Propagation and Breaking at High Altitudes of Gravity Waves Excited by Tropospheric Forcing

JOSEPH M. PRUSA
Iowa State University, Ames, Iowa

PIOTR K. SMOLARKIEWICZ AND ROLANDO R. GARCIA
National Center for Atmospheric Research, Boulder, Colorado

Manuscript received 16 August 1995, in final form 5 February 1996

ABSTRACT

An anelastic approximation is used with a time-variable coordinate transformation to formulate a two-dimensional numerical model that describes the evolution of gravity waves. The model is solved using a semi-Lagrangian method with monotone (nonoscillatory) interpolation of all advected fields. The time-variable transformation is used to generate disturbances at the lower boundary that approximate the effect of a traveling line of thunderstorms (a squall line) or of flow over a broad topographic obstacle. The vertical propagation and breaking of the gravity wave field (under conditions typical of summer solstice) is illustrated for each of these cases. It is shown that the wave field at high altitudes is dominated by a single horizontal wavelength, which is not always related simply to the horizontal dimension of the source. The morphology of wave breaking depends on the horizontal wavelength; for sufficiently short waves, breaking involves roughly one half of the wavelength. In common with other studies, it is found that the breaking waves undergo "self-acceleration," such that the zonal-mean intrinsic frequency remains approximately constant in spite of large changes in the background wind. It is also shown that many of the features obtained in the calculations can be understood in terms of linear wave theory. In particular, linear theory provides insights into the wavelength of the waves that break at high altitudes, the onset and evolution of breaking, the horizontal extent of the breaking region and its position relative to the forcing, and the minimum and maximum altitudes where breaking occurs. Wave breaking ceases at the altitude where the background dissipation rate (which in our model is a proxy for molecular diffusion) becomes greater than the rate of dissipation due to wave breaking. This altitude, in effect, the model turbopause, is shown to depend on a relatively small number of parameters that characterize the waves and the background state.

1. Introduction

For over 100 years, observations of noctilucent clouds (NLCs) have revealed the existence of wave-like structures propagating at mesopause altitudes (Gadsden and Schroder 1989). These NLCs exhibit several distinct morphologies, not all of which show wave-like motions. The most distinctive wave-like structures are known as bands. Bands typically display horizontal wavelengths of 10–75 km, periods of up to several hours, and wave speeds of 10–40 m s⁻¹. The waves generally move independently of the NLC display (and even against the prevailing winds). They have spanwise extents of several hundreds of kilometers and an almost two-dimensional structure. Intersecting groups of bands have been observed that give rise to bright knots where the waves cross (World Meteorological Organization, 1970).

Radar studies conducted at Jodrell Bank, England (Greenhow and Neufeld 1960), and elsewhere in the late 1940s and throughout the 1950s examining the ionization trails of meteors also indicated wave-like variations in winds at heights of 80–100 km. Typically, horizontal scales of 50–300 km and periods of one-half to several hours were observed for mean wind speeds of 10–40 m s⁻¹. Mean wind shears of 10 m s⁻¹ km⁻¹, with maximum shears 10 times greater, as well as rms turbulent velocities of 25 m⁻¹ were also observed. Wave activity was found to drop off rapidly at altitudes below 80 km and above 100 km. In March of 1959 an optical study of a sodium vapor trail released by a rocket as it traversed the region from 80 to 120 km (Blamont and de Jager 1962) corroborated other information about the atmosphere at these altitudes gleaned from the earlier radar studies. In particular, the sodium trail showed turbulent diffusion with a linear rate of expansion of about 3.3 m s⁻¹ and turbulent eddies ranging from 0.1- to 0.5-km diameter below an altitude of 102 km. Above this altitude, the trail was very smooth and expanded at a rate proportional...
to the square root of time—a rate that is consistent with molecular diffusion. The boundary between the turbulent and diffusive regions, the turbopause (Whitten and Poppoff 1971; Andrews et al. 1987), was rather sharp and no deeper than about 1 km. Blamont and de Jager (1962) concluded from elementary Reynolds and Richardson number analyses that the sudden cessation of turbulence was due to increasing molecular diffusion and not to wind shear. Additional rocket soundings beginning in the 1960s revealed wavelike structures in the winter mesosphere temperature field (Theon et al. 1967). Temperature oscillations with amplitudes of 30–40 K and vertical wavelengths of 10–20 km were observed in the altitude range of 50–90 km. Winter wave activity was observed to increase with latitude. By comparison, wave activity in the summer mesosphere was greatly reduced, with variations of no more than 6 K from the mean state occurring only above 75 km altitude. The rocket soundings produced one other remarkable observation—that the summer polar mesopause was extraordinarily cold, with a temperature of about 130 K. This temperature is some 40 K colder than what would be expected on the basis of radiative equilibrium alone (Wehrbein and Leovy 1983). In contrast, the winter polar mesopause is anomalously warm.

More recent studies using ground-based, Doppler radar measurements (Vincent and Reid 1983) have observed waves in the range of 70–98 km altitude at southern middle latitudes. For mean zonal winds of 10–20 m s⁻¹, a distribution of waves was detected with horizontal wavelengths ranging from just under 50 to nearly 200 km and corresponding periods of 13 to over 200 min. For waves with periods less than 1 h, the mean horizontal wavelength and phase velocity were 70 km and 70 m s⁻¹, respectively. Analysis of the velocity perturbations, averaged over 41 days of observations, revealed eddy vertical fluxes of horizontal momentum, \( \langle u'w' \rangle \), at mesopause altitudes of up to 1.0 m² s⁻² and an Eliassen–Palm flux divergence, \( \rho^{-1} \nabla \cdot \mathbf{F} \), of 10–20 m s⁻¹ day⁻¹ (deposited against the mean zonal flow). Radar studies have further indicated that considerable wave activity may extend up to at least 110-km altitude at high southern latitudes (Phillips and Vincent 1989). Still other observations have revealed the existence of enhanced regions of radar echo power with vertical scales ranging from 1 to 3 km (Czechowsky et al. 1979) at mesospheric heights. These echoes exhibit lifetimes of 1–3 h and show little if any vertical motion. Similar phenomena with vertical scales of 5–15 km have also been reported (Balsley et al. 1983). They may be attributable to regions of turbulence (Walterscheid 1984). Photographic airglow studies at temperate latitudes (Moreels and Herse 1977) have also added to the archive of data on wavelike structures in the upper mesosphere and lower thermosphere. Typically, very regular waves with horizontal wavelengths of 30–70 km were observed, with group velocities from less than 5 up to 17 m s⁻¹. The periods of these waves varied from 50 to 100 min. The spanwise extent of the waves was frequently 600 km, indicating a high degree of two-dimensional structure.

All of these observational studies yield wave properties that are consistent with those of internal gravity waves, as predicted by linear stability theory (Bretherton 1966). More sophisticated arguments based upon linear theory have indicated that gravity waves (GW) could profoundly affect the dynamics of the atmosphere at upper-mesospheric altitudes. An influential paper by Lindzen (1981), based upon the Wenzel–Kramers–Brillouin (WKB) method, established that wave breaking could deposit momentum [as an Eliassen–Palm flux divergence; Eliassen and Palm (1961)] at mesopause altitudes at the rate of 50–100 m s⁻¹ day⁻¹. Through a simple Coriolis force balance, this implies a meridional velocity of about 5–10 m s⁻¹ at extratropical latitudes, in rough accord with observations (Nastrom et al. 1982; Fraser 1989; and Phillips and Vincent 1989). Lindzen also noted that gravity wave breaking was probably the dominant source of turbulence, and hence mixing, in the extratropical mesosphere and lower thermosphere. Garcia and Solomon (1985) incorporated Lindzen’s GW drag parameterization into a zonally averaged model of the mesosphere and lower thermosphere. At mesopause altitudes, the model’s results indicated that breaking gravity waves deposited enough momentum to produce a reversal in zonal winds. The corresponding meridional circulation was poleward for the winter hemisphere and equatorward for the summer hemisphere. By continuity, the equatorward meridional circulation was accompanied by rising and adiabatically cooling air at the polar mesopause, whereas for the winter hemisphere air descended and warmed at the polar mesopause. This circulation, driven by the momentum deposition of breaking gravity waves, correctly reproduced the observed reversal in solstice mesopause temperature gradient.

In order to analyze approximately the breaking process (which goes well beyond the limits of the WKB method), Lindzen (1981) used the hypothesis that wave amplitudes are limited by saturation. That is, when gravity waves produce a negative vertical gradient of potential temperature, \( \partial \theta / \partial z \), they become unstable and turbulent breakdown ensues. This turbulence hypothesis is one of several that have been put forward to explain the dynamics of breaking waves. Additional hypotheses that have been advanced incorporate various elements of dynamical instability (Kelvin–Helmholtz instability), wave transience, and nonlinear wave–wave interactions. These various mechanisms are well summarized in Dunkerton (1989) and Walterscheid and Schubert (1990), among others. Given the range of possible physical mechanisms at work in gravity wave breaking, Walterscheid and Schubert used a compressible, fully nonlinear numerical model to investigate the dynamics of...
gravity wave saturation and breakdown. Key features of the model were that the forcing consisted of a monochromatic traveling wave and that the basic state was at rest. Overturning was found to generate convective cells with unit aspect ratios and sizes on the order of 15 km. Another important result from the simulation was that the region of overturning prior to wave breaking was large but confined. In particular, the minimum altitude of overturning descended from an initial value of approximately 85 km (at the onset of overturning) to a final value of approximately 55 km at the end of the simulation.

Bacmeister and Schoeberl (1989) used an anelastic approximation in a numerical model to simulate the dynamics of breaking gravity waves. The anelastic approximation effectively filters out transient compressibility effects (sound waves) while allowing for the exponential variation in thermodynamic properties (density, pressure, and potential temperature) actually found in the atmosphere. Another key feature of the model was that a nonzero basic-state wind was used. Finally, the waves were forced using a fixed, bell-shaped ridge at the lower boundary, exciting a spectrum of waves. Bacmeister and Schoeberl’s simulation indicated that wave overturning resulted in a nonlinear, wave–wave interaction that produced large-amplitude, downward traveling waves from the region of overturning. In addition, the minimum altitude for significant wave breaking activity descended to just above the lower boundary late in the simulation. These results are in contrast with the localized, high-altitude breaking found by Walterscheid and Schubert (1990). They appear to be more consistent with phenomena associated with very large amplitude wave forcing (e.g., downslope windstorms) than with mesospheric wave breaking.

Numerous studies have revealed three-dimensional effects in shear-driven (Kelvin–Helmholtz type) wave breaking. Fully three-dimensional numerical computations by Andreasen et al. (1994) predict that spanwise instability is a major factor in wave breaking. The spanwise, or secondary, instabilities give rise to counterrotating vortices with axes of rotation oriented more or less parallel to the direction of wave propagation (and perpendicular to the billow rolls often seen in two-dimensional simulations). Three-dimensional effects are observed to cause rapid breakdown of unstable wave structures (within one Brunt–Väisälä period) and to reduce the degree of supersaturation of waves in the breaking process. Additional computations (Fritts et al. 1993) appear to model band and billow NLC formations observed at mesopause altitudes extremely well and thus provide support for the formation of spanwise instability. Billows are a second morphological type of wave-like NLCs. They are smaller and more ephemeral than band NLCs. Typical characteristics are wavelengths of 4–9 km, spanwise lengths of 10–40 km, and lifetimes of 6–24 min (Gadsden and Schroder 1989). When silhouetted against band NLCs, they often give the appearance of a comb. These numerical computations also appear to compare reasonably well with a three-dimensional stability study by Cauffield and Peltier (1994). Other studies (Thorpe 1987; Klaassen and Peltier 1991; Scinocca 1995; to name a few) have also contributed important experimental and theoretical information on the nature of the spanwise structures that form. However, our understanding is less than complete. Limited computational resources force decisions to be made even in theoretical studies to capture only some of the possible secondary motions observed in experiments. In particular, studies often set a domain length equal to one primary wave, whereas several secondary instability mechanisms are known to operate over more than one wavelength of the primary wave (Scinocca 1995).

The present work was undertaken in part to elucidate some of the physical mechanisms involved in the development and breaking of gravity waves and their effects upon the basic state of the atmosphere. In particular, the numerical model was designed to test the effects of forcing by “realistic” tropospheric sources, that is, sources that excite a spectrum of waves in both frequency and wavenumber. Thus, a main ingredient of the formulation is the use of a time-variable geometry transformation that makes it possible to specify any type of forcing function, provided it is sufficiently differentiable. An anelastic approximation is also applied in order to filter out sound waves from the solutions. The application of this approximation to the governing equations in a time-variable geometry requires a significant generalization in that the transient term in the continuity equation is not zero. Finally, the transformed evolution equations are solved using an extremely efficient semi-Lagrangian method. The details of the model formulation are presented in sections 2 and 3 of the paper.

In our calculations waves are forced by flow over an obstacle, which can be thought of as representing a squall line or a topographic feature, depending on its longitudinal dimension. In order to capture the excitation, propagation to high altitudes, and dispersion of the wave field, a large domain (up to 720 km in the horizontal and 120 km in the vertical) is used. Furthermore, accurate modeling of the breaking process turns out to require rather fine spatial (<1 km) and temporal (<10 s) resolutions. The combination of a large domain size and fine spatial and temporal resolution requires large amounts of computer time to follow the development of the wave field from excitation at the lower boundary to breaking at mesospheric altitudes. For this reason, we have not explored three-dimensional effects in this study. We emphasize instead how forcing and vertical dispersion influence the wave field that propagates to high altitudes and what processes determine the upper and lower limits of the region of wave breaking. These issues are examined in detail in sections 4 and 5, where the results of a suite
of model calculations are discussed and given theoretical interpretation. In section 6, the effect of externally imposed damping and its implications for the altitude of the earth’s turbopause are considered. A summary and conclusions are presented in the last section.

2. Formulation

The atmospheric state is decomposed into a basic state $\Phi_0$ that is in hydrostatic equilibrium and a perturbation state $\Phi'$ that represents the departure of the atmosphere from this equilibrium. The decomposed variables $\Phi = \Phi_0 + \Phi'$ for $\Phi = p, \rho, T,$ and $\theta$, where $p$ is pressure, $\rho$ density, $T$ temperature, and $\theta$ potential temperature (the velocity does not have to be decomposed explicitly) are substituted into the equations of continuity, horizontal and vertical momentum, and energy, ignoring rotation and spherical geometry. From these equations the governing equations for hydrostatic equilibrium are subtracted off, leaving only perturbation equations for the disturbances from equilibrium.

