midlatitude results are compared to those obtained with a "generalized Rayleigh friction." In summary, for perpetual midlatitude solstice conditions, this frictional drag can provide excellent agreement if the drag coefficient is adequately large in the mesosphere. At the equinoxes or at the equator, however, a frictional drag is not appropriate, or at least very artificial.
2. Model basis

Formally speaking, the quasi-linear model outlined above contains three time scales. Individual gravity wave oscillations occur on a “fast” time (and height) scale. These oscillations are not explicitly calculated in the model; rather, the lowest-order equations in the inverse square-root Richardson number $\mu = [\tilde{R}] / N$ are used to formulate the model equations on a “slow” time (and height) scale (Dunkerton, 1981a). On the slow scale, wave amplitude changes and wave-induced mean flow accelerations occur. The definition of a third, ultra-slow time scale on which the external forcing operates, allows the gravity wave event to be visualized as if the mean flow were not under any external forcing at all.

a. Transient gravity wave event on the slow scale

Consider an individual gravity wave event on the slow scale. The initial mean flow $\tilde{u}_0(z)$, which is not a function of the slow time, enters the nondimensional wave action equation as

$$\frac{\partial A}{\partial t} + \left( \frac{\partial}{\partial z} - 1 \right) A(1 - A - \tilde{u}_0)^2 = 0,$$  \hspace{1cm} (2.1)

as described by Dunkerton (1981a). The time and height scales used to non-dimensionalize (2.1) are $NH/kc^2$ and $H$, respectively, where $N$ and $k$ are static stability and zonal wavenumber. The mean flow and wave action density $A$ are both normalized by the phase speed $c$. In general, the phase speed of the wave will undergo an “acceleration” brought about by the accelerating mean flow (Grimshaw, 1975; Coy, 1982), an effect not included here. Once again, these formulas describe waves whose orientation is parallel to the mean flow, but generalizations are possible.

Equation (2.1) may be solved by the method of characteristics as in Dunkerton (1981a). A non-unique solution first appears at the height

$$z_s = \ln \frac{4(1 - \tilde{u}_0)^3}{27 B_0},$$  \hspace{1cm} (2.2)

assuming for simplicity that $B_0$ is constant in the interval $(0, t_f)$. In general, a non-unique region always appears unless the mean flow $\tilde{u}_0$ satisfies the peculiar condition

$$\tilde{u}_0 < -\left( \frac{27}{4} B_0 \right)^{1/3} \exp \frac{z}{3},$$  \hspace{1cm} (2.3)

implying that the wave would propagate away from the critical level (non-dimensionally $\tilde{u} = 1$) at such a rate that its exponential density growth with height would always be more than cancelled.

The appearance of the non-unique region under these circumstances (viz. parallel orientation) coincides with the occurrence of saturation as defined by Lindzen (1981), as noted by Dunkerton (1982). At this level an internal shock appears (Dunkerton, 1981a) which descends in time according to an “equal-area” rule if the forcing is left on (Dunkerton, 1982). Once the forcing is turned off a trailing shock begins to ascend. If the saturation hypothesis is not invoked, the trailing shock asymptotes to $z_f$ as given by (1.3). With saturation, a residual shock is left at the lowest level of saturation (where trailing and internal shocks merge) while the trailing shock asymptotes to $z_c$ or infinity, if no critical level exists.

b. Accuracy of the “nonsaturated” solution

Fig. 2 shows an example of final mean flows for the nondimensional parameters $\tilde{u}_0 = 0, t_f = 6, B_0 = 0.0098$. The calculation of the saturated final mean flow required knowledge of the history of the various shock positions. On the other hand, the “nonsaturated” step function final mean flow could be inferred at once from (1.2) (Dunkerton, 1982). Clearly the similarity in the two solutions, together with the ease of using (1.2), motivates us to replace the more accurate saturated final mean flow with its “nonsaturated” counterpart. It is implicitly assumed, however, that saturation is responsible for damping of each wave event. Furthermore, saturation causes the final mean flow in the “accelerated” phase speed problem to resemble Fig. 2 much closer than the nonsaturated accelerated phase speed solution (Coy, 1982) for in the latter case no limit is imposed on the mean flow beneath the critical level (Grimshaw, 1975).

