

Equatorward Propagation of Inertia–Gravity Waves due to Steady and Intermittent Wave Sources

OLIVER BÜHLER

Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York

(Manuscript received 18 July 2002, in final form 6 January 2003)

ABSTRACT

A simple ray-tracing model for the equatorward propagation of inertia–gravity waves in the lower stratosphere is investigated. The model is based on a zonally symmetric wave source and incorporates radiative wave damping. It is shown that steady extratropical wave sources are able to produce spectra of potential energy that exhibit a conspicuous peak at the equator, which resembles the spectral peaks that are often observed in field data. This reinforces the recent suggestion by other authors that such peaks are caused by the latitudinal variation of the local Coriolis parameter, though these authors did not consider horizontal wave propagation. Notably, horizontal wave propagation can produce equatorial peaks even with strictly extratropical wave sources.

Wave source intermittency is then investigated by allowing for a time-dependent wave source. The source is treated statistically as a stationary random process and a number of general comments are made with respect to the impact of intermittency on gravity wave parameterizations in general circulation models and on the expected variance of observational estimates of energy spectra.

Finally, the possibility of an observational bias toward waves with lower group velocities (as recently suggested by other authors) is examined using a simple example of topographic wave generation by time-varying surface winds.

1. Introduction

In a recent paper by Alexander et al. (2002, hereafter ATV) the observed latitudinal structure of potential energy spectra in the lower stratosphere is discussed. These zonally averaged spectra show a conspicuous peak at tropical latitudes, centered more or less at the equator. Under the plausible assumption that the observed spectra are due to low-frequency, inertia–gravity waves, ATV make the original proposal that the observed peak may in some way be caused by the vanishing of the Coriolis parameter at the equator. Specifically, they discuss two separate mechanisms by which the vanishing of the Coriolis parameter can lead to peaks in the spectrum. The first is a straightforward assumption about gravity wave sources in the troposphere or at ground level: if one assumes that these sources are present at all temporal frequencies, then near the equator there are “more” gravity waves that can be excited within the admissible intrinsic frequency band $N > \hat{\omega} > f$, where N is the buoyancy frequency, $\hat{\omega}$ is the intrinsic frequency of the gravity waves, and f is the local Coriolis parameter. This leads to an energy peak at the equator, where f goes to zero. The second mechanism

discussed in ATV involves the intermittency of wave sources, which ATV argue could also lead to an equatorial peak in the observed spectrum, though this mechanism is less straightforward than the first (cf. discussion in section 3c). As in most gravity wave studies, only vertical wave propagation was considered in detail by ATV.

Now the aim of the present paper is twofold. First, the discussion of the spectral peak in ATV is complemented by demonstrating that it is in fact possible to obtain potential energy spectra with a conspicuous peak at the equator even in a situation in which there are *no* equatorial gravity wave sources at all. This becomes possible once the horizontal propagation of gravity waves is taken into account, and this is studied in some detail here by a simple toy model consisting of a zonally symmetric wave source at various extratropical latitudes. The meridional structure of the resultant wave trains is first computed using ray tracing without dissipation and then with dissipation due to radiative damping, which is significant near the equator. A robust and conspicuous equatorial peak in potential energy is found in all cases, and it is argued that similar peaks would also occur in less idealized models.

Of course, none of these results imply that the observed wave spectra are in fact predominantly due to horizontally propagating waves. But presumably horizontal propagation makes a contribution of some sort

Corresponding author address: Oliver Bühler, Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012-1185.
E-mail: obuhler@cims.nyu.edu

to the observed spectra, and it is useful to present an initial study of it in the simplest possible idealized setup, with all other effects absent purely in order to limit distraction from the new effect. Overall, the present paper reinforces the original suggestion made by ATV about an eventual explanation of the observations in terms of the variable f .

The second aim of this paper is a discussion of wave source intermittency, which can be studied in detail in the present toy model. Intermittency is well recognized to be important for consistent gravity wave parameterizations in general circulation models and also for the interpretation of observed data (e.g., the discussion and references in Fritts and Alexander 2002; Alexander and Dunkerton 1999; ATV). In the present paper, intermittency is discussed in simple terms via time-dependent wave sources that are modeled as stationary random processes. This yields average wave energies that are described by the expectation values of the corresponding random wave field. The difficult question of how to parameterize wave amplitudes for use in amplitude-dependent, nonlinear wave saturation and dissipation schemes (e.g., Alexander and Dunkerton 1999) is discussed for such random sources, and a simple approximation for wave sources with known mean and variance is proposed. It is argued that intermittent sources should be treated statistically and that estimating statistical source parameters such as the expected wave activity flux and the source autocorrelation function may be a useful objective of observational campaigns and numerical investigations.

The impact of intermittency on observed wave spectra is also discussed, noting in particular the increased variance of space-averaged estimates for intermittent waves with higher group velocity. However, no general observational bias against waves with higher vertical group velocity is found, in contrast with the “probability of observation” argument proposed in ATV. This is demonstrated here by a model example of topographic wave generation due to intermittent, time-dependent surface winds, in which such a bias is absent.

The plan for the paper is as follows. The next section describes the toy model setup and the wave train structure under nondissipative and dissipative conditions. Only steady, nonintermittent wave sources are considered there. Following this is a study of wave source intermittency in section 3, which discusses parameterizations, observations, and group-velocity-related effects. Finally, some concluding remarks are given in section 4.

2. Model wave source and wave train structure

A zonally symmetric wave source is considered at a wave launch latitude $\phi_0 < 0$ in the Southern Hemisphere. The source generates inertia-gravity waves propagating due north, that is, equatorward. No particular physical forcing mechanism is meant to correspond

to this hypothetical source, although convection could certainly launch waves with significant meridional group velocities. For convenience, the vertical propagation of the waves is neglected, either because one implicitly considers vertically integrated wave structures, or because one considers vertically trapped waves, or because of a sufficiently deep source. However, considering the vertical propagation explicitly would not change the main features discussed later.

