Toward an ultra-simple spectral gravity wave parameterization for general circulation models

Christopher D. Warner and Michael E. McIntyre

Centre for Atmospheric Science* at the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, U.K.

This paper reports first steps toward a computationally inexpensive spectral gravity wave parameterization scheme whose predictions approximate those of a full three-dimensional (in spectral space) spectral model of atmospheric gravity waves. A reduction to two dimensions, as proposed by Hines, requiring the neglect of Coriolis and non-hydrostatic effects, is explored on the basis of comparisons with a full three-dimensional power-spectral model that includes Coriolis and non-hydrostatic effects. The reduction tries to be more realistic in terms of spectral shapes, though simpler in terms of wave-breaking criteria. It works remarkably well in the absence of, but less well in the presence of, background shear. The reasons for the discrepancies are under investigation, as are the implications for two-dimensional schemes, including Hines' as well as ours.

1. Introduction

Full three-dimensional (3D) power spectral models of gravity wave propagation and breaking are too demanding, in terms of computational expense, for use in parameterizations for general circulation models (GCMs). This is because of the quasi-advective, or fluid-like, behavior of spectral elements in spectral space under Doppler shifting (Warner and McIntyre, 1996). The three dimensions of spectral space can be taken to be the vertical wavenumber m, the intrinsic frequency $\hat{\omega}$, and the azimuthal angle ϕ. Hines (1997a,b) has proposed using a simplified representation in which, in effect, the $\hat{\omega}$ dependence is integrated out. This has some justification if one neglects Coriolis and non-hydrostatic effects, so that the dispersion relation becomes

$$\hat{\omega} = \frac{Nk_0}{m}, \tag{1}$$

where N is the buoyancy frequency and k_0 the magnitude of the horizontal wavenumber $\mathbf{k_0} = k_0(\cos\phi, \sin\phi, 0)$, which is invariant under propagation in a horizontally homogeneous background. However, it is far from obvious how models based on these simplifications—which we call two-dimensional (2D) power-spectral models because they depend on m and ϕ only—will compare with full 3D power-spectral models. Note in particular that (1) cannot describe the back-reflection that takes place, in reality, when $\hat{\omega}$ reaches N, nor correctly represent the important low frequencies $\hat{\omega} \gtrsim f$, where f is the Coriolis parameter.

As part of an effort to develop practical parameterization schemes for GCMs, we present here a first comparison be-

tween 2D and 3D power-spectral models, using idealized spectral shapes designed to further increase the simplicity and computational efficiency of the 2D model. The idealization is in the spirit of Fritts and Lu (1993), but more physically realistic in that it preserves the important distinction between the conservatively-propagated and saturated parts of the spectrum. The 2D-3D comparison turns out to be remarkably good in the case of no background shear, but less so when background shear produces substantial Doppler shifting. We use the July, $40°$N, CIRA 1986 summer winds as the basis for the comparison. In this case, the 2D model, as developed so far, predicts vertical profiles of wave-induced force, turbulent energy dissipation, and pseudomomentum flux[1] that are qualitatively reasonable in most respects but not quantitatively satisfactory. Work is in progress to see how far the discrepancies can be reduced.

2. 2D Power Spectral Modelling
2.1 Launch spectrum

Following Warner and McIntyre (1996, 1997), hereafter WM96 and WM97, we consider the waves to be launched upward from an altitude $z = z_L = 19.2$ km with the same spectrum in each direction ϕ_j. For the 2D power-spectral model, we choose the launch spectrum shown by the solid curve in Fig. 1. The dotted curve shows the $\hat{\omega}$-integral of the pseudomomentum-flux launch spectrum in the full 3D power-spectral model. This last corresponds to the energy