2a. Governing equations in inertial coordinates

The potential temperature scale height of the middle atmosphere, $H_\theta$, is about 25 km (U.S. Standard Atmosphere, NOAA 1976; Houghton 1986) for solstice conditions at 40° latitude. Given the observed wavelengths of gravity waves summarized in the introduction, a characteristic vertical scale of $l_z = 2$ km may be inferred for these waves [using the inverse wavenumber $m^{-1} = \lambda_z/2\pi$ as the characteristic vertical scale (Lipps 1990)]. From this $l_z/H_\theta \approx 0.1$, and one may profitably employ the anelastic approximation of Lipps and Hemler (1982) to filter out sound waves from the model equations. This releases the model from a time step restriction not based upon the relevant physics of the problem. With the onset of wave breaking, still smaller structures become dominant and the approximation becomes even better. The full-perturbation equations that result are cast in Lagrangian form, as follows:

$$\frac{du}{dt} = -\frac{\partial p^*}{\partial x}$$

(conservation of horizontal momentum) \hspace{1cm} (2.1a)

$$\frac{dw}{dt} = -\frac{\partial p^*}{\partial z} + \frac{g\theta'/\theta_0}{\partial z}$$

(conservation of vertical momentum) \hspace{1cm} (2.1b)

$$\frac{d\theta}{dt} = 0$$

(conservation of energy) \hspace{1cm} (2.1c)

where $p^* = (p'/\rho) = c_\theta \theta \pi'$, where $\pi$ is the Exner function and $dL/dt = \partial L/\partial t + u'\partial L/\partial x'$. Since only the basic-state density $\rho_0$ appears in the formulation, the subscript "o" is dropped for convenience. From this point on, $\rho$ will denote the basic-state density. A fundamental difference between the present model equations and those of Lipps and Hemler (1982) is that in the present formulation the transient term (although here identically equal to zero) is not dropped from the continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u')}{\partial x'} = 0.$$ \hspace{1cm} (2.2)

All versions of the anelastic approximation share the feature that the basic-state density is not a function of time. The flow is, in a sense, locally steady. However, in the present formulation the geometry itself is unsteady. Thus, the transient term in the continuity equation must be retained in order to properly capture the effects of this unsteady (or time variable) geometry.

2b. Transformation to noninertial coordinates

Figure 1 illustrates the two-dimensional, unsteady geometry of the physical domain. The top of the domain is set at constant altitude $H$. The lower boundary is set at altitude $z_0(x, t)$, which is a function of both horizontal coordinate $x$ and time $t$. The vertical coordinate is denoted by $z$, and the horizontal extent of the domain is set to $\pm x_{max}$. The dependent variables to be determined are the horizontal and vertical components of velocity, $u$ and $w$, respectively; the potential temperature, $\theta$; and the normalized perturbation pressure $p^*$. Each of these dependent variables is a function of $(x, z, t)$.

Since the lower boundary may be both spatially irregular and variable in time, it is convenient to transform the physical domain into one that is constant and of uniform shape throughout the analysis. This can be

![Physical Model](fig1)

**FIG. 1.** Schematic diagram of the model geometry. Waves are forced by a time-dependent deflection of the lower boundary, $z_0(x, t)$.
accomplished with a simple algebraic stretching of the vertical coordinate. This stretching is the terrain-following coordinate transformation (Clark 1977) found in Gal-Chen and Somerville (1975). In the present study, the full generality of this stretching transformation is recognized in that it is also allowed to be a function of time. The precise form of this boundary can be quite arbitrary, provided that it is sufficiently smooth and that the inverse transformation exists (Gal-Chen and Somerville 1975; Prusa and Yao 1985). By allowing the lower boundary to vary in time as well as spatially, some control can be exercised over the frequencies as well as the wavenumbers of the gravity waves generated by the lower boundary forcing.

As time increases, gravity waves propagate away from the location where the forcing occurs. Thus a second stretching transformation $F(x, t)$ is introduced for the horizontal coordinate. The time dependence allows the horizontal extent of the physical domain to increase in time, while the spatial dependence allows the physical horizontal coordinate to increase more rapidly than the transformed coordinate as the lateral boundaries of the transformed domain are approached.

The coordinate transformation then takes the form

$$\begin{bmatrix} \tilde{t} \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} t \\ F(x, t) \\ \frac{z - z_s}{H - z_s} \end{bmatrix}, \quad (2.3a)$$

where

$$z_t = z_s(x, t) \quad (2.3b)$$

and the overbars denote the transformed coordinates. The forms for the boundary forcing function $z_s(x, t)$ and the horizontal stretching function $F(x, t)$ are given in section 3a.

Although the present study is two-dimensional, the anelastic approximation, the time variable geometry transformation, and the semi-Lagrangian solver used in this study are by no means restricted to two dimensions. The formulation is easily extended to three dimensions.

c. Governing equations in noninertial coordinates

The independent variables of physical space are expanded into those of the transformed space using the chain rule. The governing equations are then reorganized in terms of the contravariant velocities of the transformed space. Finally, fundamental geometrical identities such as $F_x = 1/f_x$ and $F_t = -f_t/f_x$, where $f = f(x, t) = x$ is the inverse function of $F(x, t)$, are used to simplify the resulting system of coupled equations. The results are

$$\tilde{u} = F_x u + F_t,$$

and

$$\tilde{w} = (G^{11} u + G^{-12} w)/F_x - \beta z_{s,s}, \quad \text{(velocities)} \quad (2.4)$$

$$\frac{d\tilde{u}}{dt} = -\left( G^{11} \frac{\partial \tilde{u}^*}{\partial \tilde{x}} + G^{12} \frac{\partial \tilde{u}^*}{\partial \tilde{z}} \right) + F_{tt},$$

$$+ \left( \tilde{u}^2 + F_t^2 \right) \frac{F_{xx}}{F_x^2},$$

$$+ 2 \left[ \left( \tilde{u} - F_t \right) F_{xx} - \tilde{u} F_t F_{xx} \right]$$

(conservation of horizontal momentum) \quad (2.5a)

$$\frac{d\tilde{w}}{dt} = \frac{\delta}{\theta_0 F_x G^{1/2}} - \left( G^{11} \frac{\partial \tilde{u}^*}{\partial \tilde{x}} + G^{22} \frac{\partial \tilde{w}^*}{\partial \tilde{z}} \right) - \beta z_{s,s},$$

$$- \left( \frac{\beta z_{s,s}}{F_x^2} \right) \left( \tilde{u}^2 + F_t^2 - 2\tilde{u} F_t \right) + \frac{2\tilde{w} z_{s,s}}{(H - z_s)} \frac{H - z_s}{H - z_s},$$

$$+ 2 \left( F_t - \tilde{u} \right) \left( \frac{\beta z_{s,s}}{F_x} - \frac{\tilde{w} z_{s,s}}{(H - z_s)} \right)$$

(conservation of vertical momentum) \quad (2.5b)

$$\frac{d\vartheta}{dt} = 0, \quad \text{(conservation of energy)} \quad (2.5c)$$

$$G^{-11} \frac{\partial (\rho G^{1/2})}{\partial \tilde{t}} + G^{-12} \frac{\partial (\rho G^{1/2} \tilde{u})}{\partial \tilde{z}} = 0$$

(mass continuity) \quad (2.5d)

where

$$\frac{d}{dt} = \frac{\partial}{\partial \tilde{t}} + \tilde{u} \frac{\partial}{\partial \tilde{x}} + \tilde{w} \frac{\partial}{\partial \tilde{z}}, \quad G^{11} = F_x^2,$$

$$G^{12} = -\beta F_x z_{s,s} = G^{21},$$

$$G^{22} = \beta^2 z_{s,s}^2 + \left[ H/(H - z_s) \right]^2, \quad \beta = \left( \frac{H - z_s}{H - z_s} \right),$$

and $G^{1/2} = (H - z_s)/HF_x$ is the Jacobian of the transformation.

The terms $F_x$, $G^{11}$, $G^{21}$, $G^{1/2}$, and $\beta$ appearing in the contravariant velocities all represent scaling factors accounting for the stretch in geometry from the physical to the transformed domain. The terms $F_t$ and $z_{s,s}$ represent grid speeds and give the relative motion of the transformed coordinates with respect to the physical coordinates. The continuity equation is actually in full tensor form (coordinate invariant form). Recall from section 2a that, despite the use of the anelastic approximation, the unsteady term of the continuity equation must be retained because of the time-variant nature of the geometry. While the thermodynamic energy equation appears the same, the momentum equations now contain a number of terms that are quadratic in velocity. These Christoffel symbol terms result from the curvature of the transformed geometry with respect to physical space. They represent apparent Coriolis and centrifugal accelerations, which fluid particles feel in the transformed space. All partial derivatives of $F$ and $z_t$,
appearing in the above formulation are with respect to the independent variables of physical space.

d. Semi-Lagrangian solution

The anelastic advection equations for momentum and potential temperature [Eqs. (2.5a–c)] are solved using the semi-Lagrangian approximation as discussed in Smolarkiewicz and Pudykiewicz (1992). An additional key feature of this approximation was the use of monotone, or nonoscillatory, interpolation (Smolarkiewicz and Grell 1992) on the advected fields in order to determine their values at the points of origination of particle trajectories at previous time steps. Monotone interpolation preserves the local topology of the flow, that is, it does not allow streamlines to cross. Without a monotonicity constraint, results in the nonlinear regime superficially resemble wave breaking, but closer inspection reveals intersecting streamlines—a nonphysical result. Ultimately, this leads to negative values of potential temperature, clearly showing that lack of monotonicity is equivalent to negative entropy production and a violation of the second law of thermodynamics.

Details of the trajectory computation and the interpolation scheme may be found in the references cited above. Each of (2.5a–c) was cast in the following form:

\[ \phi^{n+1} = \phi + R_\phi^{n+1} \Delta t/2, \]  

(2.6a)

where

\[ \phi + \phi_0 + R_{\phi,0} \Delta t/2, \]  

(2.6b)

where \( \phi = \bar{u}, \bar{w}, \) or \( \theta; \) \( n + 1 \) denotes the new time level at which a solution is sought; \( \Delta t \) is the time step size; and \( R_\phi \) denotes the right-hand sides of (2.5a–c). Subscripted variables \( \phi_0 \) and \( R_{\phi,0} \) denote the field variable and the corresponding right-hand sides of (2.5a–c) evaluated at the particle trajectory points of origination (determined from the previous time level \( n \) solution). For the momentum equations, \( R_\phi \) consists of pressure and buoyancy forcing terms and Christoffel symbols (apparent centrifugal and Coriolis accelerations). Equations (2.6a,b) are a second-order accurate, implicit trapezoidal rule integration. The pressure forcing terms at time level \( n + 1 \) are evaluated after solving the Poisson pressure equation (see section 2e). Estimates of the \( n + 1 \) time level velocities, needed for evaluating the Christoffel symbols appearing in \( R_\phi^{n+1} \) in (2.6a), were made using the explicit Euler method. The corresponding terms appearing at time level \( n \) in \( R_{\phi,0} \) [(2.6b)] were evaluated using the available second-order solution. Since \( R_\phi = 0 \), potential temperature is simply carried along with fluid particles. Backward integration of the particle trajectories followed by interpolation of the previous \( \theta \) field to the resulting points of origination immediately gives the \( n + 1 \) time level field: \( \theta^{n+1} = \theta_0 \).

The points of origination of the particle trajectories, \( x_0 = (x_0, z_0) \), were evaluated using a second-order Runge–Kutta scheme:

\[ x_0 = x^{n+1} - \bar{u}(x_m, t_m) \Delta t, \]

where

\[ (x_m, t_m) = \left( \frac{x_0 + x^{n+1}}{2}, \frac{t_0 + t^{n+1}}{2} \right) \]  

(2.7a)

and

\[ \bar{u}(x_m, t_m) = u(\bar{x}_0, t_0) + R_{\phi,0} \Delta t/2, \]

(2.7b)

where

\[ R_{\phi,0} = (R_{\phi,0,1}, R_{\phi,0,2}). \]  

(2.7b)

Here \( \bar{x}_0 \) is a first-order approximation to \( x_0 \), determined by using a simple Euler method to integrate backward in time from \( x^{n+1} = (x^{n+1}, z^{n+1}) \) using the velocity known from the previous time step, \( u_0 = (u_0, w_0) \). Note that (2.7b) is simply (2.6b) applied to the velocity vector.

The trajectory scheme (2.7) is a special case of the implicit midpoint rule, for which iterations are a fundamental part of the numerical process. Smolarkiewicz and Pudykiewicz (1992) demonstrate that the iterations converge provided the Lipschitz number \( L \) is less than one:

\[ L = \left\| \frac{\partial \bar{u}}{\partial x} \right\| \Delta t < 1, \]  

(2.8)

where \( \bar{u} \) and \( x \) are the velocity and position vectors. The Lipschitz number is the semi-Lagrangian counterpart of the Courant number for Eulerian solvers. The two iterations that occur in (2.7) suffice to ensure second-order accuracy.

The interpolation method used here is based upon the use of a nonoscillatory, constant coefficient, Eulerian solver. In the last decade such solvers have been studied and developed extensively, and have reached a high level of sophistication. The interpolator that draws from this experience (Smolarkiewicz and Grell 1992) is both accurate and fast. For the results generated in this paper, it routinely ran at nearly 200 Mflops using a single processor of a CRAY YMP (interpolations account for approximately two thirds of the computational operations).

e. Diagnostic pressure equation

The contravariant velocity vector, \( u = (\bar{u}, \bar{w}) \), is multiplied by \( \rho^* = \rho G^{1/2} \), and the divergence of this modified velocity field, \( \nabla \cdot (\rho^*u)^{n+1} \), is then computed. Expansion of \( u^{n+1} \) according to (2.6), followed by substitution into the continuity equation in noninertial coordinates (2.5d), ultimately leads to a Poisson equation for pressure:
\[
\rho^* \left( G^{11} \frac{\partial^2 p^*}{\partial z^2} + 2G^{12} \frac{\partial^2 p^*}{\partial x \partial z} + G^{22} \frac{\partial^2 p^*}{\partial x^2} \right) \\
+ \rho^* F_{xx} \frac{\partial \rho^*}{\partial x} \right]^{n+1} + \left( B_z \frac{\partial \rho^*}{\partial z} \right)^{n+1} \\
= \frac{2}{\Delta t} \left\{ \left( \frac{\partial \rho^*}{\partial t} \right)^{n+1} + \nabla \cdot \left[ (\rho^*)^{n+1} \bar{u} \right] \right\}, \quad (2.9)
\]

where

\[
B_z = \left[ \frac{\partial (\rho^* G^{12})}{\partial x} + \frac{\partial (\rho^* G^{22})}{\partial z} \right]^{n+1} \\
= \left[ \frac{1}{F_x} \frac{\partial \rho}{\partial z} - \rho^* \beta \left( z_{*,xx} + \frac{2z_{*,x}^2}{H - z_*} \right) \right]^{n+1}
\]

and \(\bar{u} = (\bar{u}, \bar{w})\) is the modified velocity vector defined by (2.7b). The term \(\left( \frac{\partial \rho^*}{\partial t} \right)^{n+1}\) in (2.9) represents explicitly the effects of the time variation of the transformed geometry. Essentially the density changes because the space is deforming. This term is evaluated in much the same way as \(B_z\); it is expanded into derivatives in terms of the physical coordinates using the chain rule. Additional coupling of the pressure field to the time variation of the geometry occurs through the pressure equation coefficients, which are all time variable.