Also for convenience, there are no background winds, the buoyancy frequency N is constant, and the vertical wavenumber of the waves is high enough to allow the neglect of scale height effects and to allow the use of slowly varying, Wentzel–Kramers–Brillouin (WKB) theory and hence ray tracing. Furthermore, the gravity wave dispersion relation is approximated by

$$\hat{\omega}^2 = N^2 \frac{l^2}{m^2} + f^2, \quad (1)$$

where $\mathbf{k} = (0, l, m)$ is the wavenumber vector lying in the local y - z plane and $f = 2\Omega \sin\phi$ is the local Coriolis parameter in terms of the earth’s rotation rate Ω and latitude ϕ . Equation (1) is valid for waves with $l^2 \ll m^2$, that is, for waves with a much shorter vertical than horizontal wavelength. Without loss of generality, it is assumed that l and $\hat{\omega}$ are both positive and then (1) can be rewritten as

$$l = \frac{|m|}{N} \sqrt{\hat{\omega}^2 - f^2}. \quad (2)$$

The meridional group velocity $v_g > 0$ is given by

$$v_g = \frac{\partial \hat{\omega}}{\partial l} = \frac{\hat{\omega}^2 - f^2}{l \hat{\omega}} = \frac{N}{|m|} \frac{\sqrt{\hat{\omega}^2 - f^2}}{\hat{\omega}} \quad (3)$$

and the waves propagate in the meridional direction according to the usual ray-tracing equations

$$\begin{aligned} \frac{d\phi}{dt} &= \frac{v_g}{R}, & \frac{d\hat{\omega}}{dt} &= 0, & \frac{dl}{dt} &= -\frac{f}{\hat{\omega}R} \frac{df}{d\phi}, \\ \frac{dm}{dt} &= 0, \end{aligned} \quad (4)$$

where R is the earth’s radius. As $\hat{\omega}$ and m are propagation invariants, the third factor in (4) expresses the rate of change in l necessary in order to keep (2) satisfied as $f(\phi)$ changes along the ray. Both the meridional wavenumber l and the meridional group velocity v_g increase as the waves approach the equator and f^2 decreases.

All parameters and wave properties at the launch altitude ϕ_0 are denoted by a subscript zero, for example, $f_0 = f(\phi_0)$ and so on. Clearly, $\hat{\omega} = \hat{\omega}_0$ and $m = m_0$ throughout in this paper, so subscripts on these quantities are not strictly needed. All waves are launched from ϕ_0 with a particular frequency given by the ratio

$$\mu = \frac{\hat{\omega}_0}{f_0} = \frac{\hat{\omega}}{f} > 1, \quad (5)$$

which will be kept the same throughout this paper. This means that the wave source at ϕ_0 is monochromatic. A numerical value for μ will be given in (12) later.

a. Nondissipative wave train

The wave amplitudes are determined from the usual WKB wave action conservation law (e.g., Andrews et al. 1987), which for nondissipative conditions (radiative damping is considered later in section 2b) away from the wave source reads

$$\frac{\partial(E/\hat{\omega})}{\partial t} + \frac{1}{R \cos \phi} \frac{\partial(\cos \phi v_g E/\hat{\omega})}{\partial \phi} = 0, \quad (6)$$

where $E(\phi, t)$ is the standard wave disturbance energy density per unit volume averaged over a wave cycle. As horizontal background density variations are neglected in this paper, one could equally well use the energy density per unit mass in (6). Equation (6) contains the $\cos \phi$ factors due to spherical geometry. This is easily done here and makes the results more accurate. However, a simpler tangent-plane approximation would lead to qualitatively the same results below; that is, the spherical geometry is not essential.

For the envisaged monochromatic wave source the invariance of $\hat{\omega}$ along group velocity rays implies that $\hat{\omega}$ has the same value everywhere and at all times and hence $\hat{\omega}$ can be factored out of (6). This implies the global conservation of wave disturbance energy under this condition. For a steady wave source (averaged over a wave cycle, of course), Eq. (6) can be simplified further by setting the time derivative to zero in order to obtain

$$\frac{\partial(\cos \phi v_g E)}{\partial \phi} = 0, \quad (7)$$

which describes the constancy of wave energy flux that must be obtained after a steady wave train has established itself. Equation (7) has solution

$$\frac{E(\phi)}{E_0} = \frac{\cos \phi_0 v_g(\phi_0)}{\cos \phi v_g(\phi)}, \quad (8)$$

where $E_0 = E(\phi_0)$ is the prescribed wave energy density at the source. Substituting for v_g from (3) and using (5) gives

$$\frac{E(\phi)}{E_0} = \frac{\cos \phi_0}{\cos \phi} \sqrt{\frac{\hat{\omega}^2 - f_0^2}{\hat{\omega}^2 - f^2}} = \frac{\cos \phi_0}{\cos \phi} \sqrt{\frac{\mu^2 - 1}{\mu^2 - \frac{\sin^2 \phi}{\sin^2 \phi_0}}}. \quad (9)$$

This shows how $E(\phi)$ varies due to two unrelated physical effects, as follows. The first factor in (9) captures the lengthening of latitude circles as the waves approach the equator. This dilutes the wave energy and decreases E . The second factor captures the increasing group velocity, which also dilutes the wave energy and hence also decreases E .

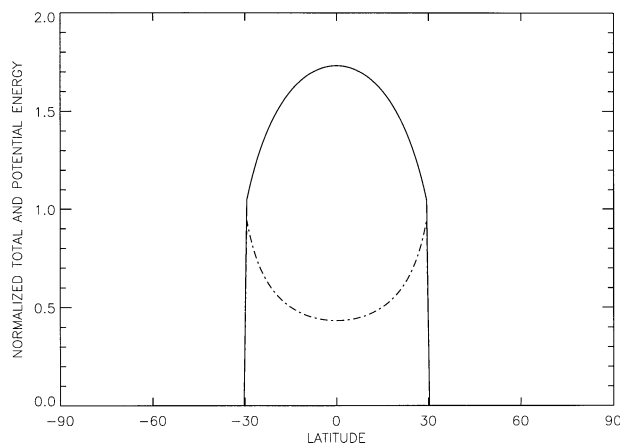


FIG. 1. Normalized total and potential energy as a function of latitude. The waves are launched at $\phi_0 = -30^\circ$ and propagate without dissipation until they reach $\phi_0 = +30^\circ$, where they are terminated. Potential energy, exhibiting maximum at equator (solid line); total energy, exhibiting minimum at equator (dash-dot lines).