---

[1]Momentum and pseudomomentum are distinct but related quantites that appear in wave-mean interaction theory. Conservation of momentum is related to translational invariance of the physics, whereas conservation of pseudomomentum is related to translational invariance of the medium in which the waves propagate, as represented in linearized wave theory. The relevant components of their respective fluxes, i.e., the vertical fluxes of horizontal momentum and pseudomomentum, are equal, to sufficient approximation, in our problem. However, the pseudomomentum flux can be equated to the vertical group velocity times the horizontal pseudomomentum density, which latter can be evaluated from linearized wave theory. For further discussion see WM96, above Eq. (13).
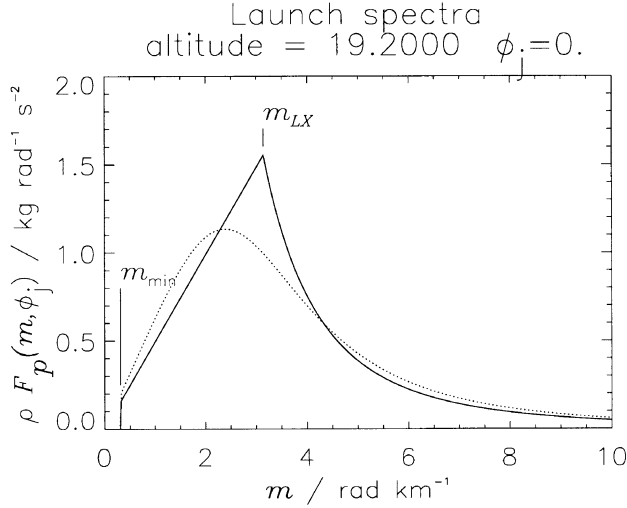
Fig. 1.   Launch spectra of the vertical flux of horizontal pseudomomentum, integrated over the jth azimuthal sector $\phi_j \pm \frac{1}{2}\Delta\phi$. The sector size $\Delta\phi = 90°$ and j runs from 0 to 3. The solid curve is (4) and the dashed curve is (3). The total (one dimensional) pseudomomentum flux $\rho(z_L)\mathbf{F}_\mathbf{p}^{(1)}(z_L, \phi_j)$ is identical for the two cases and is given by the (equal) areas under the curves. The characteristic vertical wavenumber $m_* = 2\pi/(2\,\text{km}) = 3.142 \times 10^{-3}$ rad m^{-1}, the crossover wavenumber $m_{LX} = m_*$, and the small-m cutoff $m_{\min} = 2\pi/(20\,\text{km}) = 3.142 \times 10^{-4}$ rad m^{-1}. Units for m are rad km^{-1} and for the 2D pseudomomentum flux $\rho(z_L)\mathbf{F}_\mathbf{p}^{(2)}(z_L, m, \phi_j)$ are kg (rad km^{-1})$^{-1}$km^{-1}s^{-2}.

spectrum of Fritts and VanZandt (1993) apart from the introduction of a long wave cutoff at vertical wavenumber $m = m_{\min}$.

More precisely, the dotted curve is the $\hat{\omega}$-integral of

$$\rho(z_L)\mathbf{F}_\mathbf{pL}^{(3)}(z_L, m, \hat{\omega}, \phi_j)$$

$$= \begin{cases} 0 & \text{for} \quad m < m_{\min}, \\ \rho(z_L)\beta A_0 B_0 \Phi_0 m(m_*^4 + m^4)^{-1} N_L \hat{\omega}^{1-p} & \\ \quad \times(1 - \hat{\omega}^2/N_L^2)^{1/2}(1 - f^2/\hat{\omega}^2)^{3/2} & \\ \quad \times(1 - f^2/N_L^2)^{-1} & \\ \quad \times 2\sin(\Delta\phi/2)\hat{\mathbf{k}_0} & \text{for} \quad m_{\min} \leq m < \infty, \end{cases} \quad (2)$$

which is the pseudomomentum-flux launch spectrum in the full 3D power-spectral model. Here $\rho(z_L)$ is the density at the launch altitude z_L, β is an empirical constant, taken throughout this paper as 1.047×10^{-1}, while A_0, B_0, and $\Phi_0 (= (2\pi)^{-1})$ are normalization constants for the vertical wavenumber spectrum, the intrinsic frequency spectrum and the azimuthal direction spectrum respectively and are defined in WM96, m_* is the "characteristic wavenumber" of the Fritts-VanZandt spectrum, N_L is the value of the buoyancy frequency at the launch altitude $z = z_L$, and the factors in the second and third lines of the equation come from taking Coriolis and non-hydrostatic effects into account. The $2\sin(\Delta\phi/2)$ factor is the result of integrating vectorally over the azimuthal sector $\phi_j - \Delta\phi/2 < \phi < \phi_j + \Delta\phi/2$ centered on direction ϕ_j and of angular width $\Delta\phi$ (see WM96). The notation $\hat{\mathbf{k}_0}$ means the unit vector $\mathbf{k_0}/|\mathbf{k_0}|$. The subscript L applied to any symbol indicates the value at the launch altitude. The notational conventions used for the symbols for