Equation (2.9) is solved using the MUDPACk multigrid solver (Adams 1989, 1991). Prior to application of the algorithm, the equation is normalized (or pre-conditioned) by dividing it by \(\rho^*\). This improves the robustness of MUDPACk considerably (the density decreases by a factor of \(10^8\) from the bottom to the top of the computational domain). Despite the intrinsic inefficiency of multigrid methods, the time variability of the coefficients exacts a significant time penalty and causes the pressure computation to consume over one-half of the total computing time.

\[f. \text{ Basic state}\]

Specification of basic-state potential temperature, density, and pressure is required by the anelastic approximation. These variables are a function only of the vertical coordinate \(z\) and satisfy hydrostatic balance, continuity, the equation of state, and the definition of potential temperature. The assumed profiles are

\[
\theta_0(z) = \theta_{oo} \exp \left\{ (z - z_0)/H_\theta \right\} \quad (2.10a)
\]

\[
(\rho(z), p_0(z)) = (\rho_{oo}, p_{oo}) \exp \left\{ -(z - z_0)/H_p \right\}, \quad (2.10b,c)
\]

where \(H_\theta = RT_{oo}/(\kappa g)\) is the potential temperature scale height; \(H_p = RT_{oo}/g\) is the density scale height; and \(\rho_{oo}, p_{oo}, T_{oo},\) and \(\theta_{oo}\) are the values of density, pressure, temperature, and potential temperature at the lower boundary, which was taken to coincide with the tropopause; that is, \(z_0 = 15\) km. At this altitude the density and potential temperature are \(\rho_{oo} = 0.206\) kg m\(^{-3}\) and \(\theta_{oo} = 383\) K, respectively. The specification of the basic-state profiles is completed by matching observed values at the mesopause, which lies approximately at an altitude of 90 km. The density and potential temperature at this altitude are \(2.53 \times 10^{-4}\) kg m\(^{-3}\) and 8210 K. Scale heights of \(H_p = 6.63\) km and \(H_\theta = 24.5\) km were then determined by fitting exponential profiles from the tropopause to the mesopause. The corresponding basic-state Brunt–Väisälä frequency and buoyancy period are 0.0200 s\(^{-1}\) and 5.23 min, respectively. All of these parameters are typical values for summer solstice conditions at temperate latitudes (Houghton 1986).

\[g. \text{ Boundary conditions}\]

At the bottom of the domain, the deforming boundary is assumed to be a material surface. In transformed coordinates \(\bar{w} = 0 \quad \bar{z} = 0\) [note that this implies a nonzero, physical vertical velocity, \(w \neq 0\), by (2.4)]. At the top of the domain, \(\bar{w} = w = 0\) at \(z = H\). At these boundaries the potential temperature is specified by its basic state values. A sponge layer, distributed in the vertical coordinate \(z\), is also placed adjacent to the top boundary. Although this sponge acts to produce a radiation boundary condition at the top of the domain, it is much more than just a computational device in this study. In particular, the sponge has been tuned to mimic the molecular diffusion profile of the atmosphere. In the lower thermosphere, diffusive effects actually become appreciable [the kinematic viscosity is roughly \(\nu = 100\) m\(^2\) s\(^{-1}\) at an altitude of 100–105 km (Whitten and Poppoff 1971) and increases with altitude approximately as \(\rho^{-1}\)]. Additional details about this sponge are given later in section 3b. The lateral boundary conditions are either periodic- or open-type conditions. They may be combined with lateral sponge layers, whose purpose is to force the transient solution toward the basic-state profiles given in the previous section.

Neumann pressure boundary conditions along the top and bottom of the domain were developed by substituting the vertical momentum (2.5b) into the quadrature formula (2.7) and solving for the pressure gradient terms. Unknown future values of velocity were evaluated in the resulting expression by applying the basic-state values. Open lateral boundary conditions were developed in precisely the same manner from the horizontal momentum (2.5a). Dirichlet conditions were specified for periodic lateral boundaries. Details about the boundary conditions and the sponge layers are given in sections 3a and 3b.

\[h. \text{ Initial conditions}\]

A set of initial conditions is necessary to complete the specification of the numerical problem. The initial
profiles of pressure, density, and potential temperature are set equal initially to the basic-state values given in section 2f. In addition, the vertical velocity \( w \) is set to zero, while the horizontal wind \( u \) is set to a value representative of the observed zonal mean wind in the extratropical middle atmosphere. Specific initial values chosen for \( u \) are discussed in section 3.

3. Computational details

a. Lower boundary forcing and horizontal stretching

Inspection of the equations presented in section 2c reveals that the coordinate functions \( z_c \) and \( F \) must be continuous up to their second derivatives in time and space. This ensures that the transformation and its inverse are well defined (Prusa and Yao 1985). In particular, good numerical behavior generally requires that (appropriately nondimensionalized) first derivatives be of the order of unity (Gal-Chen and Somerville 1975) and that the second derivatives be as small as possible (Dietachmayer and Droegemeier 1992). With these considerations in mind, the forcing at the lower boundary was specified as follows:

\[
z_c(x, t) = AZ(x, t) \exp \left[ -\left( \frac{t - t_m}{\sigma_t} \right)^2 \right], \quad (3.1)
\]

which represents a Gaussian pulse of amplitude \( A \) and half-width \( \sigma_t \), centered at time \( t_m \). The spatial distribution is a Gaussian "mountain" with peak half-width \( \sigma_x \):

\[
Z(x, t) = \exp \left[ -\left( \frac{x - (x_m + \gamma t)}{\sigma_x} \right)^2 \right]. \quad (3.2)
\]

If the parameter \( \gamma \) in (3.2b) is nonzero, the Gaussian peak translates horizontally with velocity \( \gamma \). Time derivatives of the forcing function \( z_c \) are evaluated analytically. Necessary spatial derivatives are evaluated numerically to second-order truncation error using standard finite difference formulas.

The horizontal stretching function was chosen following a series of tests with polynomials and hyperbolic functions. The following functional form (for the inverse function \( f \)) was found to give the most satisfactory results:

\[
f(\tilde{x}, \tilde{t}) = x = \tilde{x}_i + \tilde{L}x + (L - \tilde{L})x^p, \quad (3.3)
\]

where \( X = (x - \tilde{x}_i)/\tilde{L} \), \( L = \tilde{x}_2 - \tilde{x}_i \), and \( \tilde{L} = x_2 - x_1 \). The inverse horizontal stretching function, \( f(\tilde{x}, \tilde{t}) = x \), is specified as a \( p \)-th order polynomial that maps the transformed domain \([\tilde{x}_1, \tilde{x}_2]\) into the physical domain \([x_1, x_2]\). No experiments were conducted with time dependency incorporated. This function has desirable properties in that it is unitary (\( f' = 1 \)) and has zero curvature (\( f'' = 0 \)) at \( X = 0 \). It may then be applied to smoothly merge with a unity transformation \((x \equiv \tilde{x})\) in the interior of the domain or with another parameterization of itself to create a physical domain that has high resolution in its interior regions but distant lateral boundaries. Required spatial derivatives of \( F(x, t) \) were determined using the same finite difference formulas as for \( z_c \), in conjunction with the geometrical identities mentioned in the preceding section on noninertial coordinates. A value of \( p = 3 \) was found to be an acceptable compromise that allowed reasonable stretching yet kept the first and second spatial derivatives of \( f(\tilde{x}) \equiv x \) reasonably small, even at its extremum point, \( X = 1 \). In practice, the solution was found to show some sensitivity to the maximum value of \( f(\tilde{x}) \), and so \( x_1 \) and \( \tilde{x}_i \) were also set equal to as small a value as practicable in the effort to minimize this second derivative.

b. Lateral and vertical sponge layers

The numerical algorithm can use either periodic or open lateral boundary conditions. Either type of boundary condition may be used with the Gaussian spatial forcing (3.2). In the effort to mimic isolated forcing with the Gaussian distribution, lateral sponges were found to be extremely helpful in reducing wave transmission through the lateral boundaries when using periodic boundary conditions (Bacmeister and Schoeberl 1989). For open lateral boundaries, lateral sponges were still found to be helpful in reducing wave reflection (Davies 1983; Kosloff and Kosloff 1986). Careful sponge tuning was found to be critical to the generation of accurate solutions. In particular, the depth of the absorber was set to ensure adequate absorption of waves, while the damping was turned on gradually enough to avoid reflections. These lateral sponges relaxed the velocity and potential temperature fields toward their basic-state values.

Additional considerations played a role in determining the characteristics of the vertical sponge. Unlike tropospheric simulations, where (vertically distributed) absorbers are merely artificial numerical devices for minimizing wave reflections, in deep atmospheres they actually have physical meaning. Since the diffusion parameters for momentum and energy transport become quite significant in the lower thermosphere, the vertical sponge layer may be tuned to effectively mimic the physical damping in the upper mesosphere and lower thermosphere that is caused by molecular diffusion. An approximation for this tuning was obtained by equating the sponge-layer damping term to a molecular diffusion term:

\[
\frac{u_x - u}{\tau} = \nu \nabla^2 u = \frac{\partial u}{\partial t}, \quad (3.4)
\]

where \( u = (u, w) \), \( \tau = \tau(z) \) is the sponge-layer damping time, \( \nu = \nu(z) \) is the molecular diffusion coefficient of the atmosphere, and \( \nabla^2 \) is the Laplacian operator. Solving the two rightmost terms of (3.4) — a Stokes problem — for the lowest-order mode and assuming an initial condition for the disturbance, which remains ev-
Table 1. Parameters used in the numerical experiments. In all runs the forcing function is given by (3.1) and (3.2), with $A = 200$, $t_a = 120$ min, and $\sigma_t = 60$ min. The vertical sponge profile is given by (3.6). The width of the lateral sponges is 30 km for $\sigma_e = 2$ km (cases G1, G4, and G5), 75 km for $\sigma_e = 10$ km (G2 and G6), and 100 km for $\sigma_e = 50$ km (G3). The physical dimensions of the domain are given in all cases. In cases G2 and G3 the stretching transformation (2.3a) was used. In these cases the inner interval $[-120, 120]$ km was unstretched and had a horizontal resolution of 0.625 km in the outer intervals, $[-360, -120]$ km and $[120, 360]$ km, the resolution increased smoothly from 0.625 km at $\pm 120$ km to 2.49 km at $\pm 360$ km.

<table>
<thead>
<tr>
<th>Run</th>
<th>$\sigma_e$ (km)</th>
<th>$x_m$ (km)</th>
<th>$u_0$ (m s$^{-1}$)</th>
<th>$\gamma$ (m s$^{-1}$)</th>
<th>Vertical $\tau_{\text{min}}$ (s)</th>
<th>Lateral $\tau_{\text{min}}$ (s)</th>
<th>$\pm x_{\text{max}}$ (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>2</td>
<td>-30</td>
<td>-25</td>
<td>7</td>
<td>172</td>
<td>200</td>
<td>$\pm 120$</td>
</tr>
<tr>
<td>G2</td>
<td>10</td>
<td>-30</td>
<td>-25</td>
<td>7</td>
<td>116</td>
<td>2100</td>
<td>$\pm 360$</td>
</tr>
<tr>
<td>G3</td>
<td>50</td>
<td>-90</td>
<td>-50</td>
<td>0</td>
<td>116</td>
<td>3300</td>
<td>$\pm 360$</td>
</tr>
<tr>
<td>G4</td>
<td>2</td>
<td>-30</td>
<td>-25</td>
<td>7</td>
<td>23.3</td>
<td>200</td>
<td>$\pm 120$</td>
</tr>
<tr>
<td>G5</td>
<td>2</td>
<td>-30</td>
<td>-25</td>
<td>7</td>
<td>3.15</td>
<td>200</td>
<td>$\pm 120$</td>
</tr>
</tbody>
</table>

erywhere bounded, leads to the result $u = u_{\text{max}} J_0(\lambda r) \exp(-\eta^2 \nu t)$, where $r$ is the distance from the center of the disturbance (at which $u = u_{\text{max}}$). The variable $\lambda$ is the disturbance wavelength. The value of $\eta$ is determined by the boundary condition that $u = 0$ at $r = \lambda/4$. A table of Bessel functions (Abramowitz and Stegun 1972) then leads to $\eta^2 \approx 90/\lambda^2$. Equating the damping solution of the first and last terms of (3.4) with the exponential transient part of the Stokes solution immediately results in

$$ \tau^{-1} = \eta^2 \nu \approx \frac{90\nu}{\lambda^2}. \tag{3.5} $$

Thus, it is possible to choose the damping time to match the effect of molecular diffusion upon waves of a given wavelength. Shorter, higher-frequency waves will be underdamped, whereas longer, lower-frequency waves will be overdamped compared to viscous damping. The numerical results indicate that, prior to overturning and wave breaking, a single wavelength dominates the wave field, even though the Gaussian pulse (3.2) is used as the forcing function. It is to this dominant wavelength that the vertical and lateral sponges are tuned.

In order to simulate the diffusivity profile of the atmosphere, the damping rate in the vertical sponge layer is specified to increase exponentially with altitude throughout the whole domain. Its vertical profile is given by

$$ \tau^{-1} = \tau_{\text{min}}^{-1} \exp \left( \frac{z - z_{\text{top}}}{H_z} \right), \tag{3.6} $$

where $z_{\text{top}}$ denotes the top boundary, $\tau_{\text{min}}$ is the damping time at $z_{\text{top}}$, and $H_z = 5.7$ km. Since the molecular diffusivity of the atmosphere increases with height primarily due to decreasing density, this scale height is somewhat small (it should match the density scale height). However, this value was used in order to match smoothly diffusivity data for the lower thermosphere (Whitten and Popoff 1971) with corresponding data for the upper mesosphere (U.S. Standard Atmosphere, NOAA 1976).

Since the Prandtl number, $Pr = \nu/\alpha$ (where $\alpha$ is the thermal diffusivity), of the atmosphere is approximately $Pr = 0.7$ (Fay 1965; NOAA 1976; Whitten and Popoff 1971), the thermal diffusivity is the same order of magnitude as the kinematic viscosity. Consequently, the potential temperature field was also relaxed in the sponges. Thus, the damping of the wave field in the top sponge should be thought of as being due to the molecular diffusivity of the atmosphere and not to other processes, such as radiative relaxation [which has a much longer timescale on the order of 3–10 days for the mesosphere (Andrews et al. 1987)].

### c. Spatial resolution and domain size

Horizontal and vertical resolution of 0.625 km (in transformed coordinates) was used in all of the calculations presented in sections 4–6. Appendix A shows that this resolution is about the minimum necessary to yield converged results for the wavefield characteristics we examine. The horizontal extent of the computational domain, $\pm x_{\text{max}}$, was based upon a number of trial experiments to determine the minimum size required to closely approximate the effects of isolated forcing as well as to contain all of the relevant physical phenomena of interest. We found that $x_{\text{max}} = 120$ km was adequate when the half-width of the Gaussian forcing function was narrow [$\sigma_e = 2$ km in (3.2)]. For wider Gaussian forcing, $x_{\text{max}} = 360$ km was used.