Now, many observational techniques only measure E_p , that is, the potential energy part of E that is visible in temperature fluctuations. As is well known, inertia-gravity waves do not obey equipartition between kinetic and potential energy. Instead, the relations

$$E = E_p + E_k, \quad E_p = \frac{1}{2}E \left(1 - \frac{f^2}{\hat{\omega}^2}\right),$$

$$E_k = \frac{1}{2}E \left(1 + \frac{f^2}{\hat{\omega}^2}\right) \quad (10)$$

hold, where E_k is the kinetic energy. Indeed, in the limit of inertial oscillations with $\hat{\omega} = f$ the potential energy is exactly zero. It is clear from combining (6) and (10) that E_p is not conserved globally, in contrast with the conserved E .

The absence of equipartition has a profound impact on the inferences that can be made from observations of E_p . This is especially true for varying f , as is the case here. Explicitly, one obtains

$$\frac{E_p(\phi)}{E_p(\phi_0)} = \frac{E}{E_0} \frac{\hat{\omega}^2 - f^2}{\hat{\omega}_0^2 - f_0^2} = \frac{\cos \phi_0}{\cos \phi} \sqrt{\frac{\mu^2 - \frac{\sin^2 \phi}{\sin^2 \phi_0}}{\mu^2 - 1}} \quad (11)$$

in the present case. The second factor in (11) now leads to an increase in E_p toward the equator, contrary to the trend in E . This can be seen most clearly in a situation where $\phi - \phi_0$ is small and hence a tangent-plane approximation is valid, because then (9) and (11) imply that in fact $EE_p = \text{constant}$. This means that energy and potential energy vary inversely proportional to each other as a result of both the changes in group velocity and the fact that the waves become less affected by f near the equator.

The ratios (9) and (11) are plotted in Fig. 1 as a function of ϕ for a case with $\phi_0 = -30^\circ$ and

$$\mu = 2/\sqrt{3} \approx 1.15. \quad (12)$$

This particular value of μ is used throughout this paper. It has been chosen purely for convenience because the aim is to launch low-frequency inertia-gravity waves with $\hat{\omega}$ close to f_0 (and hence μ close to unity), and because this value makes the second factor in (11) equal to 2 at the equator (where $\phi_0 = 0$). Figure 1 clearly indicates the different behavior of E_p and E : the former peaks at the equator while the latter has a trough there.

The amplitudes are symmetric across the equator and the wave train would actually continue to propagate northward past $\phi = +30^\circ$, until it is reflected at a latitude where $\sin^2 \phi = \mu^2 \sin^2 \phi_0$ and hence l goes through zero. To avoid complications due to this reflection (where WKB theory cannot be used to predict amplitudes), the wave train is simply terminated at $\phi = |\phi_0|$. This wave train termination has no physical basis in a nondissipative theory, but will be seen to matter less once damping has been introduced in the next section. It can be noted in passing that a more complete linear theory that takes reflection into account would have to be based on meridional waveguides rather than on simple ray tracing.

b. Radiative damping

Linear dissipation due to radiative damping near the equator alters the wave train appearance significantly, as will be shown here. The temperature disturbance associated with the gravity waves upsets the radiative transfer equilibrium of the stable background temperature field and this induces radiative transfer processes that seek to dampen the temperature disturbance on an exponential time scale $1/\hat{\alpha}$, which depends on both altitude and on the vertical wavenumber of the waves (cf. Andrews et al. 1987).

The wave energy dissipation density per unit volume and unit time averaged over a wave period is $2\hat{\alpha}E_p$. This yields an amended wave action law (6) that reads

$$\begin{aligned} \frac{\partial(E/\hat{\omega})}{\partial t} + \frac{1}{R \cos \phi} \frac{\partial(\cos \phi v_g E/\hat{\omega})}{\partial \phi} &= -2\hat{\alpha} \frac{E_p}{\hat{\omega}} \\ &= -\alpha \frac{E}{\hat{\omega}}, \end{aligned} \quad (13)$$

where the effective damping rate per unit time (Bühler et al. 1999; Bühler and McIntyre 1999) is defined as

$$\alpha = 2\hat{\alpha} \frac{E_p}{E} = \hat{\alpha} \left(1 - \frac{f^2}{\hat{\omega}^2}\right). \quad (14)$$

The definition of α shows the important fact that low-frequency inertia-gravity waves are significantly less visible to radiative damping than is suggested by the bare damping rate $\hat{\alpha}$. This is because of their low relative levels of potential energy compared to kinetic energy. This fact needs to be borne in mind when using numerical values for the bare damping rate $\hat{\alpha}$, which typ-

ically increases as $\sqrt{|m|}$ in the lower stratosphere (e.g., Fels 1982, 1984). One can also note that the rate α measures the wave-action decay per unit time, so that the decay rate per unit distance is given by α/v_g . This will be used later.