fluxes with subscripts and superscripts are as follows. Taking the above example, $\mathbf{F}_\mathbf{pL}^{(3)}$, the bold font shows that this is a vector; the superscript (3) shows that the spectral space is 3D, subscript $\mathbf{p}$ shows that this is a flux of pseudomomentum (subscript E would indicate a flux of wave energy); subscript L indicates, as before, that this is a launch altitude flux.

When integrated over $\hat{\omega}$, (2) yields the equation of the dotted curve,

$$\rho(z_L)\mathbf{F}_\mathbf{pL}^{(2)}(z_L, m, \phi_j)$$

$$= \begin{cases} 0 & \text{for} \quad m < m_{\min}, \\ \rho(z_L)C_{L3D}m_*^2 m(m_*^4 + m^4)^{-1} & \\ \quad \times 2\sin(\Delta\phi/2)\hat{\mathbf{k}_0} & \text{for} \quad m_{\min} \leq m < \infty, \end{cases} \quad (3)$$

where $C_{L3D} = 0.0384$ m^{2}s^{-2}, $m_* = 2\pi/(2$ km), $m_{\min} = 2\pi/(20$ km), and $N_L^2 = 5.01 \times 10^{-4}$ rad s^{-2}, which latter is the CIRA summer value at 19.2 km altitude.

To allow efficient computation, the pseudomomentum-flux launch spectrum in the 2D power-spectral model, the solid curve in Fig. 1, is defined by two power laws in different ranges of m separated by a crossover wavenumber $m(z_L) = m_{LX}$, here taken somewhat arbitrarily to be equal to m_*:

$$\rho(z_L)\mathbf{F}_\mathbf{pL}^{(2)}(z_L, m, \phi_j)$$

$$= \begin{cases} 0 & \text{for} \quad m < m_{\min}, \\ \rho(z_L)C_L m_{LX}^{-1-s} m^s & \\ \quad \times 2\sin(\Delta\phi/2)\hat{\mathbf{k}_0} & \text{for} \quad m_{\min} \leq m < m_{LX}, \quad (4) \\ \rho(z_L)C_L m_{LX}^2 m^{-3} & \\ \quad \times 2\sin(\Delta\phi/2)\hat{\mathbf{k}_0} & \text{for} \quad m_{LX} \leq m < \infty. \end{cases}$$

Here $C_L = 0.0329$ m^{2}s^{-1}; it is chosen to make the areas under the curves in Fig. 1 the same. Note that $C_L \neq C_{L3D}$ because of the different spectral shapes.

In this paper, the power s for $m_{\min} \leq m < m_{LX}$ will be taken to be $s = 1$ as suggested in Fritts and VanZandt (1993). In WM97, we found that this value $s = 1$ gives too large a contribution from ultra-long vertical wavelengths. In WM97 we tried adjusting s (which, because it refers to the small-m end of the spectrum, is hardly constrained by gravity wave observations). However, gravity waves with vertical wavelengths greater than 20 km are rarely if ever observed in the stratosphere. Therefore, somewhat arbitrarily, we keep $s = 1$ as in WM96 but introduce a long-wave (small-m) cutoff at $m = m_{\min}$. A reviewer has suggested that, when the model is used as a GCM parameterisation, the small-m cutoff should be set to the maximum wavenumber of vertically resolvable waves in the GCM. (This would presume that the model by itself correctly generates such waves.) Without the small-m cutoff, the model predicts turbulent energy dissipation rates ε that appear to be unrealistically high, in comparison for instance with the rates observed by Lübken (1997).