Horizontal stretching was employed in several of the calculations presented in this paper (see Table 1). In these calculations the physical domain was $\pm 360$ km, but the transformed domain had an extent of only $\pm 240$ km. The outer [240, $\pm 120$] km intervals of this domain were transformed into physical domain intervals of [240, $\pm 120$] km (corresponding to $x_1 = \pm 120$ km, $x_2 = \pm 240$ km, and $x_3 = \pm 360$ km). Thus, the inner interval $[-120, +120]$ km was unstretched and had a physical horizontal resolution of 0.625 km. In the outer intervals, the physical horizontal resolution increased smoothly from its inner value to 2.49 km at the edge of the domain.
The maximum height of the computational domain was based upon the physical properties of the lower thermosphere; in particular, the facts that at higher altitudes the molecular viscosity is larger, while at lower altitudes the atmosphere is still reasonably modeled as a continuum and is well mixed. Based on these factors, the top boundary was placed at \( z_{\text{top}} = 135 \) km altitude. Above this level, diffusive effects start to become so significant that the atmosphere takes on a heterogeneous character (Whitten and Poppoff 1971). For example, at 135-km altitude the mean molecular weight is approximately 12% lower than its standard value below the turbopause (U.S. Standard Atmosphere, NOAA 1976). With regard to the continuum nature of the atmosphere at this altitude, the mean free path is approximately \( l = 15 \) m (U.S. Standard Atmosphere, NOAA 1976). Since the finest spatial scale is set by the grid resolution, the Knudsen number, \( \text{Kn} = l/\Delta x \), is 0.025. Although this is much smaller than one, it is near the upper limit for a continuum flow (Eckert and Drake 1959).

d. Time step

The effects of the time step were monitored through the growth of the Lipschitz number, \( L \). While (2.8) suggests that \( L < 1 \) is sufficient for good results, our experience indicated that \( L < 0.5 \) was a more reasonable working guide. This criterion resulted from the need to keep the time truncation error small rather than from convergence considerations. In particular, the numerical results showed a tendency to break too soon, and computational statistics such as the standard deviations of the divergence and the flow Jacobian (Lamb 1945, section 14) — \( D_{\text{ad}} \) and \( J_{\text{ad}} \), respectively — would become relatively large if the time step were chosen so that \( 0.5 < L < 1 \). These statistics were computed according to

\[
D = \Delta t \left[ \frac{\partial \rho^*}{\partial t} + \nabla \cdot (\rho^* \mathbf{u}) \right] \quad \text{(divergence)} \quad (3.7a)
\]

and

\[
J = \frac{\partial (\mathbf{x}, \mathbf{z})}{\partial (x, z)} \left( \frac{\bar{p}}{\rho} \right) \quad \text{(flow Jacobian).} \quad (3.7b)
\]

Physically, \( D \) represents mass conservation from an Eulerian viewpoint, (3.7a) following immediately from the continuity equation (2.5d), while \( J \) represents mass conservation from a Lagrangian viewpoint. Nominal, \( D = 0 \) and \( J = 1 \). Any deviations of the divergence from zero simply reflect the degree of convergence obtained on the pressure field, determined from (2.9). Deviations of the flow Jacobian from zero, on the other hand, reflect truncation errors of the semi-Lagrangian method itself. With \( L < 0.5 \), no pathological behaviors were noted, and the maximum values of \( D_{\text{ad}} \) and \( J_{\text{ad}} \) were 0.0034 and 0.0038, respectively. In general, these maximum values occurred only for brief intervals corresponding to the most intense wave activity; at other times, \( D_{\text{ad}} \) and \( J_{\text{ad}} \) were significantly smaller. For most of the calculations reported in this paper, the pressure equation was solved using MUPACK, as noted in section 2e, and the convergence limit was automatically set to the truncation error. We also used a different computational algorithm to solve the pressure equation, a conjugate residual method customized for our model (Smolarkiewicz and Margolin 1994). With this algorithm it was possible to reduce \( D \) to the level of computer precision.

Since the results always showed very strong development of the field gradients in time, this required that the time step be shortened as the computation progressed in order to keep \( L < 0.5 \) and minimize \( D_{\text{ad}} \) and \( J_{\text{ad}} \). Typically, a time step of between 10 and 20 s was used at the beginning of the computations, while a time step of between 2 and 6 s was required as the wave field approached saturation. In all cases, the maximum values of \( D_{\text{ad}} \) and \( J_{\text{ad}} \) generally occurred only for a brief interval associated with the most intense wave activity. Throughout the rest of the calculations, \( D_{\text{ad}} \) and \( J_{\text{ad}} \) were significantly smaller than their maximum values. It is noteworthy that, in all the computations, the Courant number was often greater than unity.

e. Summary of computational parameters

Table 1 lists the calculations discussed in sections 4–6 of this paper. Here, \( A \), \( \sigma \), and \( x_m \) are the amplitude, half-width, and initial location of the lower boundary forcing, respectively; \( u_0 \) and \( \gamma \) are the background wind and speed of translation of the forcing; \( \tau_{\text{min}} \) gives the minimum damping time for the top and lateral sponges; \( \pm x_{\text{max}} \) denotes the horizontal extent of the domain; and \( \Delta x \) is the horizontal resolution. The suite of calculations listed in Table 1 is intended to illustrate the effect of several basic parameters on the excitation, propagation, and breaking of the gravity waves.

In most of the experiments, the background zonal wind was set equal to a constant value of \( u_0 = -25 \) m s\(^{-1}\), chosen to approximate the middle atmosphere near summer solstice conditions at temperate latitudes (Andrews et al. 1987; Randel 1992). In one experiment a constant wind of \( u_0 = -50 \) m s\(^{-1}\) was used. This latter value is more representative of larger zonal winds that may be encountered locally or during winter (although one must then reverse the results, right to left, as the winter winds are westerly), and it serves to illustrate the effect of the background wind distribution. We do not examine the effects of wind shear (or nonuniform stability) in this study because, as will be seen in the following sections, the resulting behavior using uniform profiles is already extremely rich.

The Gaussian forcing function (3.2) was chosen to mimic the effect of a traveling summer squall line storm.
Calculations that used this type of forcing are denoted by $G_i$, where $i$ is the number of the calculation (see Table 1). The maximum amplitude of the disturbance was $A = 200$ m. Peak half-widths of $\sigma_x = 2$ and 10 km were used, with a translation velocity of $\gamma = 7$ m s$^{-1}$. One calculation was carried out with $\gamma = 0$ (nontranslating forcing) and $\sigma_x = 50$ km; this choice of parameters was meant to represent forcing by a broad topographic obstacle. The time of maximum forcing was set to $t_m = 2.00$ h and a pulse half-width of $\sigma_t = 1.00$ h was used. No attempt was made in this study to model the higher-frequency transients that appear to be generated by mesoscale storms (Fovell et al. 1992).

4. Results

In this section we first present the results of three numerical experiments designed to illustrate how the spatial extent of the forcing and the characteristics of the background state affect the excitation, propagation, and breaking of the gravity wave field. We then show in some detail how the numerical results can be understood in terms of the linear theory of gravity waves. From such theoretical considerations we also argue that the maximum altitude at which breaking occurs—the turbopause—depends in a straightforward manner on a small number of parameters characterizing the wave field and the background state. We conclude the discussion by examining the effect of horizontal resolution on the numerical simulation of wave breaking.

a. Evolution and breaking of the mesospheric wave field

1) NARROW SQUALL LINE

Figure 2 shows the development of the potential temperature field for case G1 at 140, 145, and 160 min into the computation. This case is intended to simulate a narrow squall line since the Gaussian width, $\sigma_x$, of the lower boundary deflection is only 2 km. The forcing function translates in the positive $x$ direction at speed $\gamma = 7$ m s$^{-1}$; its time dependence is given by (3.1) with $\sigma_t = 1$ h and $t_m = 2$ h. Thus, the forcing has already attained its maximum amplitude and begun to diminish by the time the fastest waves reach the upper mesosphere. The triangles along the abscissa of Fig. 2 denote the position of the lower boundary forcing, which is initially located at $x = -30$ km but has propagated to $x = +37$ km after 160 min.

The three panels of Fig. 2 are representative of the saturated ($t = 140$ min), incipient breaking ($t = 145$ min), and mature ($t = 160$ min) stages of the evolution of the wave field. Just prior to and near the initial point of wave breaking (at $t = 125$ min, $x = -60$ km, and $z = 105$ km), spectral analysis reveals that the wave power is confined to narrow peaks with horizontal and vertical wavelengths of $\lambda_x = 15.1$ km and $\lambda_z = 14.0$ km, respectively. These wavelengths characterize most of the wave field, which is almost monochromatic for the range of altitudes and times shown in Fig. 2. Compared to the Brunt–Väisälä period, the nonbreaking wave field changes only slowly in time and space. At 160 min and 105-km altitude, longwave spectral power is concentrated in a narrow peak near $\lambda_x = 18.4$ km, whereas longer wavelengths of up to $\lambda_x = 40$ km appear in a much broader peak at 60-km altitude. As a consequence of the near-unity aspect ratio, $\lambda_x/\lambda_z = 0.93$, the overturning of potential temperature surfaces that accompanies breaking involves roughly half the horizontal wavelength—a result that contrasts with the behavior of longer waves, as shown in Walterscheid and Schubert (1990).

The nearly monochromatic wave field at mesospheric altitudes indicates that considerable vertical dispersion has occurred, since in the vicinity of the nar-

Fig. 2. Potential temperature field for the "narrow squall line case" (G1) at saturation ($t = 140$ min) and during the incipient ($t = 145$ min) and mature ($t = 160$ min) stages of wave breaking. The half-width $\sigma_x$ of the lower boundary deflection is 2 km. The mesospheric wave field is nearly monochromatic and is found downwind from the source, whose position is denoted by the triangles along the $x$ axis. Contour interval is $\Delta \ln \theta = 0.0816$.
row Gaussian forcing the response must be locally confined. (This cannot be readily appreciated in Fig. 2 because the deflection of the potential temperature surfaces is too small below the stratopause.) Dispersion in the horizontal direction is also evident in the mesospheric wave train, which extends over a span of some 50–100 km. At 140 min, the region of large wave amplitudes lies well downstream from the location of the lower boundary forcing, but it gradually advances relative to the forcing until, at 160 min, its leading edge is only 20 km downstream from the forcing.

Figure 3 shows the detailed structure of the $\theta$, $u$, and $w$ fields during the mature stage of breaking ($t = 160$ min). It encompasses the region of vigorous breaking in Fig. 2c, an ellipse roughly 50 km by 85 km centered at $x = -45$ km and $z = 105$ km. The enlarged region consists of nearly 14,000 cells, each 0.625 km $\times$ 0.625 km. Once saturation occurs, vigorous breaking begins within a few Brunt–Väisälä periods. Two distinct types of morphological structure develop: The first type, the breaking front, is found windward and below the breaking region. It consists of regions of vigorous wave overturning, although fluid streamlines are still essentially continuous. The horizontal and vertical winds in the breaking front are characterized by jetlike structures with characteristic scales of 5–10 km. Locally, extremely large shears (50–100 m s$^{-1}$ km$^{-1}$) may appear in small regions only 1–2 km deep. The second type of structure, the turbulence field, is located leeward of the breaking front; it is characterized by much finer scales and contains fluid particles that have broken away from their original streamlines. The turbulence field grows steadily in size as the simulation advances in time, whereas the breaking front remains roughly the same size.$^1$

2) Broader Squall Line

We consider next the behavior of traveling forcing identical to the case just discussed except that the $\epsilon$-folding width of the lower boundary deflection is now $\sigma_x = 10$ km. This case is denoted by G2 in Table 1. The evolution of the potential temperature field is shown in Fig. 4 at $t = 175$, 185, and 200 min. As in case G1, the wave field is monochromatic at mesospheric altitudes, although with somewhat longer horizontal wavelength (22 km) and shorter vertical wavelength (11.5 km). The wave field becomes saturated at $t = 175$ min; it begins to break vigorously after 185 min, reaching the mature stage by 200 min. The waves display dispersion similar to that found in the narrow squall line case, except that the region of large wave amplitudes is not as far downstream as in the previous case. In fact, by 200 min, the leading edge of the mesospheric wave field has overtaken the lower boundary forcing.

In spite of the differences just noted, the wave field at high altitudes is remarkably similar to that of the previous case, even though there is a factor of 5 dif-

$^1$ Animation of the model results unmistakably reveals motion and dissipation of random eddies over many length scales in the turbulence field. The breaking front is seen to develop and then move rapidly upwind and gives the appearance of a moving wave front, which drastically changes the character of the monochromatic wave field through which it propagates. A VHS tape of the animation (case G1) can be obtained from the authors upon request.
is that the background flow is $-50 \text{ m s}^{-1}$, rather than the $-25 \text{ m s}^{-1}$ used in the other two cases. As in the previous two cases, the time evolution of the forcing is given by the Gaussian pulse (3.2). This case is an idealization of forcing by a strong flow over a broad topographic barrier.

The behavior of the potential temperature field for case G3 is shown in Fig. 5. The triangle denotes the position of the stationary forcing. The considerably broader forcing at the lower boundary has produced a large change in the character of the wave field in the mesosphere. As in cases G1 and G2, the wave field is predominantly monochromatic, but the dominant horizontal wavelength is now much larger—about 100 km near the time and altitude (105 km) of incipient breaking. At the lowest altitudes where the deflection of potential temperature surfaces is noticeable, the horizontal wavelength is even longer (e.g., 145 km at 60-km altitude). The change in horizontal wavelength between cases G2 and G3 (22 vs 110 km) matches the factor of 5 increase in the half-width of the forcing, $\sigma_x$.

The vertical wavelength ($\lambda_z \approx 19 \text{ km}$) remains relatively small and comparable to that found in cases G1 and G2. Since $\lambda_z$ is now much longer, the aspect ratio of these waves is much different from those found in the squall line cases. Thus, as seen in Figs. 5b,c, overturning of potential temperature surfaces affects a much smaller portion of the horizontal wavelength than in cases G1 and G2. The long waves that characterize case G3 do not reach saturation until $t = 225 \text{ min}$, by which time the forcing at the lower boundary has already decreased substantially. The breaking region, and indeed the entire region where wave amplitudes are large, now occurs upstream of the lower boundary forcing, rather than downstream, as in the previous two cases.