Near the equator $f^2/\hat{\omega}^2$ will become small and the wave train will hence be significantly damped, as will be shown now. For a steady wave train with uniform $\hat{\omega}$ the solution of (13) in terms of E_p is

$$\begin{aligned} \frac{E_p(\phi)}{E_p(\phi_0)} &= \frac{\cos \phi_0}{\cos \phi} \sqrt{\frac{\mu^2 - \frac{\sin^2 \phi}{\sin^2 \phi_0}}{\mu^2 - 1}} \\ &\times \exp \left[- \int_{\phi_0}^{\phi} \frac{R\alpha(\phi')}{v_g(\phi')} d\phi' \right]. \end{aligned} \quad (15)$$

This differs from (11) by the exponential attenuation factor, which is less than unity if $\alpha > 0$. Notably, the effective damping rate per unit distance α/v_g is proportional to $v_g \hat{\alpha}$ in the present case, because $\alpha \propto v_g^2 \hat{\alpha}$. This means that the effective damping rate per unit distance is actually highest at the equator, where the wave is fastest. Indeed, combining (3), (5), and (14), the integrand in the exponential factor in (15) can be rewritten as

$$\frac{R\alpha(\phi')}{v_g(\phi')} = \frac{R\hat{\alpha}|m|}{N} \sqrt{1 - \frac{\sin^2 \phi'}{\mu^2 \sin^2 \phi_0}}, \quad (16)$$

which yields an incomplete elliptical integral of the second kind in (15). This can easily be evaluated numerically. The parameters chosen for this purpose were

$$\begin{aligned} R &= 6300 \text{ km}, & |m| &= 2\pi/2 \text{ km}, \\ \hat{\alpha} &= 0.1/\text{day}, & N &= 2\pi/7 \text{ min}. \end{aligned} \quad (17)$$

These were meant to be typical for the lower stratosphere and they produce a maximal damping rate per unit distance at the equator of $R\alpha/v_g = 0.027$ per degree latitude.

The results are plotted in Fig. 2 for three different launch latitudes $\phi_0 \in \{-60^\circ, -30^\circ, -15^\circ\}$. The solid line gives E_p without damping and the dashed line with damping. The $\phi_0 = -60^\circ$ case is peculiar because for the chosen value of μ it just so happens that in this case the undamped E_p profile is exactly constant, as can be readily verified from (11). As ϕ_0 moves closer to the equator the peak in the undamped E_p gets bigger and eventually approaches an upper limit, which is equal to 2 for the chosen value of μ . The damped curves are clearly not symmetric across the equator. It is obvious from Fig. 2 that the damping has a significant effect on the latitudinal structure of E_p and hence needs to be taken into account if gravity waves are indeed allowed to travel large horizontal distances. Figure 2 also shows that once damping is considered the termination of the wave train at $\phi = |\phi_0|$ introduces only a small error, as the wave amplitudes are quite small there anyway.

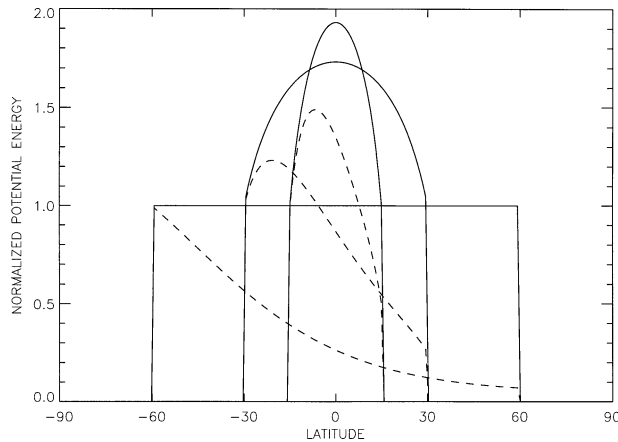


FIG. 2. Normalized potential energy as a function of latitude. The waves are launched at $\phi_0 = \{-60^\circ, -30^\circ, -15^\circ\}$ and terminated at $\phi_0 = \{+60^\circ, +30^\circ, +15^\circ\}$. Potential energy without dissipation (solid lines); potential energy with dissipation due to radiative damping (dashed lines).

c. Multiple wave sources

The previous computation can be repeated for multiple wave sources at different latitudes and the resulting wave energy profiles can be added if one assumes that the wave sources are mutually incoherent. With μ kept constant at all latitudes, each source actually generates waves with a slightly different intrinsic frequency $\hat{\omega} = \mu f_0$. The resultant cumulative energy density at a given latitude hence consists of contributions from a range of intrinsic frequencies.

Such cumulative densities are plotted in Fig. 3. The solid line is again the undamped case and the dashed line is the damped case. The left density results from placing sources of equal strength [as measured by $E_p(\phi_0)$] at all integer latitudes from -60° to -30° . The right density results from doing the same over the larger range from -60° to -15° . Finally, the resultant densities have been symmetrized across the equator by the artifice of putting equal sources at the corresponding positive latitudes. (Such artificial symmetrization could presumably be avoided in a modal approach, as all modes would be symmetrical across the equator.) The dip in the center of the left plot arises because of the radiative damping suffered by the waves. Similar dips can sometimes be seen in observed spectra (cf. references in ATV).

It is quite obvious that by varying the latitudinal source strengths and the other wave parameters a multitude of energy densities can be generated, which could be tuned to agree closely with observations. As noted before, this is no proof that meridional propagation of inertia-gravity waves is relevant to these observations, though it clearly could make a contribution of some sort. The key point that has been illustrated here is that even wave sources outside the equatorial region can robustly produce E_p spectra that peak at the equator.

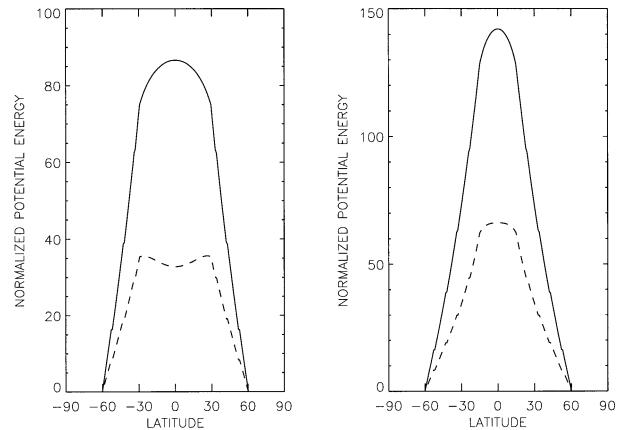


FIG. 3. Normalized potential energy as a function of latitude. (left) Multiple wave sources with uniform strength located symmetrically across the equator at $|\phi_0| \in [30^\circ, 60^\circ]$. (right) The same as in (a) with $|\phi_0| \in [15^\circ, 60^\circ]$. Nondissipative solution (solid lines); with dissipation due to radiative damping (dashed lines).