As a word of caution, there may exist situations in which this approximation is not entirely suitable. For example, if a wave were to initially experience some wavebreaking at a lower level, then propagate away from the critical level and back toward it only to experience more wavebreaking, a permanent mean flow change would appear at the lower level of wavebreaking, which would thereby reduce the net change above. (Perhaps this is analogous to ocean wavebreaking at the edge of a coral reef?) This final mean flow would in this case differ more significantly from that implied by (1.2). Such a situation was, in fact, observed in Fig. 3 of Dunkerton (1982) over two narrow ranges of $B_0$ in that particular case.

c. Validity of the time-scale separation

Our separation of slow and ultra-slow time scales is motivated largely by the resulting simplicity. Supposing that mesospheric mean flows are diabatically induced on a time scale of a few days (Holton, 1982), dimensional values to $t_f$ are constrained to be less than O(1 day). This requirement appears best satisfied by relatively short-lived atmospheric forcings, such as the radiation of gravity waves from shear
layers (Fritts, 1982a). For those forcings of several days length, if they exist, it would be less appropriate to make such a scale separation. Instead, the interaction of these waves with a *diabatically-changing* mean flow will be important (Holton, 1982).

3. Specification of model parameters

a. Radiative equilibrium

A simplified form of relaxation to radiative equilibrium, following Holton (1982), is to adopt the mean potential vorticity equation in a beta-plane channel confining attention to the lowest-order mode $\bar{u} \propto \sin y$:

$$\frac{\partial}{\partial t} \left\{ \left( \frac{\partial}{\partial z} - \frac{1}{H} \right) \bar{u}_z - \frac{f^2 N^2}{f^2} \bar{u}_z \right\} + \left( \frac{\partial}{\partial z} - \frac{1}{H} \right) \alpha (\bar{u}_z - \bar{u}_z^R) = 0. \quad (3.1)$$

It should be noted, however, that there is no entirely satisfactory way of modeling the interaction in one dimension, since: 1) the steep shear zones observed by Jones and Houghton (1971), Dunkerton (1981a, 1982), Fritts (1982b), and Coy (1982) may not be geostrophic; 2) it may be inappropriate to prespecify the cross-channel dependence of the mean flow evolution; and 3) the gravity-wave forcings might also exhibit such a $y$-dependence.

In (3.1), $l = 3a^{-1}$, $a$ is the radius of the earth, $f = 1.03 \times 10^{-4} \text{ s}^{-1}$, $N = 0.02 \text{ s}^{-1}$, and $\tau = 1$ year:

$$\bar{u}_z^R(z, t) = (z - 20) \left\{ 17.5 \cos \frac{2\pi t}{\tau} + 9.5 \bigg| \cos \frac{2\pi t}{\tau} \bigg|^3 \right\} + 10 \text{ m s}^{-1}, \quad (3.2a)$$

$$\alpha(z) = 1.5 \times 10^{-6} \text{ s}^{-1} \left\{ 12 + \tanh \frac{z - 50}{12} \right\}, \quad (3.2b)$$

which are similar to Holton's (1982) except that $\alpha$ is reduced by a factor of 2 and the radiative equilibrium profile is relatively stronger in winter and experiences a more sudden equinoctial transition.

b. Smoothing of steep shear zones

Eq. (1.2) implies a discontinuity in the mean flow, or "shock," following the gravity wave event. In re-
ality, discontinuities of this sort may not exist, or may not last, for a variety of reasons. These include: 1) wave saturation; 2) eddy diffusion; 3) the uncertainty principle; and 4) zonal mean transmissivity. The last reason could turn out to be the most important one, and the most likely to be overlooked. Further discussion is given in Section 5 below.

In the model, computational stability was aided by the inclusion of a diffusive smoothing routine designed to mimic the above effects. After each gravity wave event the mean flow above the stratopause was adjusted in a momentum-conserving manner to render it continuous. However, the effect of the adjustment was negligible outside \( z_f \pm 5 \) km, and the adjusted mean flow was virtually identical in appearance to Fig. 1.

c. Simplest stochastic model

In the simplest model the transient gravity wave, mean-flow interaction can be described with two wave parameters: the phase speed \( c \), and the time-integrated momentum flux across the lower boundary \( \Delta M \). In general, the momentum flux involves several other wave parameters including zonal wavenumber, angle of orientation, intrinsic phase speed in the source region, static stability, forced wave action amplitude, and length of forcing. On the ultra-slow time scale a third relevant parameter is \( \Delta T \), the length of the quiet interval between gravity wave events. While any or all of these parameters could be explicitly determined, the approach taken here is one of maximum simplification: I consider, for each day of the year, \( n \) equally spaced pairs of gravity wave events. One wave has a positive phase speed with an associated positive \( \Delta M \), while the other wave is of stationary or negative phase speed, and with an associated negative \( \Delta M \) (note that the sign of \( \Delta M \) is determined by the intrinsic phase speed in the source region). The procedure is to determine either the phase speed or the momentum flux of each wave with a pseudo-random number generator and probability distribution.