Figure 6 shows details of $\theta$, $u$, and $w$ in the breaking region 260 min after the onset of forcing. Many of the features seen in the narrow squall line case (Fig. 3) are also present here. However, the turbulent region containing these features appears more organized than its counterpart in Fig. 3. An overall quasi-periodic structure may be discerned that is most obvious in the vertical velocity field, which shows a cellular structure with horizontal and vertical scales of order 10 km. This organization is evidently a consequence of the small aspect ratio of the waves in the present case. These results are strikingly similar to those shown in Figs. 1–3 of Walterscheid and Schubert (1990). The close match strongly suggests that this structure may be set only by the scale height of the atmosphere and not by the characteristics of the forcing. In Walterscheid and Schubert (1990), the basic-state atmosphere is at rest, and the forcing source is a traveling sine wave with a wavelength of 300 km (this sets the horizontal wavelength of the prebreaking wave field) and a phase velocity of $41.7 \text{ m s}^{-1}$. Finally, the sharply defined breaking front that is observed in Figs. 2–4 is much less apparent in Figs. 5 and 6. The long horizontal wave-

3) Broad topographic forcing

The foregoing results are complemented by case G3, wherein the forcing is a broad ($\sigma_x = 50 \text{ km}$), nontranslating ($\gamma = 0$) Gaussian deflection of the lower boundary, centered at $x = -90 \text{ km}$. An additional difference in the width of the lower boundary displacement ($\sigma_x = 2 \text{ km}$ vs $\sigma_x = 10 \text{ km}$). In particular, the horizontal wavelength of the monochromatic wave field in the mesosphere increases by less than 50%, from 15 to 22 km. This is strikingly different from the behavior of tropospheric mountain waves, whose wavelength is known to scale with the horizontal dimensions of the source (Smith 1979). At lower altitudes (below 65 km) rather longer horizontal wavelengths ($\approx 30 \text{ km}$) are seen in the present case at 200 min. The detailed development of $\theta$, $u$, and $w$ (not shown) is similar to that found for the narrow squall line. This is not surprising since in both cases the breaking wave field is dominated by waves of short horizontal wavelength, which have aspect ratios of order unity.
length of the present case acts to spread out the breaking front over a much larger horizontal region.

b. Horizontal and vertical evolution of the wave breaking region

The two panels of Fig. 7 illustrate how the breaking region develops in the vertical and horizontal directions, respectively, as a function of time for the three cases, G1–G3, discussed above. The breaking region is defined in terms of the location of the maximum and minimum heights of saturated waves in the computational domain for the three cases. We consider waves to become saturated when $\partial \theta / \partial z$ first vanishes. Note that this is a local characteristic of particular waves and not a global characteristic of the wave field. The locations of the saturation points have been obtained from inspection of the potential temperature fields. Since saturation is a precursor to breaking, Fig. 7a approximates the transient vertical development of the breaking region, although the extent of the breaking region may be overestimated somewhat (at the edges of the breaking region the waves are saturated but do not quite break).

The lower edge of the breaking region descends quickly to a global minimum value of 65 km in case G1 and to 60 km in case G2, a drop of about seven density scale heights. In both cases, saturated waves first appear in a small region less than 10 km in extent. The onset of saturation in experiment G1 occurs at 128 min at an altitude of 108 km. For G2 the onset of saturation occurs at 161 min at an altitude of about 107 km. A brief time interval of explosive growth follows, with the vertical interval of saturated waves tripling in size in just 5 min—about one Brunt–Väisälä period. The growth in the vertical extent of the saturation region occurs primarily by extension to lower altitudes; the upper edge of the region reaches a maximum altitude of about 120 km within a few minutes after the
inception of breaking and remains constant thereafter. Note that the minimum breaking amplitude in case G2 is somewhat lower than in G1. This is consistent with the greater total forcing power of the broader Gaussian ($\sigma_x = 10$ km) of case G2.

In case G3, saturation is reached at 107 km and 216 min, about the same altitude but much later than in the previous two cases. The maximum breaking height (125 km) is also similar to that of the previous cases, but the minimum height was still decreasing by the end of the calculation at 270 min. The calculation had to be terminated at this time because of leakage of wave activity across the side boundaries. The very red forcing of case G3 ($\sigma_x = 50$ km) generates large amounts of power at long wavelengths that cannot be satisfactorily contained within the computational domain once dispersion becomes significant.

The horizontal development of the region of saturated waves is depicted in Fig. 7b. The same set of calculations as in Fig. 7a are shown. The filled circles indicate the position of the forcing in cases G1 and G2. For these cases, saturation is seen to begin well down-
wind of the traveling disturbance. As in Fig. 7a, the first saturated waves appear in a small region less than 10 km wide. This initial development is again followed by extremely rapid growth for a short time—the horizontal extent is over 40 km within 5 min after the onset of saturation. The leeward edge represents the boundary of the turbulence field that is generated by vigorous breaking; its location is subject to considerable oscillation as various eddies enter this area and are dissipated. The windward edge eventually overtakes the forcing, especially in case G2.

In case G3 breaking begins some 90 km upstream of the (stationary) forcing, whose position is denoted in Fig. 7b by the open circles. The breaking region then expands rapidly in the horizontal as the dominant long waves reach their breaking altitudes. As time goes on, the trailing edge of the breaking region moves downstream, toward the forcing. By the end of the calculation, the trailing edge of the breaking region has propagated past the position of the forcing.

c. Self-acceleration

In a study of gravity wave saturation in a background shear flow, \( u_0(z) \), Fritts and Dunkerton (1984) found that, as the waves approached a critical level, their phase velocity \( c \) appeared to adjust so as to keep the Doppler-shifted phase speed \( u_0 - c \) approximately constant. They called this phenomenon self-acceleration, and noted that it could not be explained in terms of linear theory, which instead predicts a rapid change in wavelength near a critical level (e.g., Jones and Houghton 1972). Walterscheid and Schubert (1990) found similar behavior in their study of breaking gravity waves. In their calculations the vertical wavelength remained fairly constant despite very large changes in the background wind, implying no significant change in \( u_0 - c \). Walterscheid and Schubert also determined \( c \) from the movement of constant-phase lines and noted that the result was consistent with the approximate constancy of the vertical wavelength.

We find evidence of self-acceleration in all of our calculations. Examination of Figs. 2, 4, and 6 reveals no large changes in the combined wavelength\(^2\) as the waves approach saturation, or indeed at any time during the breaking process, even though the background wind undergoes very large changes. In common with Walterscheid and Schubert, we find that the near constancy in \( u_0 - c \) implied by the approximately constant wavelength is consistent with the results of a direct determination of \( c \). Obtaining the phase velocity from our results is complicated by the fact that the wave field is not exactly monochromatic, even at high altitudes. Accordingly, quadratic regression fits were made to the calculated potential temperature fields at 60 and 105 km to determine the position of extrema and zeros. The curve fits included sufficient points to smooth out finescale variations due to breaking. Phase velocities were then determined by computing the horizontal displacement of particular extrema/zeros during specified time intervals. Typically, four to eight such points were analyzed in the domain \(-120 \text{ km} < x < 120 \text{ km} \). These local values were then averaged to give an estimate of the phase velocity.

The results for case G3 are shown in Fig. 8 as a function of time at \( z = 105 \text{ km} \). The near constancy of \( u_0 - c \) is especially striking after 220 min, when the

---

\(^2\) In general, the Doppler-shifted phase velocity is related to the combined wavelength, \( \lambda_c = 2\pi/(k^2 + m^2) \), where \( k \) and \( m \) are the horizontal and vertical wavenumbers, respectively. This is equivalent to the vertical wavelength \( 2\pi/m \), when \( m \gg k \). See the discussion in section 5c.
deceleration of the background wind becomes large. The phase velocity changes from about 8 m s\(^{-1}\) in the early stages of the simulation to over 33 m s\(^{-1}\) at 237 min, and its evolution tracks that of the zonal wind so changes in \(u_0 - c\) are always smaller than changes in \(u_0\) or \(c\) individually.

Both Fritts and Dunkerton and Walterscheid and Schubert have raised the question of whether self-acceleration represents an actual change in the phase speed of the wave form or whether it is instead due to the presence of a spectrum of frequencies, each of which has its own critical level. Our results support the former possibility. Although the forcing in our calculations evolves in time, it does so on a relatively long timescale. Recall from section 3 that the time dependence of the forcing is Gaussian, with a characteristic timescale, \(\sigma_t = 60\) min. The corresponding power spectrum, shown in Fig. 9, indicates that the forcing amplitude falls by a factor of 100 at \(\omega = \pm 0.05\) min\(^{-1}\) compared to its value at zero frequency. Even for the relatively long zonal wavelength (110 km) that dominates the solution in case G3, this corresponds to phase velocities of only \(\pm 14.6\) m s\(^{-1}\), much smaller than the 33 m s\(^{-1}\) attained at 237 min. Furthermore, we show in section 5c that the vertical group velocity in case G3 is a strong function of frequency, such that waves with large, positive phase velocities should reach high altitudes ahead of waves with slower phase velocities. This is just the opposite of the behavior seen in Fig. 8, where \(c\) remains nearly constant at less than 8 m s\(^{-1}\) through \(t = 200\) min and then increases sharply (and in concert with the deceleration of the zonal wind) in the latter stages of the simulation.

5. Interpretation in terms of linear wave theory

Although self-acceleration does not appear to be explainable in terms of linear theory, many of the results discussed in sections 5a, b can be readily interpreted in those terms. For example, one would expect the longer waves of case G3 to propagate to high altitudes more slowly than the short waves of cases G1 and G2, since the vertical group velocity of the waves is known to be proportional to the horizontal wavenumber. More detailed consideration of the results reveals that most features of the behavior of the waves can be explained in terms of linear theory.

\[ a.\: Quasi-stationary\: waves \]

We begin by showing that, to a first approximation, the development of the wave field can be understood by considering the waves to be stationary with respect to the forcing. That is, we ignore the fact that the growth and decay of the forcing will excite a spectrum of frequencies and assume instead that the wave field can be characterized as having zero frequency relative to the forcing but a (slowly) varying amplitude that follows the time development of the latter. It turns out that this assumption is a rather good one for cases G1 and G2, wherein the wave field at high altitudes is dominated by short horizontal wavelengths. The assumption is less justified for the long waves of case G3 but is still useful even in that case.

Since the waves are not exactly stationary with respect to the forcing, we refer to them as quasi-stationary, linear waves. Linear theory for nonhydrostatic gravity waves (Bretherton 1971) then leads to the following dispersion relation when the basic-state wind is uniform:

\[ m^2 = \left( \frac{N}{u_{rel}} \right)^2 - k^2 - M^2, \]

(5.1)

where \(M = 0.5/H_\omega\), \(m\) is the vertical wavenumber, \(k\) is the horizontal wavenumber, and \(u_{rel} = u_0 - \gamma\) is the relative wind velocity, which must be used in the dis-

![Normalized Frequency Spectrum](image)

**Fig. 9.** Normalized frequency spectrum of the lower boundary forcing. The power falls by a factor of 100 at \(\omega = \pm 0.05\) min\(^{-1}\).
Table 2. Characteristics of the dominant quasi-stationary waves.

<table>
<thead>
<tr>
<th>Run</th>
<th>$\sigma_*$ (km)</th>
<th>$u_{rel}$ (m s$^{-1}$)</th>
<th>Measured wavelength (km)</th>
<th>$\lambda_c$ (theory)</th>
<th>$c_{eq}$, $c_{ec}$ (m s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>2</td>
<td>-32</td>
<td>15.1</td>
<td>14.0</td>
<td>10.12</td>
</tr>
<tr>
<td>G2</td>
<td>10</td>
<td>-32</td>
<td>22.0</td>
<td>11.5</td>
<td>10.12</td>
</tr>
<tr>
<td>G3</td>
<td>50</td>
<td>-50</td>
<td>110</td>
<td>18.6</td>
<td>15.99</td>
</tr>
</tbody>
</table>

Dispersion relation to properly account for the Doppler shift caused by the translation of the forcing source. The corresponding vertical and horizontal group velocities are

$$c_{ge} = kN \left( 1 - \frac{k^2 + M^2}{K^2} \right)^{1/2} = \frac{kNm}{K^3} \quad (5.2a)$$

and

$$c_{ge} = u_{rel} - \frac{N}{K} \left( \frac{k^2}{K^2} \right) = u_{rel} - \frac{Nm^2 + M^2}{K^2}, \quad (5.2b)$$

where

$$K^2 = m^2 + k^2 + M^2 = \left( \frac{N}{u_{rel}} \right)^2. \quad (5.2c)$$

Our sign convention is that $k$, $K$, and $\lambda_*$ have the same sign as $u_{rel}$. The parameter $K^2$ is the total wavenumber, defined by the middle sum in (5.2c). For the present case of quasi-stationary waves, this sum equals the extreme rhs of (5.2c). Note that, as $k \to 0$, (5.1)–(5.2) collapse to the familiar hydrostatic relations. The dispersion relation may be simplified if one considers the combined wavenumber in the direction perpendicular to the lines of constant phase:

$$m_c = \sqrt{m^2 + k^2} = \sqrt{\frac{N^2}{u_{rel}^2} - \frac{1}{4H^2}}. \quad (5.3)$$

Note that it is determined solely by the properties of the basic state. The combined wavenumber is also known as the Scorer parameter; it appears in the linearized vertical structure equation for gravity waves (Keller 1994). It is through the Scorer parameter that the vertical distributions in density, zonal wind, and stability influence the waves. The combined wavenumber for nonhydrostatic waves has the same value as the vertical wavenumber predicted by hydrostatic theory. The corresponding combined wavelength may be determined from $\lambda_c = 2\pi / m_c$ or according to

$$\lambda_c = \frac{\lambda_* |\lambda_*|}{(\lambda_*^2 + \lambda^2)^{1/2}}. \quad (5.4)$$

Once a horizontal wavelength has been excited by the forcing function, the basic state dictates the permitted vertical wavelength through the dispersion relation. This behavior has also been observed by Keller (1994) in a theoretical study on the effects of wind shear on the vertical propagation of gravity waves.

b. Wavelengths in the breaking region

Equations (5.1–5.4) provide a set of relations that may be used to check the consistency of the computations with the predictions of linear theory. To determine values of the horizontal, vertical, and combined wavelengths, the perturbation fields in the region where breaking occurs were analyzed just prior to breaking. In practice, it is somewhat difficult to pinpoint precise wavelengths for the dominant waves because their characteristics change in time and space. In case G3, there is the additional difficulty that only a few wavelengths are present within the breaking region. For these reasons, the determination of wavelengths involves a certain amount of subjectivity. Table 2 lists the vertical and horizontal wavelengths that were determined near the point of incipient breaking from the three numerical experiments presented in section 4.1. Although the horizontal scale of the forcing and/or the relative wind differ significantly among the three cases, the measured combined wavelengths are predicted well by linear theory. In cases G1 and G2, $\lambda_*$ lies within 2% of the theoretical value given by (5.3); in case G3, the difference is 14%. (The reason for the greater discrepancy in case G3 is explored in section 5c.) Also shown in Table 2 are the horizontal and vertical group velocities of the dominant waves. These are considerably larger for the short waves of cases G1 and G2 than for the longer waves of case G3.

c. Dispersion

According to the linearized form of the thermodynamic equation, the perturbation potential temperature $\theta'$ at the lower boundary is proportional to the forcing $z_0'$; that is,

$$\theta' = -z_0' \frac{\partial \theta_0}{\partial z} \text{ at } z = z_0, \quad (5.5)$$

where $\theta_0$ is the background potential temperature field and $z_0$ is given by (3.1)–(3.2). Since the wavenumber spectrum of $z_0$ is Gaussian, the spectrum of the response at $z_0$ must also be Gaussian. The range of wavelengths

Unauthenticated | Downloaded 07/30/22 06:24 AM UTC
excited depends on the half-width of the forcing, $\sigma_x$, although not all of these wavelengths correspond to vertically propagating waves. Equation (5.1) indicates that, for sufficiently large $k$, the vertical wavenumber becomes imaginary, so the waves are vertically evanescent.