3. Intermittent wave sources

Real atmospheric gravity wave sources are usually intermittent, that is, they are active only at certain times and in certain places. Therefore, intermittency presents an important issue both for the consistent parameterization of gravity wave sources in general circulation models and for the interpretation of observational field data (cf. Holton 1983; Dunkerton 1989; ATV; Alexander and Dunkerton 1999).

The simple wave propagation problem studied in this paper has the advantage that temporal intermittency can be examined in it in full detail. This leads to a certain clarification of the implications that intermittency has for parameterization and observation, which makes this worthwhile. Temporal intermittency is incorporated here via time-dependent sources, that is, the constant E_0 is replaced by $E_0(t)$. As always in this context, this implicitly presupposes a scale separation between the frequency of the forced waves and the intermittency time-scales present in $E_0(t)$. Similar implicit assumptions underlie the spatial variability that may occur in $E(\phi, t)$. For instance, these assumptions impose well-known observational constraints (e.g., Alisse and Sidi 2000).

Now, it is straightforward to obtain the time-dependent solution of (13), but the important points can already be examined much easier in a local tangent-plane approximation, which gives (if damping is also neglected)

$$\frac{\partial E}{\partial t} + \frac{\partial(\mathbf{v}_g E)}{\partial y} = 0, \quad (18)$$

where $y = R\phi$. For simplicity, the group velocity \mathbf{v}_g is also taken to be constant over the y region of interest. Notably, all the discussion that follows also applies directly to the case of vertical wave propagation and pseudomomentum transport after making the substitutions $y \rightarrow z$, $\mathbf{v}_g \rightarrow \mathbf{w}_g$, and $E \rightarrow kE/\hat{\omega}$.

The elementary solution of (18) is

$$E(y, t) = E_0(t - y/v_g), \quad (19)$$

provided the wave source is located at $y_0 = 0$ and $v_g > 0$, as before. This means that source energy values are simply advected to the right with constant speed v_g .

Now, if the function $E_0(t)$ is treated as a stationary random process (e.g., Yaglom 1962) in time, then (19) defines the random process $E(y, t)$ in time and space. Clearly, the process $E(y, t)$ is stationary in t and homogeneous in y . By using the tangent-plane approximation and by treating the source as a stationary random process, we in essence focus on timescales and space scales over which the statistical character of $E_0(t)$ can be approximated as constant. This models the typical behavior of an atmospheric wave source under slowly varying external conditions, say during a certain month.

Some assumptions about the nonnegative random process $E_0(t)$ are required now to make progress. Let $E_0(t)$ have expected value

$$E_{00} \equiv \mathcal{E}\{E_0(t)\}, \quad (20)$$

and temporal autocorrelation function

$$\begin{aligned} B(t - s) &= B(s - t) \\ &\equiv \mathcal{E}\{[E_0(t) - E_{00}][E_0(s) - E_{00}]\}. \end{aligned} \quad (21)$$

such that the variance of $E_0(t)$ is $B(0)$. Here $\mathcal{E}\{\cdot\}$ denotes the probabilistic expectation operator. The wave energy flux

$$F(y, t) \equiv v_g E(y, t) = v_g E_0(t - y/v_g) \quad (22)$$

is used below, and it is trivially proportional to $E(y, t)$ in the present setting with fixed v_g . Its expected value is $v_g E_{00}$ and its autocorrelation function is $v_g^2 B(t)$.

a. Parameterization

Gravity wave parameterizations are usually formulated based on wave theory with steady (i.e., nonintermittent) sources. That is to say, although the parameterized sources may change slowly at a given location in response to changes in the overall conditions (prevailing surface winds, strength of convective activity, etc.), the (vertical) wave propagation is usually modeled in terms of a steady wave train. This means that parameterizing an intermittent source requires a nontrivial modeling of the intermittent physical source by a steady parameterized source. How to do this consistently and efficiently is subject to some current debate (e.g., ATV and references therein; Fritts and Alexander 2002; Alexander and Dunkerton 1999).

In general the parameterized wave flux should be equal to the expected value of the true flux $\mathcal{E}\{F(y, t)\}$, which in the present case of a stationary source does not depend on y or t . When this condition is satisfied, then the steady model wave train will, on average, deliver the same wave flux as the true intermittent source.

This is the natural condition to use if one considers, for instance, the vertical transport of zonal pseudomomentum: delivering, on average, the correct amount of pseudomomentum to higher altitudes means that, on average, the correct amount of zonal mean-flow forcing is exerted on the middle and upper atmosphere.

As a simple idealized example, consider the three different wave fields depicted in Fig. 4. These are snapshots at a fixed time t , and the left panel shows the wave field due to a steady, nonintermittent source at $y_0 = 0$. Disregarding amplitudes for a moment, the middle panel shows the same wave but now due to an intermittent source, which is switched off half of the time and switched on during the other half. The right panel shows the result of an even more intermittent source, which is only switched on one-quarter of the time. Such on-off sources are described by an intermittency parameter γ that is equal to the fraction of time that the source is switched on. In the three presented cases

$$\gamma = \{100\%, 50\%, 25\%\}. \quad (23)$$

Of course, there are many different random processes that have the same intermittency parameter γ . However, knowledge of γ is enough to determine E_{00} and the variance $B(0)$ [though not $B(s)$ for nonzero s]. This is possible because here $E_0(t)$ takes only two values: zero and E_s , say, with corresponding fractions of time $(1 - \gamma)$ and γ . This means that E_s is the source energy spectrum in the present context, corresponding to the ‘‘on’’ phase of the physical wave source. (One can note in passing that in the parameterization literature the term ‘‘source spectrum’’ often refers to something different, namely, to the prescribed steady values of E at a certain lower level in the vertical.)