To motivate this approach, let us generalize slightly the discussion of Lindzen (1981). Lindzen speaks of “permitted phase speeds” during summer and winter. In summer, according to Lindzen, the relevant waves are those of phase speed \( \sim 20 \) m s\(^{-1} \) or greater. These waves are presumably excited in connection with the upper tropospheric jet. The stationary waves, together with other waves of phase speed less than that of the jet, are not relevant in view of the distribution of critical levels in the troposphere and lower stratosphere. In winter, on the other hand, the stationary waves and those of negative phase speed do not encounter critical levels in the troposphere and are therefore relevant in the mesosphere. At this time, those 20 m s\(^{-1} \) waves, important in summer, are absorbed in the lower stratosphere, and are therefore unimportant.

In viewing the middle atmosphere as a whole (and also when discussing equinoctial and equatorial flows), it might be preferable to clarify that a “permitted” wave phase speed is really a “relevant” phase speed for the mesosphere. Actually all reasonable values of phase speed are being generated at one time or another in the atmosphere. In extratropical latitudes, we might crudely consider a Gaussian distribution of phase speeds about some stationary or weak westerly values, say, which is independent of season (Matsuno, 1982). Now those waves having phase speeds near the center of the distribution will be the least important in the mesosphere, being significantly absorbed in the lower atmosphere. Only those waves whose phase speeds lie in the wings of the Gaussian will be important in the mesosphere. At solstice conditions, the relevant phase speeds are opposite to the sense of the stratospheric flow, while at the equinoxes or at the equator, both wings could prove important.

For the purposes of numerically modeling an annual cycle, then, it is necessary to force both westerly and easterly waves, although only one will be operating in the mesosphere at the solstices. Furthermore, it was found desirable to eliminate those waves having phase speeds near the value of the zonal wind at the lower boundary. In the model, \( \bar{u} \) is held fixed at 10 m s\(^{-1} \) at 20 km, the lower boundary. In cases where the momentum flux was specified, the following distributions of phase speed were chosen:

\[
c = c_0 \ln(1 + x),
\]

\[
c = c_s - c_0 \ln(1 - x),
\]

for the negative and positive phase speed waves, respectively, and \( x \) is a pseudo-random number on the interval (0, 1). Eqs. (3.3a,b) define a probability distribution \( P(c) = dx/dc \). Noting, however, that the functions in (3.3) are exponential, the \( P(c) \) are also exponential, and may be viewed in Fig. 3 by making the replacement \( P = 1 - x \). One could regard these distributions as similar to a Gaussian with its center removed.

A pseudo-random number generator is

\[
x_{n+1} = \text{frac}\{x_n + \pi\}^5.
\]

For each gravity wave event, (3.4) determines \( x \), which determines \( c \) via (3.3a,b), which in turn gives \( z_f \) and, via (1.3), \( z_f \). Experiments have been performed with various spectral widths and values of \( \Delta M, \Delta T \), and also with constant \( c \) and \( \Delta M \) determined stochastically.

The upper boundary condition is to set \( \partial \bar{u}/\partial z = 0 \). A finite difference grid is used with

\[
\Delta z = 1 \text{ km},
\]

\[
\Delta t = 1 \text{ h},
\]
and the density scale height is uniformly 7 km. The model is initialized with $\tilde{u}(z) = 10$ m s\(^{-1}\).

Note that if no critical level exists before a given gravity wave event, (1.3) must be truncated to the top level of the model, which is taken to be 120 km. For realistic values of $\Delta M$, however, the error in this approximation is extremely small since

$$\rho_0(120) = \exp - \frac{1000}{\gamma} = 6.25 \times 10^{-7}. \quad (3.6)$$

4. Results and discussion

The model was integrated under three conditions: perpetual winter, perpetual summer, and an annual cycle. For perpetual conditions only the relevant half of the $P(c)$ distribution was retained.

a. Perpetual winter

The winter radiative equilibrium profile in Eq. (3.2a) is

$$\tilde{u}^R(z) = 3z - 50. \quad (4.1)$$

Model runs of 1000 days length were not found very sensitive to the initial pseudo-random number.

In this and the following subsection two statistics are plotted: the average zonal wind after gravity wave events $\tilde{u}^A$, and the average zonal wind before gravity wave events $\tilde{u}^B$. A number of experiments for different values of $n$, the number of events per day, where

$$n\Delta T = 1 \text{ day}, \quad (4.2a)$$

$$n\Delta M = \Delta M_{\text{tot}}, \quad (4.2b)$$

revealed that when $\Delta M_{\text{tot}}$ is fixed the average values $\tilde{u}^B$ were largely dependent upon $n$, while the $\tilde{u}^A$ were not. This is because values of $\tilde{u}^B$ were determined by the amount of time for radiative equilibrium to establish itself. On the other hand $\tilde{u}^A$ was not dependent upon $n$ because, even though higher $n$ values implied lower $\Delta M$, $(c - \tilde{u}_0)$ was approximately proportional to $n^{-1}$ (i.e., proportional to the time allowed for radiative forcing to act).