Figure 10 shows wavenumber power spectra of the forcing function (the square of the Fourier transform of (3.2)) for the three cases under consideration. In cases G1 and G2, power is distributed over a very wide range of $k$, but in case G3 there is negligible forcing of horizontal wavelengths $<60$ km ($k > 0.1$ km$^{-1}$). The extreme concentration of power at long wavelengths in case G3 helps explain why only waves of very long zonal wavelengths are seen in the breaking region in Fig. 5: there is essentially no forcing power at the shorter wavelengths. The situation is not this simple in cases G1 and G2. The power spectrum of the forcing is much "redder" in case G2 than in case G1, but the mesospheric wave field is dominated in both cases by monochromatic wave trains of short zonal wavelength (15–22 km). To understand this behavior, we need to consider the dependence of the vertical group velocity on zonal wavenumber, which may be computed from (5.2). The permitted range of horizontal wavenumbers is $0 \leq k \leq m_t$, with $m_t$ given by (5.3). The limit $k \rightarrow 0$ corresponds to hydrostatic waves, while the limit $k \rightarrow m_t$ corresponds to evanescent waves. The corresponding range in horizontal wavelength is $\infty > \lambda_x \approx \lambda_c$.

Figure 11 shows the variation of the components of the group velocity as functions of zonal wavenumber for $u_{ref} = -32$ and $-50$ m s$^{-1}$. The vertical group velocity $c_{gz}$ increases rapidly from zero at the evanescent asymptote $m_t$ to a maximum value at $\lambda_x = 2\lambda_c$, and then decreases more gradually toward zero as $\lambda_x \rightarrow \infty$. The horizontal group velocity $c_{hx}$ varies monotonically from zero at $\lambda_x \rightarrow \infty$ to $-c_{gz}$ at $\lambda_x = \sqrt{2}\lambda_c$. Its magnitude continues to increase even beyond the evanescent asymptote, but this is not relevant to the present problem, since the corresponding waves do not propagate vertically. Figure 11 suggests that the early appearance of short waves in the mesosphere in cases G1 and G2 is a result of their large vertical group velocities. In both cases, sufficient power is available at wavelengths $<30$ km ($k > 0.2$ km$^{-1}$) to allow these rapidly propagating waves to dominate the wave field and initiate breaking at high altitudes. Note that long waves ($\lambda_x > 50$ km, $k < 0.125$ km$^{-1}$) are also excited in cases G1 and G2. However, because of their smaller $c_{gz}$, these waves reach mesospheric altitudes considerably later than shorter waves. The integration of case G1 was not carried out long enough to capture the propagation and breaking of these longer waves. As suggested by Fig. 11, waves with $\lambda_x > 50$ km would take at least twice as long as the fastest waves to reach breaking altitudes.

The power associated with the wavelengths that break in cases G1 ($k \approx 0.416$ km$^{-1}$, $\lambda_x \approx 15$ km) and G2 ($k \approx 0.286$ km$^{-1}$, $\lambda_x \approx 22$ km) is similar (cf. Fig. 10). The average numerical value is 7.1 m$^2$, which corresponds to a displacement amplitude of 2.65 m. If the waves propagate conservatively, this will grow to over 2000 m by 110 km because of the density stratification of the atmosphere. A displacement of this magnitude, which is comparable to the inverse vertical wavenumber m$^{-1}$, should be sufficient to destabilize the potential temperature gradient since the linear saturation condition,

$$|\theta_x'|/|\theta_x| = 1,$$

(5.6a)
implies

$$|c'| = m^{-1}. \tag{5.6b}$$

According to Table 2, the dominant waves of case G1 have approximately equal horizontal and vertical wavenumbers, \(m_1/\sqrt{2}\). These waves also have vertical and horizontal group velocities that are nearly equal in value and opposite in direction. The fact that \(c_{gx}/c_{gz} \approx -0.9\) accounts for the appearance of the mesospheric wave train downstream of the forcing, at a zenith angle of approximately \(-42^\circ\) (Fig. 2). In case G2, the mesospheric wave field is dominated by somewhat longer horizontal wavelengths (about 22 km). The ratio of vertical to horizontal velocities in this case is \(c_{vx}/c_{gz} \approx -0.52\). This implies that the mesospheric wave field should be found at a zenith angle of about \(-27^\circ\), in good agreement with Fig. 4. In the later stages of the evolution of cases G1 and G2, the mesospheric wave field approaches and begins to overtake the surface forcing. This can be understood by noting that, at these later times, the longer wavelength components of the response make a greater contribution to the wave activity reaching the mesosphere. These longer waves have slower vertical group velocities and, therefore, a longer propagation time to the high altitudes.

The time at which saturated waves first appear provides additional support for the interpretation that the waves with largest \(c_{gx}\) are responsible for breaking at mesopause altitudes. For example, in experiment G1 the measured horizontal wavelength is 15.1 km, which corresponds to \(c_{gz} = 15.7\) m s\(^{-1}\). Based upon this vertical group velocity, the travel time for wave activity to propagate to the observed initial breaking altitude (\(\approx 110\) km) should be a little over 100 min. For case G2, the initial breaking altitude is also approximately 110 km. Since the dominant mesospheric wavelength is 22 km in this case, \(c_{gz} = 12.9\) m s\(^{-1}\), which gives a travel time of about 122 min. The onset of saturation actually occurs at 128 and 160 min in cases G1 and G2, respectively (Fig. 7). Nevertheless, the agreement between the estimates based on vertical group velocity is entirely satisfactory, especially in view of the fact that there is no direct wave activity to propagate upward in the early stages of the simulation. We show below that these arguments can be further refined through a more detailed analysis of the linear far-field response.

Before proceeding with this analysis, it should be noted that some of the results of case G3 cannot be explained if the waves are assumed to be quasi-stationary with respect to the forcing. In case G3, the mesospheric wave field is found upstream of the forcing (Fig. 5), an inconsistent result since for a horizontal wavelength of 110 km (5.2) predicts \(c_{vx}/c_{gz} \approx -0.15\), that is, a zenith angle of \(-8^\circ\). If the waves of Fig. 5 were really quasi-stationary, the mesospheric response should have been centered very slightly downstream from the source rather than upstream from it.

To understand the behavior of the mesospheric wave field in case G3, we need to consider the dependence of the group velocity on frequency \(\omega\), as well as horizontal wavenumber \(k\). When the frequency dependence is taken into account, the group velocities are still given by (5.2), but with \(K^2 [\text{(5.2c)}] \) now equal to

$$K^2 = N^2 / \left( \mu_{rel} - \omega^2 k \right). \tag{5.7}$$

Figure 12 shows \(c_{gx}\) and \(c_{gz}\) as functions of wavenumber and frequency for a background wind speed \(u_{rel} = -50\) m s\(^{-1}\). The group velocities are shown for the frequency range \(\pm 0.05\) min\(^{-1}\). At these frequencies, the power spectrum of the Gaussian pulse (3.1) drops by a factor of 100 with respect to the value at zero frequency (see Fig. 9). Thus, beyond the limits \(\pm 0.05\) min\(^{-1}\), waves are not excited efficiently by (3.1). Inspection of Fig. 12 reveals that the assumption of quasi steadiness is much better justified for short zonal wavelengths than for long wavelengths. At the shorter wavelengths (\(k > 0.2\)), \(c_{gx}\) is essentially independent of frequency over the range \(\pm 0.05\) min\(^{-1}\), while \(c_{gz}\) varies by no more than 10%. At longer wavelengths, on the other hand, \(c_{gx}\) varies by a relatively large amount over the same frequency range, and \(c_{gz}\).
can even change sign when the zonal wavelength becomes greater than \( \approx 60 \text{ km} \) \((k < 0.1 \text{ km}^{-1})\). In particular, waves with \( \omega > 0 \) have positive \( c_{zz} \); in addition, \( c_{zz} \) is larger for these waves than for steady waves.

These considerations provide a satisfactory explanation for the behavior of the wave field in case G3. The dominant wavenumber in this case is 110 km \((k \approx 0.057 \text{ km}^{-1})\), which, from Fig. 12, implies a positive horizontal group velocity \( c_{zz} = 8.4 \text{ m s}^{-1} \) at \( \omega = 0.035 \text{ min}^{-1} \). Although the power at this frequency is a factor of 10 smaller than at \( \omega = 0 \) (Fig. 9), the wavenumber spectrum for zero frequency (Fig. 10) indicates a value of about 100 m² at \( k = 0.057 \text{ km}^{-1} \). Thus, even for \( \omega = 0.035 \text{ min}^{-1} \), the power at \( k = 0.057 \text{ km}^{-1} \) should be of the order of 10 m², comparable to that of the waves that break in the mesosphere in cases G1 and G2. Note that the vertical group velocity at \( \omega = 0.035 \text{ min}^{-1} \) is 10 m s⁻¹, compared with 6.93 m s⁻¹ at \( \omega = 0 \). Thus, waves of relatively low, positive frequency will arrive in the mesosphere before the stationary waves. Note also that \( c_{zz}/c_{zz} = +0.85 \), which is equivalent to a zenith angle of +40°. This prediction is consistent with the position of the mesospheric wave train in Fig. 5.

The inference that the mesospheric wave train in case G3 is dominated by waves with \( \omega \approx 0.035 \text{ min}^{-1} \) is also consistent with the results of the analysis presented in Fig. 8, which indicates that the phase velocity is initially about 8 m s⁻¹. This compares fairly well with the estimate of 10 m s⁻¹ obtained by calculating \( c = \omega/k \), with \( k = 0.057 \text{ km}^{-1} \) and \( \omega = 0.035 \text{ min}^{-1} \).

The combined wavelength obtained from (5.4) and (5.7) at \( \omega = 0.035 \text{ min}^{-1} \) is 19.4 km, much closer to the measured value of 18.3 km (Table 2) than the value computed at zero frequency. Note finally that, although \( c \) increases sharply after the onset of wavebreaking, the Doppler-shifted phase velocity remains approximately constant, so all the foregoing conclusions should still hold.

**d. Evolution of the far-field response**

Further insight into the development of the wave field at high altitudes can be gained by using linear theory to characterize in more detail the far-field response. We show in appendix B that, over a narrow range of zonal wavenumber centered on wavenumber \( k \), the time-dependent amplitude of the vertical displacement field associated with the waves can be expressed as

\[
|\zeta(k; z, t)| = \frac{A \sigma_z}{\sqrt{2z|m_{zz}}\cdot \exp \left( \frac{z}{2H_z} \right) \cdot \exp \left( \frac{-k^2}{4\sigma_z^2} \right)} \\
\times \exp \left[ -\frac{(t - t_m - t_d)^2}{\sigma_t^2} \right], \quad (5.8)
\]

where \( m_{zz} \) denotes the second partial derivative of the vertical wavenumber with respect to \( k \) evaluated at \( \omega = 0, z \) is the altitude above the lower boundary; \( A \) is the amplitude of the lower boundary displacement \( (3.1) \); \( \sigma_z \) is the horizontal half-width, \( \sigma_t \) and \( t_m \) are the temporal half-width and the time of maximum displacement, respectively; and \( t_d(k, z) = z/c_{zz}(k, \omega = 0) \). Note that \( t_d \) represents a time delay in the evolution of the wave field at altitude \( z \) relative to the evolution of the forcing at the lower boundary. This delay is simply the time it takes the waves to reach the altitude \( z \).

Figure 13 shows a plot of \( |\zeta(k; z, t)| \) evaluated at \( z = 110 \text{ km} \), which is close to the altitude where the waves first break in all of the calculations discussed above. The filled circles indicate the wavenumber and the time when saturation first occurs in each case (cf. Fig. 7). In all three cases, (5.8) predicts quite well the wavelength of the waves that break in the mesosphere. It is also evident from the figure that, especially in cases G1 and G2, the largest response to the forcing occurs at longer wavelengths. However, the vertical group ve-

![Fig. 13](image-url)
Locality of these longer waves is slower, so only at later times could they begin to make a significant contribution to the response at 110 km. For example, Fig. 13a implies that the dominant wavelength at 110 km should gradually increase to nearly 30 km (k = 0.21 km\(^{-1}\)) by 260 min. Although not shown here, the numerical results do show that longer waves appear later in the simulation. Spectral analysis for case G1 reveals that at 240 min (the end of the numerical integration) the dominant horizontal wavelength is 25–30 km at breaking altitudes.

Figure 13 shows that the maximum values of |\(\zeta\)| at the onset of breaking are about 1100 m at \(k = 0.4\) km\(^{-1}\) \((\lambda_\theta = 15.7\) km\) in case G1 and 1600 m at \(k = 0.28\) km\(^{-1}\) \((\lambda_\theta = 21.6\) km\) in case G2. Straightforward application of the linear saturation criterion [\((5.6b)\)], indicates that breaking should begin when \(\zeta \approx 2000\) m. Although the saturation values of |\(\zeta\)| shown in Fig. 13 are smaller than this, it must be borne in mind that waves over a range of \(k\) attain such amplitudes almost simultaneously. For example, in case G1 waves in the range \(0.34 \leq k \leq 0.47\) km\(^{-1}\) \((13.4 \leq \lambda_\theta \leq 18.5\) km\) have amplitudes \(>1000\) m by the initial breaking time, while in case G2 waves in the range \(0.23 \leq k \leq 0.35\) km\(^{-1}\) \((17.9 \leq \lambda_\theta \leq 27.3\) km\) achieve such amplitudes at the start of breaking. Superposition of wave trains over these extended, but still fairly narrow, ranges of wavenumber can easily produce total amplitudes in excess of that needed to precipitate breaking.

In case G3, the dominant wavenumber (\(k = 0.057\) km\(^{-1}\), \(\lambda_\theta = 110\) km\) is also well predicted by (5.8). In this case, the far-field response is concentrated over a very narrow wavenumber range, as expected from the “redness” of the forcing spectrum (Fig. 8). The displacement at the onset of breaking (\(t = 216\) min) is about 2000 m, which compares reasonably well with that estimated from the linear saturation criterion \((\approx 2900\) m\) for the longer vertical wavelength found in this case.

Equation (5.8) can also be used to predict the vertical development of the breaking region shown in Fig. 7. To do this, we plotted the locus of points \((z, t)\) that satisfies (5.8) with \(|\zeta| = 1100, 1600,\) and \(2000\) m for cases G1, G2, and G3, respectively. The results, shown in Fig. 14, are in excellent agreement with the calculations for cases G1 and G2. Note that in case G1 breaking does not occur at 60 km; this is consistent with Fig. 7, where the lower boundary of the breaking region never descends below 65 km. The comparison is much less favorable for case G3. In this case, (5.8) predicts breaking almost simultaneously at all altitudes, in sharp contrast to the numerical results of Fig. 7. The discrepancy is not altogether surprising since, as noted in appendix B, the conditions necessary for the derivation of (5.8) are not met closely at the long wavelengths characteristic of case G3. In particular, at long wavelengths (5.8) underestimates the dispersion due to the dependence of the vertical wavenumber on frequency.