The expectation E_{00} is then given by

$$\begin{aligned} E_{00} &= \mathcal{E}\{E_0(t)\} = (1 - \gamma)0 + \gamma E_s \\ &= \gamma E_s \quad \text{and} \end{aligned} \quad (24)$$

$$\begin{aligned} B(0) &= \mathcal{E}\{[E_0(t) - E_{00}]^2\} \\ &= (1 - \gamma)(-\gamma E_s)^2 + \gamma(E_s - \gamma E_s)^2 \\ &= \gamma E_s^2(1 - \gamma). \end{aligned} \quad (25)$$

The nonintermittent $\gamma = 1$ case has zero variance, as it should. Now, the values of the expected spectra for the wave trains in Fig. 4 are given by

$$\mathcal{E}\{E\} = \gamma E_s \quad \text{and} \quad \mathcal{E}\{F\} = v_g \gamma E_s. \quad (26)$$

However, the wave amplitudes in Fig. 4 have in fact been adjusted such that the second wave has an amplitude that is larger than the first by a factor of $\sqrt{2}$, and the third amplitude has been increased by another factor of $\sqrt{2}$. This means that the value of E_s (which is proportional to the wave amplitude squared) increases by a factor of 2 from wave to wave. This means that the product γE_s is in fact the same in all three cases. In other words, all three cases gives rise to the same pa-

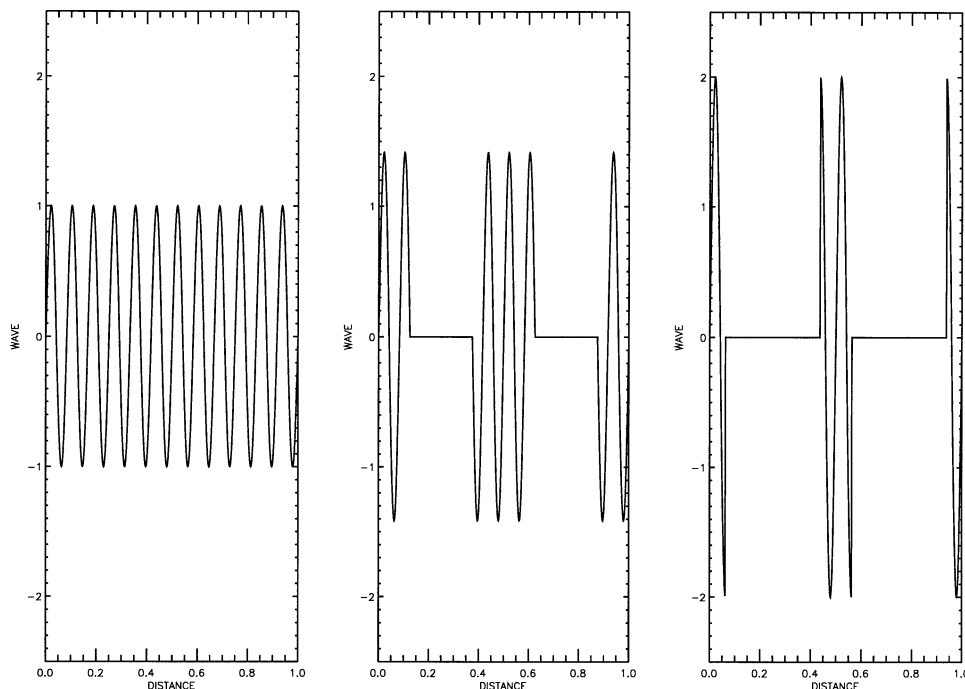


FIG. 4. Wave packets due to hypothetical intermittent wave source at $y_0 = 0$. (left) Steady, nonintermittent source. (middle) Intermittent source with $\gamma = 50\%$. (right) Intermittent source with $\gamma = 25\%$. The wave amplitude increases by a factor of $\sqrt{2}$ from wave to wave, keeping the expected wave energy flux the same in all three cases.

parameterized flux $\mathcal{E}\{F\}$. So all three wave trains are mapped onto the first wave train for the purpose of parameterization with equal expected flux.

This is despite the fact that the wave amplitudes are markedly different in the different cases. This is important for nonlinear effects, such as the amplitude-dependent dissipative processes to do with wave saturation by instability or breaking. Clearly, the individual wave packets in the second and third cases are more likely to break earlier because of their larger amplitudes. This means that the likely profiles of flux convergence due to dissipation are going to be different, which is important for applications such as vertical transport of momentum (Alexander and Dunkerton 1999).

A more general random process $E_0(t)$ may take many different values, and it is not obvious how to choose “the” source spectrum value that characterizes the active source and hence the wave packet amplitudes. Basing a probability distribution for $E_0(t)$ on the expected value and the variance of $E_0(t)$ could be a useful intermediate step in this context. For instance, knowing the values of E_{00} and $B(0)$ allows computing an intermittency parameter γ now defined via (24)–(25) as

$$\gamma = \frac{E_{00}^2}{E_{00}^2 + B(0)}. \quad (27)$$

So this equation could be used for an arbitrary source $E_0(t)$ in order to define an intermittency parameter γ that approximates the source with an on–off source hav-

ing the same expected value and variance. This corresponds to a source spectrum defined as

$$E_s = \frac{1}{\gamma} E_{00} = E_{00} \left(1 + \frac{B(0)}{E_{00}^2} \right). \quad (28)$$

Modeling of wave amplitudes and of nonlinear wave saturation could then follow based on this approximation as if the source were of the simple on–off type.

b. Observation

Determining the statistics of $E_0(t)$ requires suitable observations to be made either in field data or, if feasible, in high-resolution numerical simulations of the physical source processes. Because of wave propagation, there are some interesting answers to the question of how quickly averaged measurements converge to their expected value. By definition, a single observation of $E(y, t) = E_0(t - y/v_g)$ at some y and t has expected value E_{00} and variance $B(0)$. Dividing $B(0)$ by E_{00}^2 gives the normalized variance. Now, repeating such an observation M times (and averaging the results) under the assumption that all observations are mutually uncorrelated reduces the variance to $B(0)/M$. For example, this means that for the on–off model described earlier the normalized root-mean-square error after M such observations is equal to

$$\sqrt{\frac{B(0)}{ME_0^2}} = \sqrt{\frac{1-\gamma}{M\gamma}}. \quad (29)$$

For substantially intermittent sources this shows that a large number of measurements M are needed to bring the expected error below a reasonable threshold. For instance, to estimate E_{00} to within 30% error requires about $M = 10$ measurements for a source with $\gamma = 0.5$, but $M = 30$ if $\gamma = 0.25$. To achieve 10% accuracy requires roughly 10 times more measurements in either case.