In a highly idealized model such as this one it is inappropriate to draw specific conclusions from actual values of $n$, but the general remark could be made that in the mesosphere the mean wind can easily undergo significant day-to-day variations under the competing influences of gravity wave stresses and the diabatic circulation. It seems significant that the

![Diagram showing zonal wind profiles](image)

**Fig. 4.** Time-averaged mean zonal flows for perpetual winter.
steady wave models of Holton (1982) and Matsuno (1982) cannot differentiate the observed dependence on event frequency in the transient wave model.

Unless otherwise specified, the remaining experiments to be discussed involve a single event per day. The preceding remarks concerning the sensitivity of $\tilde{u}^g$ should, however, be kept in mind.

Fig. 4 displays the time-mean statistics for the perpetual winter case with spectral width $c_0 = 10$ m s$^{-1}$ and fixed $\Delta M = -245$ m$^2$ s$^{-1}$. The upper mesospheric and lower thermospheric flow is slightly negative in accord with the narrowness of the $P(c)$ spectrum in this case. For this value of $\Delta M$, wave-breaking (approximately indicated by $z_f$) occurs as low as 60 km. The stratospheric flow is significantly less than radiative equilibrium due to the mean meridional circulation (Holton, 1982). As anticipated from the theory of gravity wave, mean-flow interaction a strong shear zone appears in the range 65–70 km. Because planetary waves and the semidiurnal tide are present in the total atmospheric flow, it is doubtful that such strong shear zones could become organized on a global scale (Section 5).

Fig. 5 shows the corresponding case with fixed $\Delta M = -90$ m$^2$ s$^{-1}$ (reduced by a factor of $e$). As expected, the shear zone now appears about one scale height higher. The jet is also stronger by about 30 m s$^{-1}$. The lower thermospheric flow is similar to before.

Fig. 6 displays the broader spectrum case $c_0 = 30$ m s$^{-1}$ and fixed $\Delta M = -245$ m$^2$ s$^{-1}$. Below 65 km things are much the same as in Fig. 4. At higher levels, however, the flow asymptotes towards a much larger easterly value. These and other experiments indicated that the height of the gravity wave, mean-flow interaction was controlled largely by $\Delta M$, while the magnitude of the upper level flow was determined by $c_0$. A stationary wave experiment, for example, generated very weak westerlies at upper levels ($c_0 = 1$ m s$^{-1}$, not shown). For the cases we considered, the spectral width did not significantly affect the height of interaction, while $\Delta M$ did not significantly affect the magnitude of the upper level flow.

It is unclear whether observations support the notion of a strong easterly flow as shown in Fig. 6 and also as modeled by Matsuno (1982). The overall reduction with height of the polar night jet to weak westerly and easterly values tends to indicate the relevance of waves having phase speeds at or close to the stationary value $c = 0$.

As a variation of this experiment, $c$ was held fixed at $c = 0$, i.e., stationary waves, while $\Delta M$ was determined stochastically by the formula

$$\Delta M = 245 \text{ m}^2 \text{ s}^{-1} \ln(1 - x),$$

\[4.3\]
with the restriction that $|\Delta M| \leq 670 \text{ m}^2 \text{ s}^{-1}$. As shown in Fig. 7, wavebreaking now occurs occasionally throughout the mesosphere and the time-averaged shear zone is more realistic. This is due to the fact that smaller (larger) wave amplitudes saturate at higher (lower) levels. This result seems to indicate that mesospheric wavebreaking cannot generally occur as low as the 50 km height suggested by Lindzen (1981).

b. Perpetual summer

The summer radiative equilibrium profile in Eq. (3.2a) is

$$\tilde{u}^R(z) = 50 - 2z,$$

and is chosen to agree with the expectation that summer latitudinal temperature gradients are less than those in winter. Fig. 8 displays the time-average statistics with $c_0 = 10 \text{ m} \text{ s}^{-1}$ and fixed $\Delta M = +33 \text{ m}^2 \text{ s}^{-1}$. The phase speed spectrum of Eq. (3.3b) has generated moderately strong westerlies above the mesopause. Regardless of the value of $c_0$, the magnitude of these westerlies (with sufficient $\Delta M$) will always be at least $c_0$; therefore, the present gravity wave mechanism might be explicitly tested by observations. Fortunately, there is somewhat better evidence of the zonal wind reversal in the summer lower thermosphere.