![Fig. 14. Altitude and time of the onset of breaking predicted by (5.8) for cases G1–G3. The predictions are excellent for the short-wavelength cases G1 and G2 but poor for the long-wavelength case (G3). See text for details.](image)

The foregoing analysis suggests that linear wave theory is capable of explaining the evolution of the wave field in all the numerical experiments discussed above up to the time when breaking begins. Not surprisingly, the subsequent development cannot be captured by linear theory, since breaking extensively modifies the background state and prevents wave amplitudes from growing much beyond their saturated values. Nonetheless, the linear analysis provides a very useful tool for sorting out how vertical propagation and dispersion give rise to a nearly monochromatic wave field in the mesosphere even though the forcing at low altitudes is localized in both space and time.

The analysis also makes clear why short horizontal wavelengths dominate breaking at high altitudes, provided that the forcing has enough power at these smaller wavelengths. Thus, in both case G1 and case G2, breaking occurs at wavelengths \(<25\) km, even though there is a factor of 5 difference in the horizontal scale of the forcing at the lower boundary. Only in case G3, where power is virtually nil at wavelengths \(<60\) km, does the dominant wavelength of the breaking waves become large. Since the efficiency of vertical mixing due to wave breaking appears to depend on the fraction of the wavelength affected by breaking (Fritts and Dunkerton 1985), and since we have found that this fraction is larger for short wavelengths, such considerations could be important for assessing the impact of different wave sources on vertical mixing in the upper atmosphere.

6. External damping and the altitude of the turbopause

a. Effect of the vertical sponge layer

To test the effects of damping by the vertical sponge layer, we carried out additional calculations using the
same computational parameters as in case G1, except for the value of the minimum damping time. These calculations are denoted by G4 and G5 in Table 1; they have minimum vertical damping times of 23.3 s and 3.15 s, respectively. Recall that the linear damping associated with the vertical sponge is not merely a numerical device; it is in fact a proxy for molecular diffusion, which becomes important in the upper mesosphere and lower thermosphere. As noted in section 3b, the vertical damping time in case G1 was chosen to match the effect of molecular diffusion on the dominant waves at high altitudes. Thus, the smaller damping times used in cases G4 and G5 correspond to atmospheres that are more dissipative by factors of $e^2$ and $e^3$, respectively. In effect, the vertical damping profile is displaced downward by two scale heights in case G4 and by four scale heights in case G5.

Figure 15 presents a comparison of the vertical evolution of the breaking region in cases G1, G4, and G5. It is clear that the maximum height at which wave breaking occurs is set by the vertical damping and is not due to the presence of the top boundary of the domain. This height decreases steadily with decreasing damping time, from 126 km in case G1, to 115 km in G4, and to 104 km in case G5. The average vertical displacement between these three experiments is 11.33 ± 0.14 km, which matches precisely the displacement in the damping profiles (which have a scale height of 5.7 km). It is also clear from Fig. 15 that the minimum height at which saturated waves are present is the same for the three experiments. Evidently, the different values of damping at these lower altitudes have no significant effect on the dynamics—the computations are effectively inviscid.

In contrast to the strong influence of the vertical damping rate, the amplitude of the forcing has little effect on the highest altitude at which breaking occurs. Although the total power associated with the forcing varies considerably among cases G1–G3, as can be appreciated from Fig. 10, the maximum breaking altitude (Fig. 7) is nearly the same in all three cases. To gain some insight into this behavior, we consider the linearized equation for the vertical propagation of gravity wave activity (Andrews et al. 1987):

$$\frac{\partial A}{\partial t} + \frac{\partial (c_{gr} A)}{\partial z} = -2\delta A,$$  \hspace{1cm} (6.1)

where $c_{gr}$ is the vertical group velocity, $\delta$ is a linear dissipation rate, and the wave activity is given by (Bretherton 1971; Andrews et al. 1987)

$$A = \frac{\rho}{2} \left[ \langle u'^2 \rangle + \langle w'^2 \rangle + \frac{g^2}{N^2} \frac{\partial^2}{\partial \theta^2} \right] \left( u_{rel} - \frac{\omega}{k} \right).$$  \hspace{1cm} (6.2)

where $\langle \cdot \rangle$ denotes horizontal (zonal) averages, $u_{rel} = u_u - \gamma$, $u_0$ is the background wind, and $\gamma$ represents the translational speed of the gravity wave source.

Now, after the onset of breaking, wave amplitudes cease growing with time. Under such quasi-steady-state conditions, the divergence of the wave activity flux must be balanced by dissipation; that is,

$$\delta = -\frac{1}{2A} \frac{\partial (c_{gr} A)}{\partial z} \approx -\frac{c_{gr}}{2} \frac{\partial (\ln A)}{\partial z},$$  \hspace{1cm} (6.3)

where we have assumed that the flux divergence is dominated by changes in $A$ itself. If wave breaking prevents wave amplitude growth above the saturation level (Lindzen 1981), it follows from (6.1)–(6.3) that the corresponding dissipation rate must be given approximately by

$$\delta = \frac{c_{gr}}{2H_b}.$$  \hspace{1cm} (6.4)

As noted by Garcia (1991), (6.4) implies that the dissipation rate due to breaking depends on the group velocity of the gravity waves and not on their amplitude. Since this value remains relatively constant while the vertical profile of damping $\tau^{-1}$, given by (3.6), increases exponentially, it may be anticipated that breaking will cease at the altitude where $\delta = \tau^{-1}$. This is consistent with our finding that the upper limit of wave breaking is approximately constant despite the different amounts of forcing power available in cases G1–G3, but decreases systematically as $\tau^{-1}$ is increased in cases G4 and G5.
To test these ideas more precisely, we estimated the maximum breaking altitude \( z_{\text{max}} \) for each case by setting \( \delta \) equal to \( \tau^{-1} \) and solving for \( z \). This gives

\[
 z_{\text{max}} = z_{\text{top}} + H_s \ln \left( \frac{c_{se}}{2H_p} \tau_{\text{min}} \right),
\]

(6.5)

where \( H_s \) is the density scale height and \( H_s \) is the scale height of the damping profile ([3.6]). The estimates obtained from (6.5) are compared in Table 3 with model results from Figs. 7 and 15. Two estimates are given for case G3. The first uses the vertical group velocity at zero frequency (Table 2), while the second uses the value at \( \omega = 0.035 \text{ min}^{-1} \) (9.9 m s\(^{-1}\), see Fig. 12). The latter choice is motivated by the argument presented in section 5c that, at long horizontal wavelengths, the mesospheric wave train is dominated by waves of positive phase velocity. The agreement between (6.5) and the model results is remarkably good in all cases; for case G3, the agreement improves when the larger value of \( c_{se} \) at \( \omega = 0.035 \text{ min}^{-1} \) is used.

We have also verified that the calculations obey (6.4) in the breaking region by computing the vertical profile of wave activity at different times before and during breaking. We used the \( u, w, \) and \( \theta \) fields produced by the model to calculate \( A(z) \) from (6.2), with \( u_{\text{rel}} = -32 \text{ m s}^{-1} \) and \( \omega = 0 \) in cases G1 and G2, and \( u_{\text{rel}} = -50 \text{ m s}^{-1} \) and \( \omega = 0.035 \text{ min}^{-1} \) in case G3. The use of these values is justified by the finding (section 4c) that the Doppler-shifted phase velocity of the waves remains approximately constant throughout the calculations, even when the phase speed itself undergoes large variations during wave saturation and breaking.

Figure 16 shows the results for case G1. Panel (a) of the figure displays the evolution of the wave activity profile from 140 to 220 min, at intervals of 20 min. At \( t = 140 \text{ min} \), \( A \) varies smoothly with altitude but at later times it decreases very rapidly above 65–70 km. This is consistent with Fig. 2, which shows that the onset of vigorous wave breaking occurs immediately after \( t = 140 \text{ min} \) and extends down to near 70 km. In Fig. 16b we have plotted \( \ln|A| \) for \( t = 200–240 \text{ min} \), at 10-min intervals. The logarithmic plot enhances details at the higher altitudes and facilitates evaluation of the vertical \( e \)-folding height. Note that \( \ln|A| \) is approximately constant with altitude up to 65–70 km, indicating that the waves propagate conservatively below the minimum altitude where breaking occurs. Above this altitude, the wave activity decreases with a scale height very close to the density scale height, \( H_\rho = 6.63 \text{ km} \), used in the calculations and indicated by the dashed lines in the figure.

The results for case G3 are shown in Fig. 17. The wave activity profiles evolve more slowly than in case G1, consistent with the smaller vertical group velocity of the long waves of case G3. When plotted on a logarithmic scale (Fig. 17b), the wave activity is not constant with altitude anywhere, although the rate of decrease is faster above 80 km than below this altitude. Evidently, the vertical propagation of the long waves excited in case G3 is too slow to establish a steady-state wave activity profile before \( t = 270 \text{ min} \). Nevertheless, the behavior above 80 km is very similar to that shown in Fig. 16b. In particular, wave activity de-

---

**Table 3. Maximum breaking altitudes.**

<table>
<thead>
<tr>
<th>Run</th>
<th>( \frac{c_{se}}{2H_p} ) (s(^{-1}))</th>
<th>( \tau_{\text{min}} ) (s)</th>
<th>( z_{\text{min}} ) (km)</th>
<th>( z_{\text{max}} ) (km)</th>
<th>( z_{\text{max}} ) (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>( 1.18 \times 10^{-3} )</td>
<td>172</td>
<td>126</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>( 9.73 \times 10^{-4} )</td>
<td>116</td>
<td>123</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>( 5.23 \times 10^{-4} )</td>
<td>116</td>
<td>119</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>G3*</td>
<td>( 7.54 \times 10^{-4} )</td>
<td>116</td>
<td>121</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>( 1.88 \times 10^{-3} )</td>
<td>23.3</td>
<td>115</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>( 1.18 \times 10^{-3} )</td>
<td>3.15</td>
<td>103</td>
<td>103</td>
<td></td>
</tr>
</tbody>
</table>

* Uses vertical group velocity corresponding to \( \omega = 0.035 \text{ min}^{-1} \). See section 5c.
where $H_p$ is the scale height of the atmosphere and $\nu_0$ is the diffusivity at some reference level $z_0$, then the dissipation rate associated with $\nu$ can be written as

$$\nu(k^2 + m^2 + M^2) = \nu K^2$$

(6.7)

for a wave varying as $\exp[i(kx + mz) + Mz]$, where $M = 0.5/H$, and $k$ and $m$ are the horizontal and vertical wavenumbers, respectively. Since the wave field in the breaking region is nearly monochromatic, (6.7) is a good approximation to the damping time due to molecular diffusion.

Using (6.4) and (6.7), we can derive the atmospheric counterpart of the model result (6.5):

$$z_{\text{turb}} = z_0 + H_p \ln \left[ \frac{c_{zz}}{2H_p \nu_0 K^2} \right],$$

(6.8)

where the factor $1/(\nu_0 K^2)$ plays the same role as $\tau_{\text{min}}$ in (6.5). This expression follows from equating the dissipation rate due to molecular diffusion (6.7) with that due to breaking (6.4), which is in the spirit of the definition of the turbopause as the altitude where molecular diffusion becomes as effective as eddy diffusion (McGrattan 1967). Equation (6.8) can also be written as

$$z_{\text{turb}} = z_0 + H_p \ln \left[ \frac{K_{zz}}{\nu_0} \right],$$

(6.9)

where

$$K_{zz} = \frac{k N m}{K^3} \frac{1}{2H_p} \frac{1}{K^2} = \frac{k u_{\text{max}}^2}{2H_p N^3} \left( 1 + \frac{k^2 + M^2}{m^2} \right)^{-1/2}$$

(6.10)

may be interpreted as a “diffusion coefficient” due to wave breaking. Equation (6.10) is the counterpart of Lindzen’s (1981) expression for $K_{zz}$. It follows from (5.2a) and (5.3), assuming $K \approx m$, and reduces to Lindzen’s formula when $M$, $k \ll m$.

Equations (6.8)–(6.10) imply that the altitude of the earth’s turbopause depends on a relatively small number of parameters characterizing the background atmospheric state and the gravity wave field at high altitudes. If the properties of the wave field (viz., the distribution of power as a function of horizontal wavenumber $k$) are assumed to remain more or less constant with time, then systematic seasonal variations in the altitude of the turbopause may be expected to occur principally as a result of changes in the properties of the background state. Since the magnitude of the zonal mean wind varies significantly between the solstices (when mean zonal winds are strong) and the equinoxes (when they are weak), large changes in the height of the turbopause may be expected as a function of season. Other things being equal, (6.8) predicts that a factor of 2 increase in the magnitude of $u_{\text{max}}$ between equinox and solstice would raise $z_{\text{turb}}$ by $2.77H_p$, or almost 20 km.

It would be of interest to investigate whether this prediction is borne out by observations.
7. Summary

We have used a two-dimensional, nonhydrostatic model to study the propagation and breaking at high altitudes of gravity waves forced at tropopause level. The waves are forced by a deflection of the lower boundary, which is Gaussian in time as well as in space (with a time half-width $\sigma_t = 1$ h and spatial half-widths $2 \leq \sigma_x \leq 50$ km). They propagate into a background state with uniform easterly zonal winds in the range of 25–50 m s$^{-1}$. Although idealized, the simulations may be viewed as representative of what happens to waves excited in the upper troposphere, above the tropospheric jets, at midlatitudes. The lower speed background wind and static stability profiles are typical of summer solstice conditions. The higher wind speed is more typical of winter solstice (the actual winter flow is westerly, so one must reverse our results right to left). More realistic wind profiles would add the effects of wave filtering and refraction. However, the behavior documented in our study is unlikely to be qualitatively affected by such influences. The principal findings include the following:

1) Although the lower boundary displacement is Gaussian in shape, the wave field at mesopause altitudes is nearly monochromatic. The dominance of a single horizontal wavelength at high altitudes is a consequence of the dispersion of the wave packet forced at the lower boundary. Short horizontal wavelengths ($\lambda_x \approx 20$ km) dominate the initial response at high altitudes whenever the forcing contains sufficient power at the smaller horizontal scales. For broader forcing ($\sigma_x > 10$ km) the power available at short wavelengths is negligible, and the response at high altitudes is dominated by wavelengths longer than 100 km. Examination of the linear dispersion relation indicates that the bias toward short horizontal wavelengths is due to the fact that the vertical group velocity $c_g$ maximizes when the horizontal and vertical wavelengths are equal [see (5.2a) and Fig. 11]. These results may help explain monochromatic wave motions often observed at mesospheric altitudes with horizontal wavelengths under 100 km.