As is well known, gravity wave energy spectra derived from observations often have an m^{-3} behavior at large vertical wavenumber m. This can be rationalized in terms of saturation by wave breaking (Dewan and Good, 1986; Smith *et al.*, 1987), though the observed spectra may in the end

require a more sophisticated explanation. For further discussion see WM96, also Hines (1991), Broutman *et al.* (1997), and Eckermann (1997). The right hand part of the solid curve in Fig. 1, i.e., (4) in the range $m_{LX} \leq m < \infty$, has just this m^{-3} behavior, and is what we mean by the "quasi-saturated part" of the model spectrum.

2.2 Propagation

In the full 3D spectral description, each spectral component has a wavevector $\mathbf{k} = (k, l, -m)$, with zonal component k, meridional component l and vertical component $-m$, to make $m > 0$ for upward propagating waves with $\hat{\omega} > 0$ (i.e., sign conventions as in WM96). The horizontal projection of $\mathbf{k}$, or the horizontal wavevector, is $\mathbf{k_0} = (k, l, 0) = k_0(\cos\phi_j, \sin\phi_j, 0)$, with magnitude $k_0 = (k^2 + l^2)^{1/2} > 0$. Each such plane wave spectral component has an absolute frequency, or frequency relative to the ground, denoted by ω_0, which is invariant under propagation in a steady background.

The intrinsic and absolute frequencies satisfy the Doppler-shift relation

$$\hat{\omega}(z) = \omega_0 - \mathbf{k_0} \cdot \mathbf{U}(z) = \omega_0 - k_0 U(z), \qquad (5)$$

where $U(z)$ is the component of $\mathbf{U}$ in the direction of $\mathbf{k_0}$, equivalently in the ϕ_j direction. Because by definition $k_0 > 0$ and, for upward propagating waves, $\hat{\omega} > 0$, a positive vertical shear $\partial U/\partial z > 0$ means that $\hat{\omega}$ will decrease in magnitude as altitude z increases, corresponding to an approach toward critical-level conditions, promoting wavebreaking.

The vertical wavenumber m also varies with z under Doppler shifting; we use the abbreviations $\hat{\omega}_1 = \hat{\omega}(z_1)$, $m_1 = m(z_1)$ and $\hat{\omega}_2 = \hat{\omega}(z_2)$, $m_2 = m(z_2)$ at altitude z_1, z_2, and similarly for the other z-dependent quantities like U and N. Substituting values $U = U_1$ and U_2 into (5), and subtracting to eliminate ω_0, we get

$$\hat{\omega}_1 - \hat{\omega}_2 = k_0 U_2 - k_0 U_1 \qquad (6)$$

for the Doppler shift in intrinsic frequencies between z_1 and z_2. Substituting for $\hat{\omega}_1$, $\hat{\omega}_2$ using the dispersion relation (1) and dividing by k_0, we obtain

$$\frac{N_1}{m_1} - \frac{N_2}{m_2} = U_2 - U_1. \qquad (7)$$

Under the hydrostatic, non-rotating approximation, therefore, the Doppler shifting of the vertical wavenumber m does not depend on the magnitude of the horizontal wavenumber k_0.

This is crucial to the simplification used here and in Hines (1997a,b). It can be shown, and will be explained more fully in a forthcoming paper, that under such simplifications we can treat spectral propagation as a 2D problem in the same way as we treated the full 3D problem in WM96. The essence of the matter is that, under the approximation (1), the flow in $(m, \hat{\omega})$ spectral space is such that elements do not tilt and stretch under Doppler shifting, as they do in the full evolution illustrated in Fig. 2 of WM96.