The converse of this result is that, given such moderately strong westerlies above the mesopause, the importance of stationary waves in the summer upper atmosphere is thereby diminished, in agreement with our expectations. It will be very interesting to see whether future observations confirm this point. Previous observations of noctilucent clouds (Fogle and Haurwitz, 1966) suggest nonzero gravity-wave phase speeds.

c. Annual cycle

The model was also integrated with the time dependence in Eq. (3.2a) and with both wave forcings simultaneously. To aid computation, gravity wave events below 30 km were ignored. Fig. 9 shows a time–height cross section of $\tilde{u}$ for the case $n = 1$, $c = 0$ and 20 m s$^{-1}$, $\Delta M_{w0} = -245 \text{ m}^2 \text{ s}^{-1}$, and $\Delta M_{s0} = +33 \text{ m}^2 \text{ s}^{-1}$. Model output was smoothed with a ten-day running mean.

The results compare fairly well to the perpetual condition cases. The time–height cross section has the advantage of showing the temporal variability. In both summer and winter this variability is most prominent in the regions of wavebreaking. In the transient wave model the equinoctial transition at

---

Fig. 7. As in Fig. 4, but for $c = 0$ and $\Delta M$ determined stochastically.

Fig. 8. As in Fig. 4, but for perpetual summer.
upper levels is quite sudden and first occurs when the opposite gravity waves are able to “sneak” through the flow profile at lower levels. Just prior to this transient there is a tendency in the lower thermosphere for easterlies (westerlies) to form before the spring (autumn) equinox, a behavior more evident in Matsuno (1982). This is due to the absence of a diabatic circulation to counteract the gravity wave acceleration.

In general, the transient variability of the upper mesosphere will depend upon the number of events per day and other factors. However, there is this kind of variability in observations, part of which could also be attributed to variations in planetary wave activity.

d. Other experiments

A straightforward generalization of this model would be to include planetary waves forced from below. There are at least two good reasons for such an effort; first, a critical level for stationary Rossby waves is implied by the model results in the winter upper mesosphere; second, low-level reversals in the mean zonal wind in connection with sudden warmings should theoretically affect vertical gravity wave propagation, as Lindzen (1981) notes.

Although this point is still speculative, it would not be surprising if gravity wave propagation were to have an effect on the occurrence of sudden warmings! We note the tendency for departures from radiative gravity wave, mean flow interaction in the present model; therefore, given the sensitivity to $\tilde{u}^R$ observed by Holton and Dunkerton (1978), the occurrence of vacillation, including warmings, should be affected by the background state.

5. Comments on “zonal mean transmissivity”

There is a fundamental point that needs to be made in connection with one-dimensional models of vertical gravity wave propagation, including the present model. While in a one-dimensional model the mean wind is considered to be a zonal mean, it must be recognized that small-scale gravity waves will not necessarily “see” the zonal mean wind, but are instead affected by the local wind profile. In the jargon of wave physics the following holds true:

**PRINCIPLE:** The zonal mean transmissivity of gravity waves of a certain phase speed is not, in general, equal to the transmissivity of the same waves in the zonal mean wind profile.

For transient, conservative waves in three dimensions what is then relevant is the three-dimensional critical surface

$$u(x, y, z) = c.$$  \hspace{1cm} (5.1)

It follows that

**COROLLARY:** Phase speeds not permitted in the zonal mean profile may nevertheless experience sig-
nificant transmission through the zonal mean critical level, \( \bar{u}(y, z) = c \).

One application of these results is that the effective critical layer absorption shown in Fig. 1 would, in a statistical sense, be spread over a broader range of heights, thus smoothing out the seemingly unrealistic shocks simulated in the stochastic transient wave model. This is significant, for it is then not necessary to invoke the uncertainty principle, i.e., the spectral broadening due to extremely transient and/or localized forcings, in order to explain any lack of steep shear zones in zonal mean wind profiles (Holton, 1975; Fritts, 1982b). On the other hand, local mean winds can exhibit strong vertical shears consistent with theory (Manson and Meek, 1976).

Clearly one must not be too rigid in the interpretation of permitted phase speeds. Probability distributions of phase speed should therefore not contain sudden cutoffs, although this was done in our Fig. 3 for simplicity. This point may have some special significance for the vertical wave propagation at the equator, which is the topic of the next section.

6. Tropical model

The gravity-wave problem in the tropics is indeed a difficult one. In addition to an annual cycle, the zonal mean flow undergoes a complicated periodic motion consisting of a quasi-biennial oscillation (QBO) in the lower stratosphere and a semiannual oscillation above (e.g., Holton, 1975; Hirota, 1980; and references therein).

Although Lindzen (1981) discussed gravity wave saturation in the extratropical mesosphere and tidal saturation in the tropical lower thermosphere, there is no reason to exclude gravity waves from the tropical mesosphere and lower thermosphere. Perhaps these waves are more prevalent here than elsewhere. Monsoonal and other convective activity can be added to the usual list of forcing mechanisms for these waves.