2) The position of the mesospheric wave train relative to the lower boundary forcing is strongly dependent on zonal wavelength. For short waves ($k > 2\pi/50$ km$^{-1}$), the group velocity is nearly independent of frequency, and the response always appears downwind of the forcing (Figs. 2 and 4). For longer waves (and especially for wavelengths > 100 km) the response is found upwind of the forcing (Fig. 5). To explain the behavior of the long waves, the dependence of the group velocity on frequency must be taken into account. For wavenumbers $k < 2\pi/100$ km$^{-1}$, the frequency $\omega$ makes a significant contribution to the intrinsic frequency $k\sqrt{u_{rel} - \omega}$ (where $u_{rel}$ is the background wind relative to the forcing). The response at high altitudes is then dominated by waves with positive $\omega$ and $c_g$, and the resulting zenith angle for group propagation lies upstream of the forcing.

3) For the forcing amplitude employed in the calculations ($A = 200$ m in (3.1)), the waves break at altitudes between about 65 and 120 km. The morphology of the wave field in the breaking region is sensitive to the ratio of horizontal to vertical wavelengths. When the latter are approximately equal, breaking involves almost the entire wave. When the zonal wavelength is much larger than the vertical wavelength, breaking is confined to a relatively small fraction of the wave (cf. Figs. 2, 4, and 5). These differences may be important insofar as the efficiency of mixing due to wave breaking depends on the localization of the ensuing turbulent motion (Fritts and Dunkerton 1985).

4) Once breaking begins, the phase velocity $c$ evolves so as to keep the Doppler-shifted phase velocity $u_{rel} - c$ approximately constant. This behavior, which Fritts and Dunkerton called self-acceleration, has been reported in previous studies (Fritts and Dunkerton 1985; Walterscheid 1990). Our results suggest that self-acceleration constitutes an actual change in the speed of the wave form and is not due to the presence of a spectrum of frequencies, each with its own critical level.

5) Throughout the breaking region, the wave activity profiles decay with a scale height very close to the atmospheric scale height $H_x$, much as predicted by linear saturation arguments (e.g., Lindzen 1981). This behavior is present in all of our calculations, although the time required to set up a “saturated profile” of wave activity depends on the vertical group velocity of the waves (cf. Figs. 16 and 17).

6) Finally, wave breaking ceases abruptly at the altitude where external dissipation becomes faster than the dissipation rate due to breaking, $c_g/2H_x$, derived from saturation arguments [(6.4)]. In the model, external dissipation is specified as a linear damping rate intended to mimic the effect of molecular diffusion in the real atmosphere. We find that the altitude where wave breaking ceases—the model turbopause—can be estimated rather accurately as the altitude where the external dissipation rate equals $c_g/2H_x$. Since $c_g$ has a strong dependence on $u_{rel} - c$, this implies that the altitude of the earth’s turbopause may change significantly as a function of season.

The calculations and analysis presented in this paper have emphasized the behavior of gravity waves as they propagate away from the forcing region, disperse, and finally break at high altitudes. As such, they may be viewed as complementary to the recent work of Alexander et al. (1995), who explored in detail forcing mechanisms due to a simulated squall line, and that of Andreassen and collaborators (Andreassen et al. 1994; Fritts et al. 1994; Isler et al. 1994), which addressed the role of wave breakdown in three dimensions. All of these aspects of the gravity wave problem may need
to be taken into account to provide a complete understanding of wave breaking in the middle atmosphere and its effect on the atmospheric mean state.

Acknowledgments. We would like to thank D. R. Durr and R. L. Walterscheid for their comments on the original manuscript. This work was supported in part by National Aeronautics and Space Administration Grants L-99120C and NAGW-4060. The National Center for Atmospheric Research is supported by the National Science Foundation.

APPENDIX A

The Effects of Grid Resolution

To illustrate that our standard resolution of 0.625 km is sufficient for determining the wave field characteristics discussed in sections 4–6, we compare case G1 with two other cases, one at lower resolution (LR) and one at higher resolution (HR). Except for the grid resolution, all parametric settings in cases LR and HR are set identical to those of case G1. The resolutions are 1.25 km in case LR and 0.3125 km in case HR. While all three cases began with a time step of 20 s, minimum time steps of 10 and 3 s were used in cases LR and HR, respectively. The minimum time step for case G1 was 6 s. Since our numerical model's truncation error is formally second order, case HR represents about a factor of 4 improvement in accuracy over case G1, and about a factor of 16 improvement in accuracy over case LR.

Figures A1a,b show the maximum zonal and vertical winds, respectively, computed in each of the three cases as functions of time. These wind extrema provide a conservative test for convergence because they are extremely sensitive to the details of the flow. For the first 60 min of the simulation, the three cases are indistinguishable at the scale of the figure. At 100 min, case LR shows noticeably smaller wind perturbations (compared to the basic state, which is \( u_0 = -25 \text{ m s}^{-1} \)) than the other two cases. After this time, the winds in cases G1 and HR depart rapidly from the basic state, whereas those in case LR change slowly. This rapid departure from the basic state represents the period of nonlinear wave growth immediately preceding breaking. Whereas mature breaking morphology is established by 160 min for case G1 (section 4a), in case LR this occurs much later, at 230 min. By contrast, the wind extrema in cases G1 and HR remain close throughout the nonlinear wave growth phase. Case HR reaches a state of mature breaking slightly earlier, at 155 min. Although wind extrema during breaking are slightly greater than in case G1, the overall behavior is much the same. Figures A2a,b show the potential temperature fields when mature breaking is first established for cases LR (at 230 min) and HR (at 155 min). These figures should be compared to Fig. 2c for case G1. Once again, we see that cases G1 and HR are similar, whereas the result for case LR is markedly different.

![Fig. A1. Evolution of the extrema of the vertical and horizontal wind for cases LR, G1, and HR. Cases LR and HR are identical to G1 except that the grid resolutions are 1.25 and 0.3125 km, respectively, instead of 0.625 km. The wind field evolution is nearly identical until breaking begins at \( t = 120 \text{ min} \). After that time, the LR calculation underestimates the extrema.](image-url)
0.625 km, serves to illustrate this point. This final, low horizontal resolution case (LHR) uses the same parameters as case G2 (Table 2). Although not shown here, the results from LHR show qualitative behavior very similar to LR above. Even though the vertical resolution is sufficient to resolve breaking, the solution shows pathological behavior and breaking is severely repressed, due to insufficient resolution.

APPENDIX B

The Far-Field Response to Lower Boundary Forcing

The general solution to the linearized version of the gravity wave equations (2.1) and (2.2) can be written in terms of a Fourier integral over wavenumber $k$ and frequency $\omega$:

$$\zeta(x, z, t) = A \exp \left( \frac{z}{2H_p} \right) \int b(\omega) \exp(-i\omega t) d\omega$$

$$\times \int a(k) \exp \{i[kx + m(k, \omega)z]\} dk,$$

where $\zeta$ is the vertical displacement field; $a(k)$ and $b(\omega)$ are Fourier transforms in wavenumber and frequency, respectively; $A$ is the amplitude; $z$ denotes height above the lower boundary; and the factor $\exp(z/2H_p)$ accounts for the growth of conservative waves in an atmosphere with density scale height $H_p$.

At $z = 0$, (B1) must satisfy the lower boundary condition (3.1) and (3.2); that is,

$$A \int b(\omega) \exp(-i\omega t) d\omega \int a(k) \exp(i k x) dk$$

$$= A \exp \left( \frac{-x^2}{\sigma_x^2} \right) \exp \left[ \frac{-(t - t_m)^2}{\sigma_t^2} \right].$$

It follows that the wavenumber and frequency Fourier transforms are

$$a(k) = \frac{\sigma_x}{2\sqrt{\pi}} \exp \left( \frac{-k^2\sigma_x^2}{4} \right)$$

and

$$b(\omega) = \frac{\sigma_t}{2\sqrt{\pi}} \exp \left( \frac{-\omega^2\sigma_t^2}{4} \right) \exp(i\omega t_m).$$

Now consider the behavior of the integral over wavenumber $k$ in (B.1) for fixed frequency $\omega$:

$$I_t = \int a(k) \exp[i\phi(k)] dk,$$

where $\phi(k) = kx + m(k, \omega)z$ is the phase of the Fourier component of wavenumber $k$. In the vicinity of the forcing, the response is locally confined, but as $z$ increases, $\exp(i\phi)$ varies more rapidly with $k$ than does $a(k)$. Thus, for $z \gg 0$, the major contributions to the integral $I_t$ must come from narrow bands of $k$ where $\phi(k)$ is approximately constant.

Expanding the phase about wavenumber $k$,

$$\phi(k') = \phi(k) + \frac{\partial \phi}{\partial k} (k' - k) + \frac{1}{2} \frac{\partial^2 \phi}{\partial k^2} (k' - k)^2$$

and assuming that $\partial \phi/\partial k = 0$, (B5) can be rewritten as

$$I_t = \exp[i(kx + mz)]a(k)$$

$$\times \int \exp \left[ \frac{i\phi_{kk}(k' - k)^2}{2} \right] dk' \quad (z \gg 0),$$

where $\phi_{kk} = \partial^2 \phi/\partial k^2$.

The solution of (B7), obtained by the method of steepest descent (e.g., Lighthill 1980), is

$$I_t = \sqrt{\frac{2\pi}{|m_{kk}|}} a(k) \exp[i(kx + mz)]$$

$$\times \exp \left[ \frac{i\pi}{4} \text{sgn}(m_{kk}) \right].$$

Recall that (8) holds for $\partial \phi/\partial k = 0$, that is, it holds for values of $x$ and $z$ that satisfy
\[ x + \frac{\partial m}{\partial k} z = 0. \]  \hspace{1cm} (B9)

Now, for fixed frequency,
\[ \frac{\partial \omega}{\partial k} dk + \frac{\partial \omega}{\partial m} dm = c_{\varepsilon} dk + c_{\varepsilon} dm = 0, \]  \hspace{1cm} (B10)

so (B9) is equivalent to
\[ \frac{x}{z} = \frac{c_{\varepsilon}}{c_{\varepsilon}}. \]  \hspace{1cm} (B11)

Thus, the "stationary phase" solution (B8) is valid along the group velocity vector for wavenumber \( k \). If we now substitute (B8) into (B1), the general solution for the far-field response becomes
\[
\zeta(x, z, t) = A \cdot \exp \left( \frac{z}{2H_p} \right) \cdot a(k) \exp(ikx) \\
\times \int b(\omega) \exp(-i\omega t) \sqrt{\frac{2\pi}{z|m_{\varepsilon}}} \exp \left\{ i \left[ m(k, \omega)z + \frac{\pi \text{sgn}(m_{\varepsilon})}{4} \right] \right\} d\omega. \]  \hspace{1cm} (B12)

Equation (B12) can be simplified if it is assumed that \( b(\omega) \) falls off rapidly with \( \omega \), that is, if the frequency spectrum is sufficiently "red." Then, the vertical wavenumber \( m \) can be approximated by expanding about its value at zero frequency:
\[ m(k, \omega) = m(k, 0) + \frac{\partial m}{\partial \omega} \bigg|_0 \omega. \]  \hspace{1cm} (B13)

If, in addition, the frequency dependence of \( m_{\varepsilon} \) is ignored, the integral over frequency in (B12) may be written as
\[
I_z = \sqrt{\frac{2\pi}{z|m_{\varepsilon}}} \exp \left\{ i \left[ m(k, 0)z + \frac{\pi \text{sgn}(m_{\varepsilon})}{4} \right] \right\} \\
\times \int b(\omega) \exp \left\{ -i\omega(t - \frac{\partial m}{\partial \omega} \bigg|_0 \omega) z \right\} d\omega \\
= \sqrt{\frac{2\pi}{z|m_{\varepsilon}}} \exp \left\{ i \left[ m(k, 0)z + \frac{\pi \text{sgn}(m_{\varepsilon})}{4} \right] \right\} \\
\times \int b(\omega) \exp[-i\omega(t - t_\varepsilon)] d\omega, \]  \hspace{1cm} (B14)

where
\[ t_\varepsilon = \frac{\partial m}{\partial \omega} \bigg|_0 \cdot \frac{1}{c_{\varepsilon}(k, 0)} \cdot z \]  \hspace{1cm} (B15)

and \( c_{\varepsilon}(k, 0) \) is the group velocity at zero frequency.

The validity of (B14) depends on the assumption that the dependence of \( m_{\varepsilon} \) on frequency can be ignored. This condition does not hold equally well for all values of the wavenumber \( k \). Figure B1 shows \( m_{\varepsilon}(\omega) \) for different values of the horizontal wavenumber. At wavenumbers typical of the dominant waves in cases G1 and G2 of section 4, \( m_{\varepsilon} \) is indeed nearly independent of frequency. However, the same is not true of the smallest wavenumber \( k = 0.06 \text{ km}^{-1} \), characteristic of the long waves of case G3. At this wavenumber, \( m_{\varepsilon} \) varies strongly with frequency and is much larger at most frequencies than it is at \( \omega = 0 \). Thus, (B14) and any results that depend on it must be applied with caution to long waves.

With this caveat in mind, we proceed by substituting (B4) into (B14) and carrying out the integration:
\[
I_z = \sqrt{\frac{2\pi}{z|m_{\varepsilon}}} \exp \left\{ i \left[ m(k, 0)z + \frac{\pi \text{sgn}(m_{\varepsilon})}{4} \right] \right\} \\
\times \exp \left[ \frac{-(t - t_\varepsilon - t_\varepsilon)}{\sigma_i^2} \right]. \]  \hspace{1cm} (B16)

Finally, using (B3), (B4), and (B16) in (B12), the far-field solution for wavenumber \( k \), time \( t \), and altitude \( z \) can be written as follows:
\[
\zeta(z, t; k) = \frac{A\sigma_i}{\sqrt{2\pi|m_{\varepsilon}|}} \exp \left( \frac{z}{2H_p} \right) \\
\times \exp \left\{ i \left[ kx + mz + \frac{\pi \text{sgn}(m_{\varepsilon})}{4} \right] \right\} \\
\times \exp \left( -\frac{k^2 \sigma_i^2}{4} \right) \cdot \exp \left[ \frac{-(t - t_\varepsilon - t_\varepsilon)}{\sigma_i^2} \right], \]  \hspace{1cm} (B17)

where it is understood that \( m \) and \( m_{\varepsilon} \) are evaluated at \( \omega = 0 \).

---

**Ft. B1.** Variation of \( m_{\varepsilon} \) as a function of frequency evaluated at the dominant wavenumbers of cases G1–G3. Here, \( m_{\varepsilon} \) is nearly independent of frequency in cases G1 and G2 \((k = 0.028, 0.42 \text{ km}^{-1})\), but it varies strongly with frequency in case G3 \((k = 0.057 \text{ km}^{-1})\).
For some purposes, it is more convenient to know the amplitude, $|\psi| = \sqrt{\xi z^3}$, at wavenumber $k$. From (B17) this is

$$|\psi(z, t; k)| = \frac{A_s}{\sqrt{2\pi} \sigma_0} \exp\left(\frac{z}{2H_s}\right) \exp\left(-\frac{k^2\sigma_0^2}{4}\right) \times \exp\left(-\frac{(t - \tau - t_0)^2}{\sigma_t^2}\right).$$

(B18)

REFERENCES


from NCAR Information Services, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.