As in WM96, we can derive a spectral Jacobian to transform (2D) spectral elements in (m, ϕ_j) spectral space to the propagation-invariant spectral elements in (ω_0, ϕ_j) spectral space. We can then use that Jacobian to calculate how the 2D

pseudomomentum flux $\rho(z)\mathbf{F}_{\mathbf{p}}^{(2)}(z, m, \phi_j)$ varies from one altitude z to another under conservative propagation. We find that

$$\rho(z_2)\mathbf{F}_{\mathbf{p}}^{(2)}(z_2, m_2, \phi_j)$$
$$= \rho(z_1)\mathbf{F}_{\mathbf{p}}^{(2)}(z_1, m_1, \phi_j)\frac{dm_1}{dm_2}$$
$$= \rho(z_1)\mathbf{F}_{\mathbf{p}}^{(2)}(z_1, m_1, \phi_j)\frac{N_2}{N_1}\left(\frac{m_1}{m_2}\right)^2. \qquad (8)$$

Having conservatively propagated the 2D spectrum from altitude z_1 up to altitude z_2, we then impose as a ceiling the m^{-3} quasi-saturated behavior, that is, we require for all m_2 that

$$\rho(z_2)\mathbf{F}_{\mathbf{p}}^{(2)}(z_2, m_2, \phi_j) \leq \rho(z_2)\mathbf{F}_{\mathbf{pS}}^{(2)}(m, \phi_j), \qquad (9)$$

where

$$\rho(z)\mathbf{F}_{\mathbf{pS}}^{(2)}(m, \phi_j) = \rho(z)C_L\frac{m_{LX}^2}{m^3} \times 2\sin(\Delta\phi/2)\hat{\mathbf{k}}_0. \qquad (10)$$

In other words, just as for the 3D case in WM96, we use the quasi-saturation spectrum (10) as a "chopping function" for $\rho(z)\mathbf{F}_{\mathbf{p}}^{(2)}(z, m, \phi_j)$ at altitude z in the 2D model. We give the name "evolved spectrum" to the final 2D spectrum after propagation and chopping. The constant C_L is the same as in (4), ensuring that the chopping function coincides with the right-hand part of the model launch spectrum—that is, coincides with (4) for $m \geq m_{LX}$, i.e., to the right of the crossover wavenumber m_{LX}. Although it can be argued that the constant C_L in (10) should be taken to be a function of altitude, via its dependence on $N(z)$, we choose for simplicity—as was done in WM96—to keep C_L constant, so that the altitude variation of $\rho(z)\mathbf{F}_{\mathbf{pS}}$ is due only to the altitude variation, $\rho(z)$, of the density for both the 2D and full 3D power-spectral models. In future work we plan to make C_L a function of altitude as a further step toward realism.

2.3 Computational procedure

The key to efficient computation is to assume that the evolving spectrum retains a piecewise structure as in (4), with smooth shapes except at crossover and cutoff wavenumbers $m_X(z)$ and $m_{min}(z)$. We may then take advantage of the power-law behavior of, respectively, the launch spectrum to the left of, and the chopping function to the right of, the crossover wavenumber.

The actual computational procedure is now described for the case of positive shear, with reference to Fig. 2. Panels (a) and (b) respectively show the simplified launch spectrum (again for $z = z_1 = z_L = 19.2$ km) together with the result of conservatively propagating it up to $z = z_2 = 20.7$ km through a positive wind increment $U_2 - U_1 = +5$ m s^{-1}. Consistent with (7), the image of $m_{LX} = m_X(z_L)$ under Doppler shifting has gone off-scale to the right of panel (b), at $m = 11.2$ rad km^{-1}, whereas the image of m_{min} has hardly shifted, by only $7\frac{1}{2}\%$ in this example, illustrating the extreme sensitivity of Doppler shifting to m values.