According to Lindzen (1981) the diurnal tide breaks above the tropical mesopause (>85 km). The tidal wavebreaking level ascends steeply away from the equator (K. Hamilton, personal communication, 1981). Below the equatorial mesopause the tidal-induced mean flow accelerations are relatively small (Fels and Lindzen, 1974). In an attempt to discuss gravity waves in the tropical upper mesosphere, then, it seems justifiable to ignore the diurnal tide, at least in our initial experiments. In this section, a variation of the stochastic transient wave model is designed and described for tropical conditions. It will be shown that this model provides a plausible explanation of the mesopause semiannual oscillation observed by Hirota (1978).

a. Stratospheric flow

In an attempt to model the low-level flow of the model semi-realistically, a mean flow equation of the following form has been used:

\[ \bar{u}_x + \frac{1}{\rho_0} \frac{\partial}{\partial z} \rho_0 (B_+ + B_-) = \frac{1}{\rho_0} \frac{\partial}{\partial z} \rho_0 \frac{\partial \bar{u}}{\partial z} + G(z, t). \]

(6.1)

Two forcing mechanisms enter this equation. The first is a simplified version of the Holton-Lindzen (1972) equation of the quasi-biennial oscillation:

\[ \rho_0 B_+ = B_{20} \exp \left( \int_{z_0}^z \frac{\alpha N}{k(c_e - \bar{u})} dz \right), \]

(6.2a)

\[ \rho_0 B_- = B_{20} \exp \left( \int_{z_0}^z \frac{\alpha N}{k(c_e + \bar{u})} dz \right), \]

(6.2b)

where

\[ B_{20} = 1.2 \times 10^{-2} \text{ m}^2 \text{ s}^{-1}, \quad \alpha = 5 \times 10^{-7} \text{ s}^{-1}, \]

\[ c_e = 25 \text{ m s}^{-1}, \quad N = 0.02 \text{ s}^{-1}, \quad k = a^{-1} \]

and, partly following Lindzen (1981),

\[ \nu = \begin{cases} 200 \exp \frac{z - 85}{7} + 0.3 \text{ m}^2 \text{ s}^{-1}, & z < 85 \\ 200.3 \text{ m}^2 \text{ s}^{-1}, & z \geq 85 \text{ km}. \end{cases} \]

This is a purely symmetric forcing as in Plumb (1977). In reality the oscillation decays with height in the middle stratosphere due to a wavelength-dependent damping rate (Hamilton, 1981; Fels, 1982). It is not our aim to simulate a realistic QBO here; instead, we want a periodic low-level flow not a rational multiple of the semiannual oscillation period above. Following Holton and Lindzen (1972) a pre-specified semiannual oscillation is forced above 30 km which in this case is of the form

\[ G(z, t) = \alpha (\bar{u}_{sa} - \bar{u}), \quad z > 30 \text{ km}, \]

(6.3a)

\[ \bar{u}_{sa} = U_{sa} \cos(2\pi t/180 \text{ days}), \]

(6.3b)

\[ U_{sa} = 25 \text{ m s}^{-1} \times \left( 1 + \tanh \frac{z - 36}{4} \right). \]

(6.3c)

Explicit calculation of the semiannual oscillation (Dunkerton, 1979) was avoided here, but further discussion will be given in a future paper.

In the interest of simplicity the annual cycle, together with the relatively small time-mean easterlies, were ignored here although the inertial stability of the tropical winter mesosphere appears to be a relevant issue (Hunt, 1981; Dunkerton, 1981b). (This difficult question probably should not be addressed without extreme caution.)
b. Gravity waves

Lacking any precise observational data a symmetric forcing was used of the form

$$\Delta M = 33 \text{ m}^2 \text{s}^{-1} \ln(1 - x);$$
$$c_+ = -25 \text{ m s}^{-1}, \quad (6.4a,b)$$
$$\Delta M_+ = -33 \text{ m}^2 \text{s}^{-1} \ln(1 - x);$$
$$c_+ = +25 \text{ m s}^{-1}, \quad (6.4c,d)$$

which might be regarded as statistically analogous to a standing wave forcing (Plumb, 1977). Easterly and westerly components, however, will separate with height due to selective transmissivity (Matsuno, 1982). The phase speed is of particular interest, thus $c$ was set to a constant value and one event per day was employed. The latter seemed acceptable since the basic state required a couple of weeks to develop; however, by itself it is not expected to be realistic. In order to guarantee the low-level flow, gravity wave events below 40 km were ignored. In general, one might wish to include the amplitude of these waves in the quasi-biennial forcing term.

c. Results and discussion

Because the quasi-biennial oscillation experiences almost a 180° phase tilt with height, this has the effect of excluding all waves of phase speed

$$|c| < \max(\bar{u}_{\text{QBO}}), \quad (6.5)$$

except during a brief "window" around the time of QBO phase reversal. Therefore, as shown in Fig. 10, when the amplitude of the semiannual oscillation is greater than that of the quasi-biennial oscillation, it is possible to achieve a semiannual oscillation at upper levels whose phase is opposite to that of the forced semiannual oscillation. Model data was smoothed with a monthly running mean. For this value of $\Delta M$, wavebreaking and the associated mean flow reversal is observed in the region above about 70 km.