Time-averaged observations of $E(y, t)$ have substantially reduced variance once the averaging time T is substantially larger than the integral timescale of the source

$$\tau = \int_0^\infty \frac{B(t)}{B(0)} dt. \quad (30)$$

Specifically, an observational time average

$$\frac{1}{T} \int_t^{t+T} E(y, t') dt' = \frac{1}{T} \int_t^{t+T} E_0 \left(t' - \frac{y}{v_g} \right) dt' \quad (31)$$

has expectation E_{00} and variance

$$\frac{1}{T^2} \int_0^T \int_0^T B(t-s) dt ds, \quad (32)$$

which for $T \gg \tau$ is $\approx B(0)2\tau/T$. So, in terms of variance a single time-averaged observation with averaging time T is equivalent to averaging over $M = T/(2\tau)$ uncorrelated observations of $E(y, t)$.

Space-averaged observations of $E(y, t)$ using an averaging length L are essentially identical to time-averaged observation, after an important scaling. This is because the statistics of

$$\frac{1}{L} \int_y^{y+L} E(y', t) dy' = \frac{1}{L} \int_y^{y+L} E_0 \left(t - \frac{y'}{v_g} \right) dy' \quad (33)$$

are identical to the statistics of a time average with averaging time $T = L/v_g$. This means that the normalized variance of (33) for sufficiently large L will be $\approx B(0)2v_g\tau/L$. There is now an explicit dependence on v_g , which will be discussed next.

c. Group-velocity effects

For intermittent wave sources there is a need to be vigilant with regard to errors that may result from observational biases, such as regarding group-velocity effects (ATV). For instance, as the normalized variance of (33) is proportional to v_g , it is clear that space-averaged observations of fast waves are less accurate than those of slow waves. This is simply because if one considers estimating the expected energy spectra of two different waves with $v_{g1} > v_{g2}$ by using a space average with fixed L , then the estimate of the expected spectrum of the first kind of wave will be less accurate, because

the effective source observation time $T = L/v_g$ will be less. (Of course, for a *time* average with fixed T the variances in the two cases would be equal.) Reduction of variance requires observing the source for longer times, which is why lower values of v_g give lower variance and hence more accuracy.

However, in the present context there appears to be no observational bias against waves with large group velocity, that is, large values of v_g do not bias observational estimates of E_{00} from (31) or (33).

For instance, consider the particular atmospheric example of a time-varying surface wind $U(t)$ over topography and the resultant gravity wave generation. Clearly, all of the earlier comments apply directly to a vertical space average over the topography, with the relevant group-velocity component now being w_g . Now, the wave source statistics are all determined by the statistics of the random process $U(t)$, which one can idealize as an on-off process with an intermittency parameter γ and an integral timescale τ . Depending on the shape of the topography, a more or less broad spectrum of gravity waves will be generated, with a corresponding broad range of w_g values. Then it is clear that space-averaged estimates of E_{00} for fast waves (in w_g) will be less accurate than those for slow waves. However, once the relevant E_{00} has been determined to required accuracy for any particular kind of wave, one can then compute the source spectra E_s (which corresponds to the “on” phase of the surface wind) simply from (24); that is,

$$E_s = \frac{1}{\gamma} E_{00}. \quad (34)$$

This holds for all waves, that is, regardless of w_g . In other words, in the present example w_g affects only the accuracy of space averages used to estimate E_{00} , but it does not affect time averages or the link between the observed mean spectrum E_{00} and E_s .

Notably, this disagrees with the probability of observation scaling between observed and source spectra that has been proposed by ATV. These authors started from the premise that waves with higher group velocity traverse a given observation region of length L faster than waves with slower group velocity. Hence they argue that, for intermittent sources, observable spectra (corresponding to E_{00} here and denoted by E_o in ATV) should be biased toward waves with slower group velocity, as these are more likely to have been observed. Overall, for fixed intermittency parameter γ , ATV propose an additional rescaling $E_s \propto w_g E_{00}$. As they make clear, such a scaling would have important consequences for gravity wave parameterizations and the interpretation of field data.

However, that their proposed scaling cannot be valid for all types of wave sources is borne out by the above counterexample of topographic wave generation. Physically, the reason why ATV’s argument is circumvented here lies in the fact that during an “on” phase of the

surface wind, say of duration t_1 , wave trains of vertical depth $t_1 w_g$ are generated. This means faster waves leave longer trails in this phase. This factor of w_g then neatly cancels the probability of observation effect of shorter transversal time. The upshot is that waves with different w_g are just as likely to be observed in any random observation in this problem.

Finally, one should note the caveat that these considerations (as were ATV's) are essentially one-dimensional in nature. For instance, two-dimensional propagation may lead to other observational biasing effects, such as the effect that nearly horizontal rays (i.e., $m^2 \gg l^2$) may be more likely to be observed by a network of stations or soundings (T. Dunkerton 2002, personal communication). This remains to be investigated.