Panel (c) shows the chopping function (10) at $z = z_2 = 20.7$ km. The positive shear and density decrease have lowered it. Panel (d) shows the evolved spectrum, as defined above, at 20.7 km. This is the result of using (10), i.e., the
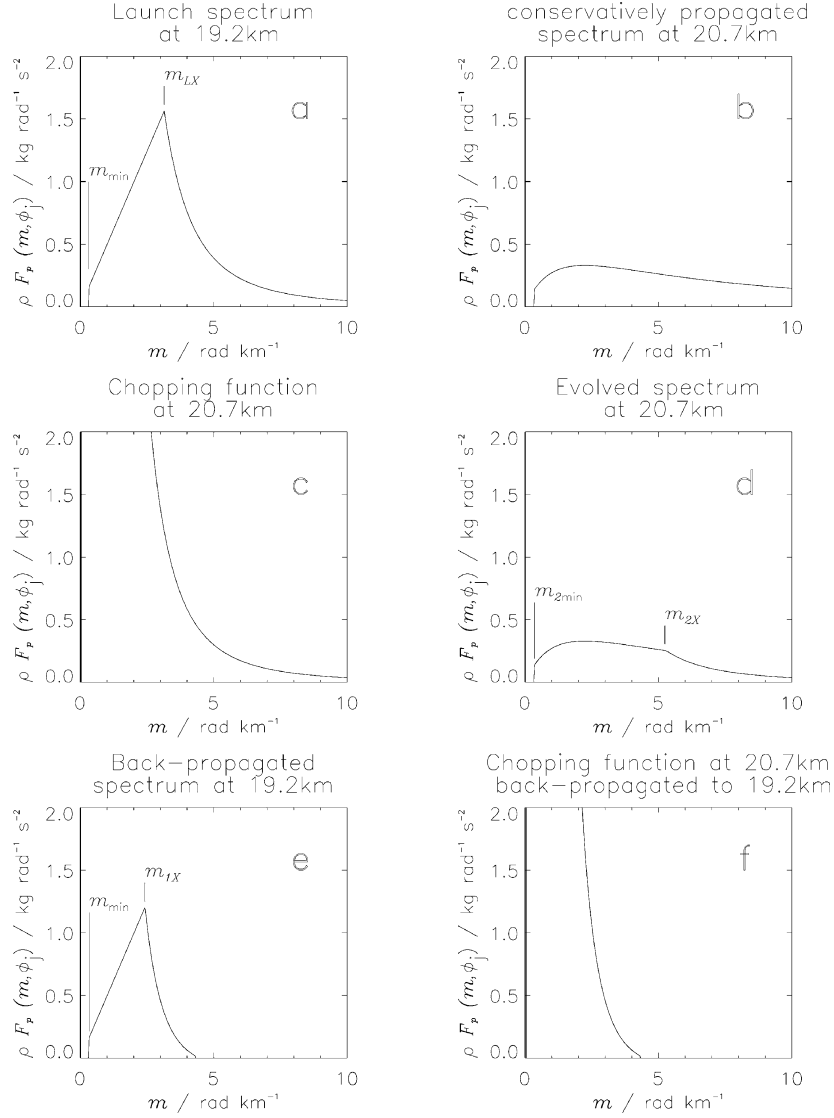
Fig. 2. Computational procedure for the ultra-simple spectral parameterisation for a positive background wind shear with increment $U_2 - U_1 = +5\,\mathrm{m\,s^{-1}}$; see text.

curve in panel (c), to chop the conservatively propagated spectrum shown in panel (b).

Panel (e) shows what will be called the artificially "back-propagated spectrum" at 19.2 km or, for brevity, "back-propagated spectrum". By this we mean a spectrum that would result from running time backwards and conservatively propagating the evolved spectrum back down from 20.7 km to 19.2 km altitude. In other words, the back-propagated spectrum at $z = z_L$ (panel (e)) is just such as will propagate conservatively to produce the spectrum in panel (d) when time runs forwards. Note that the new crossover wavenumber m_{1X} in panel (e), which is the image of that in panel (d), m_{2X}, under time-reversed Doppler shifting, lies to the left of the original crossover wavenumber m_{LX}. That is, $m_{1X} < m_{LX}$. Finally, panel (f) shows the chopping function at 20.7 km (panel (c)) back-propagated to 19.2 km.