Fig. 10 does not show any quasi-biennial oscillation at upper levels since $|c|$ is greater than the QBO amplitude. There is limited observational evidence of a quasi-biennial signal at upper levels (Belmont and Nastrom, 1979). Although this is a conjecture on my part, it is conceivable that selective transmissivity could provide for such a quasi-biennial component above the mesopause, because waves of phase speed

$$|c| < \max(\bar{u}_{\text{QBO}}), \quad (6.6)$$

are allowed to propagate vertically around the time of phase reversal. More generally, waves having phase speeds close to, or slightly greater than, the QBO amplitude would be subject to either saturation (Holton, 1982) or viscous damping (Matsuno, 1982) during their respective phases of the quasi-biennial oscillation. Such a modulation was observed by Dunkerton (1979) in connection with the stratopause semiannual oscillation. The latter effect could, of course, further affect the mesopause oscillation. Whether this mechanism is sufficient to generate a QBO signal of adequate amplitude at upper levels remains to be determined.

d. Diurnal tide

As a variation of the tropical experiment a fixed diurnal tidal forcing component was included of the form

$$\bar{u}_t = -16 \text{ m s}^{-1} \text{ day}^{-1} + \cdots z \geq 85 \text{ km}, \quad (6.7)$$

as suggested by Lindzen (1981). Results of this experiment, shown in Fig. 11, differed from those of Fig. 10 in that easterlies were much stronger above the mesopause. As noted by Lindzen (1981), diffusion associated with the breaking diurnal tide would remove most of the acceleration in (6.7). But there is, nevertheless, a residual component which would cumulatively produce a strong, and possibly barotropically unstable, easterly jet above the mesopause. This result, incidentally, depends upon the diffusion being written in momentum conserving form. Also, no other thermospheric processes are incorporated in the present model, which would presumably act to reduce the easterly flow at the uppermost levels.
The westerly gravity waves easily "erase" these residual easterlies at semianual intervals. On the basis of this experiment, and others, it seems that the breaking diurnal tide could explain the observed time-mean easterly flow around the mesopause observed by Hirota (1978, and personal communication). On the other hand, the observed westerly time-mean flow in the lower mesosphere could be explained by the Kelvin wave which drives the stratosphere semiannual oscillation (Dunkerton, 1979). This will be further explored in a forthcoming paper.

7. Conclusion

This paper has demonstrated, first, that there is a very good agreement between the mesospheric circulations predicted by this model and those of Matsuno (1982) and Holton (1982), despite the widely differing model formulations. Provided that the spectra of phase speeds and momentum flux are similar, model results are also similar. The reason for this is essentially found in our Fig. 2. It may be traced to the fact that the predicted gravity wave accelerations at upper levels are initially so large (Lindzen, 1981) that, in effect, what one sees at upper levels is a rapid evolution of the mean flow toward the gravity wave critical level, regardless of whether the waves are dissipated by diffusion (Matsuno, 1982), saturation (Holton, 1982), or neither (Dunkerton, 1982). In other words, the effective critical layer absorption is implied by all models and it is therefore not sur-

prising that the simplest "nonsaturated" solution is able to capture the essential physics of this problem, at least in cases relevant to the mesosphere.

In retrospect, this good agreement might help to justify the use of steady gravity-wave forcings in some general circulation models. At the very least, proper simulation of the mesospheric wave forcing would require some $O(20 \text{ m s}^{-1})$ waves in summer, in addition to the stationary waves of winter.

For numerical modellers seeking even greater simplicity, with less concern for rigor, it is possible that a "generalized Rayleigh friction"

$$\vec{u}_t = \alpha_M(z) \cdot \{ \vec{u}^A(z) - \vec{u}(z) \},$$

may be useful for perpetual midlatitude solstice conditions. Quantitative consideration of this parameterization is further discussed in the Appendix.

It was also demonstrated in this paper that the mesopause semianual oscillation first observed by Hirota (1978) could be viewed as a selective transmission effect involving gravity waves. Both easterly and westerly gravity waves might be important, at semianual intervals, in depositing easterly and westerly momentum around the mesopause. Superimposed on this oscillation is an easterly and westerly time-mean flow in the upper and lower mesosphere, respectively, which might be explained by the breaking diurnal tide and the Hirota-Kelvin wave, respectively.

In conclusion it should not be forgotten that gravity wave breaking in the tropical mesosphere is closely related to the inertial stability of the region (Hunt, 1981; Dunkerton, 1981). It remains to be seen whether inertial instabilities and breaking gravity waves are as mutually exclusive as is clearly implied by the marginal stability criteria on an equatorial beta-plane.