4. Concluding remarks

In essence, the main point about nonuniform potential energy spectra that is made in this paper is already contained in (10), which quantifies the increasing share of potential energy in an inertia-gravity wave as $f^2/\hat{\omega}^2$ decreases. This shows that all wave sources with fixed $\hat{\omega}$ "produce" more potential energy (per total energy) as $|f|$ decreases. The horizontally propagating waves described in the present paper are just one example of such sources, in which this tendency overwhelms the opposing energy-reducing tendencies due to latitude-circle widening and group-velocity increase. Even if only vertical wave propagation is considered, one might still find more examples of this effect. For instance, convection might be active over a range of latitudes with roughly the same emission spectrum in $\hat{\omega}$. That is to say, roughly the same amount of wave energy is produced by convection at a certain $\hat{\omega}$, regardless at which latitude the convection takes place. Then the latitudinal footprint of these waves would again show a peak in potential energy at the equator, because the same $\hat{\omega}$ corresponds to a higher share of potential energy there than at higher latitudes. Another possible effect was pointed out by an anonymous referee: the poleward propagation of waves generated by equatorial wave sources could also lead to a potential energy maximum near the equator. This is basically the thought experiment of this paper in reverse.

This ubiquity of effects that can contribute to potential energy peaks raises the question of how to distinguish between them in extended observations that go beyond $E_p(\phi)$. In horizontal propagation, for instance, relative peaks in potential energy should be accompanied by relative troughs in kinetic energy, according to (10). In vertical propagation, on the other hand, presumably peaks would occur in both spectra. Also, observations of $E_p(\phi, \hat{\omega})$ would be useful: horizontal propagation of waves generated at higher latitudes would correspond to a thinning out of the spectrum at $\hat{\omega}$ values near the local Coriolis parameter $f(\phi)$. Other wave sources might leave a different footprint in such spectra.

It is again stressed how surprisingly weakly radiative damping acts on low-frequency inertia-gravity waves [cf. (14) above, Bühler and McIntyre 1999]. Otherwise, waves forced say at $\phi_0 = -30^\circ$ would have been obliterated before they reach the equator. This is because their travel time to the equator can be roughly estimated [based on the values in (17)] to be about 15 days, which is long compared to the "bare" radiative damping time scale $1/\hat{\alpha}$ but much shorter compared to the effective damping timescale $1/\alpha$ encountered along the way. This is the key to the longevity of inertia-gravity waves in the lower stratosphere.

It seems that the predictions of the simple model presented here should carry over to more complicated models. For instance, a less simple model would have waves propagating zonally as well as meridionally, that is, the wavenumber vector would be $\mathbf{k} = (k, l, m)$ with nonzero k . However, fundamentally the same behavior would still be observed, as long as $l \neq 0$. For instance, under ray tracing as in section 2 the zonal wavenumber k would be invariant, l^2/k^2 would increase toward the equator, and hence the wave's path would be attracted to the equator, producing a latitudinal ray path much as the ones considered here. Other model improvements would remove the ray-tracing restriction in the meridional direction by developing the wave field in terms of meridional waveguide modes, which are bounded in latitude by the turning points where the wave rays are reflected. This would allow wave amplitudes to be computed everywhere, that is, including the turning point regions. On the downside, this approach would presumably necessitate taking vertical wave propagation into account explicitly.

For time-dependent sources a less simple model of wave packets would take the wavenumber and concomitant v_g spread across the wave packet into account, which leads to wave packet dilation or focusing effects that affect the local wave amplitudes. This again shows how delicate the whole question of predicting local wave amplitudes is: even packet-integrated wave activity might be a poor guide to local amplitudes. In addition, local amplitudes may or may not be significantly affected by wave superposition and nonlinear interaction effects. Expected pseudomomentum fluxes that are summed over all waves, on the other hand, are unaffected by such events, unless they lead to wave breaking. This follows from the well-known nonlinear extensions of wave activity conservation laws. It appears that local wave amplitudes will always be less predictable than mean wave fluxes.

It is quite clear from the discussion in section 3 that intermittent wave sources cannot be properly characterized nor reliably modeled in parameterization schemes based only on a single amplitude parameter. At some extreme, the full probability distribution of the source could be used. On the other hand, in section 3a the simple procedure of approximating an intermittent source with known mean energy and variance by an

equivalent on-off source sharing those values was noted. Estimating statistical source parameters [e.g., the probability density function for a wave source spectrum $E_0(t)$, and its autocorrelation] and their dependence on seasonal conditions, etc., is clearly a worthwhile aim for observational campaigns, whether they use field data or numerical simulations.

Acknowledgments. I wish to thank Tim Dunkerton and an anonymous reviewer for their constructive and insightful comments, which helped improve this paper.

REFERENCES

- Alexander, M. J., and T. J. Dunkerton, 1999: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.*, **56**, 4167–4182.
- , T. Tsuda, and R. A. Vincent, 2002: Latitudinal variations observed in gravity waves with short vertical wavelengths. *J. Atmos. Sci.*, **59**, 1394–1404.
- Alisse, J. R., and C. Sidi, 2000: Experimental probability density functions of small-scale fluctuations in the stably stratified atmosphere. *J. Fluid Mech.*, **402**, 137–162.
- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Bühler, O., and M. E. McIntyre, 1999: On shear-generated gravity waves that reach the mesosphere. Part II: Wave propagation. *J. Atmos. Sci.*, **56**, 3764–3773.
- , —, and J. F. Scinocca, 1999: On shear-generated gravity waves that reach the mesosphere. Part I: Wave generation. *J. Atmos. Sci.*, **56**, 3749–3763.
- Dunkerton, T. J., 1989: Body force circulations in a compressible atmosphere: Key concepts. *Pure Appl. Geophys.*, **130**, 243–262.
- Fels, S. B., 1982: A parameterization of scale-dependent radiative damping rates in the middle atmosphere. *J. Atmos. Sci.*, **39**, 1141–1152.
- , 1984: The radiative damping of short vertical scale waves in the mesosphere. *J. Atmos. Sci.*, **41**, 1755–1764.
- Fritts, D. C., and M. J. Alexander, 2002: A review of gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, in press.
- Holton, J. R., 1983: The influence of gravity wave breaking on the general circulation of the middle atmosphere. *J. Atmos. Sci.*, **40**, 2497–2507.
- Yaglom, A. M., 1962: *An Introduction to the Theory of Stationary Random Functions* (English translation). Prentice Hall, 235 pp.