Acknowledgments. The author thanks K. Hamilton for helpful discussions on the tropical model and also thanks I. Hirota for providing some unpublished data on the mesopause semianual oscillation.

APPENDIX

A Generalized Rayleigh Friction

A parallel series of experiments was performed in which the explicit gravity-wave-induced mean flow accelerations were replaced with a generalized Rayleigh friction:

$$\frac{\partial}{\partial t} \left( \frac{1}{H} \frac{\partial}{\partial z} \left[ \left( \frac{\partial}{\partial z} - \frac{1}{H} \right) \vec{u}_z - \frac{F \bar{N}^2}{f^2} \vec{u}^A \right] \right) + \left( \frac{\partial}{\partial z} - \frac{1}{H} \right) \alpha(\vec{u}_z - \vec{u}^A) - \frac{F \bar{N}^2}{f^2} \alpha_M(\vec{u} - \vec{u}^A) = 0. \tag{A1}$$

Here it was assumed that the stochastic model-generated $\vec{u}^A$ in Figs. 4–8 could be taken as the relevant
equilibrium zonal wind in (7.1). The frictional drag coefficient was given the functional form

$$\alpha_M(z) = \alpha_M(z_{\ast}) \left| \frac{\bar{u}^d(z) - \bar{u}^{\ast}(z)}{\bar{u}^d(z_{\ast}) - \bar{u}^{\ast}(z_{\ast})} \right|, \quad (A2)$$

where $z_{\ast}$ is some reference height. Thus, $\alpha_M$ was essentially zero below the wavebreaking region. There are of course other ways of constructing this profile, but this method was best suited in the present case. It should be emphasized that a frictional drag parameterization is pragmatic but not rigorous. Furthermore it is not appropriate in the tropics, where the gravity-wave effect is not to be viewed as a drag.

Fig. 12 shows the steady solution to (A1) for the perpetual winter case corresponding to Fig. 4. We took $z_{\ast} = 120$ km and specified $\alpha_M(120)$ as either $(1$ hour$)^{-1}$ or $(1$ day$)^{-1}$. In the wavebreaking region both profiles of $\alpha_M$ were considerably greater than those used previously (Leovy, 1964; Schoeberl and Strobel, 1978; Holton and Wehrbein, 1980), but there are nevertheless significant differences in the steady-state mean flows in the two cases. For the faster damping rate, $\bar{u}$ is very nearly equal to $\bar{u}^d$ in the wavebreaking region. The maximum amplitudes of $\bar{u}$ and $\bar{u}^d$ are about equal, although the height of the maximum $\bar{u}$ is shifted downward about 5 km and the shear zone is consequently slightly weaker. However, the stratospheric flow exhibits a departure from radiative equilibrium in this case (as before). This is due entirely to the mean meridional circulation and has nothing to do with the frictional drag in this region which is essentially zero there.

The relatively slower damping rate $(1$ day$)^{-1}$ causes the steady state mean flow to follow the overall pattern of $\bar{u}^d$, but the amplitude of $\bar{u}$ is slightly stronger, and the mesospheric shear zone is significantly weaker.

As indicated in the figure, the model domain could be separated into three regions: a radiatively-dominated region (I), a frictionally-dominated region (III), and a transition region into which the mean meridional circulation extends (II). Importantly, the frictional drag coefficient need not extend into region II in order for the gravity waves to have a significant effect there.

This experiment was repeated for the broad spectrum case (Fig. 6), and for perpetual summer (Fig. 8) and the results were very similar in nature. Both of these cases differ significantly from the traditional use of Rayleigh friction (1.1) because the equilibrium profiles differ significantly from zero. Physically, gravity waves other than those of stationary phase speed are important in these examples.

A Rayleigh friction parameterization may or may not be appropriate for the equinoctes. In Matsuno's (1982) "wave effect" experiment, for example, easterlies and westerlies are observed to form in the lower thermosphere at the equinoctial transitions; in this case the gravity waves appear to act in a positive forcing sense (as opposed to a "drag"), much like our tropical experiment, and in this example a frictional drag parameterization would not be conceptually appropriate. On the other hand, in Holton's (1982) simulation this behavior was not as prominent, due to the much narrower wave spectra employed; here, it might suffice to "connect" the winter and summer drag coefficients and equilibrium profiles in some smooth manner.

It must be admitted, however, that at some point the use of a frictional drag parameterization begins to take on an artificial appearance when using it for other than perpetual midlatitude solstice conditions, and it is doubtful whether any real insight or benefits can be obtained from such an approach, other than the computational expediency of this method.

REFERENCES


Dunkerton, T. J., 1979: On the role of the Kelvin wave in the