The aim is to compute the area under the curve in Fig. 2, panel (d), and other curves like it at other altitudes z. Note first that this area is the same as the area in panel (e) because

back-propagation is by definition conservative. Second, the areas of the right and left hand parts—i.e., the parts to the right and left of the new crossover wavenumber—are separately equal, because the relevant values m_{1X} and m_{2X} of the crossover wavenumber are connected via the Doppler relation (7). Third, the right hand area is easiest to compute analytically at $z = z_2$ because the chopping function (10) is a simple power law. By contrast, the left hand area is easiest to compute at $z = z_1 (= z_L)$, i.e., using panel (e) rather than panel (d), because it is the left half of panel (e) that has the simple power law behavior. The result is

$$\rho(z_2)\mathbf{F}_{\mathbf{p}}^{(1)}(z_2, \phi_j) = \int_{m_{\min}}^{m_{1X}} \rho(z_L)\mathbf{F}_{\mathbf{pL}}^{(2)}(z_L, m, \phi_j)\,dm$$
$$+ \int_{m_{2X}}^{\infty} \rho(z_2)\mathbf{F}_{\mathbf{pS}}^{(2)}(z_2, m, \phi_j)\,dm$$

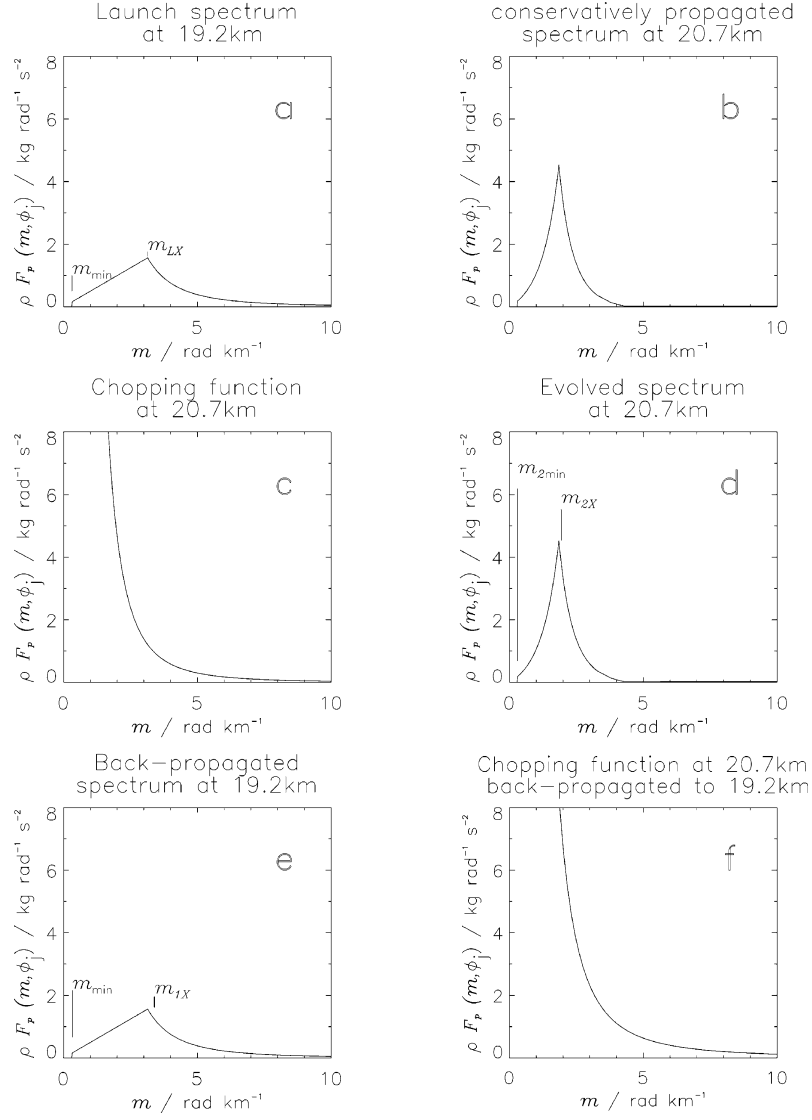Fig. 3. Computational procedure for the ultra-simple spectral parameterisation for a negative background wind shear with increment $U_2 - U_1 = -5\,\mathrm{m\,s^{-1}}$; see text.

$$= \left[\frac{\rho(z_\mathrm{L})C_\mathrm{L}}{m_\mathrm{LX}^{s+1}(s+1)} \left(m_\mathrm{1X}^{s+1} - m_\mathrm{min}^{s+1} \right) \right.$$
$$\left. + \frac{\rho(z_2)C_\mathrm{L}m_\mathrm{LX}^2}{2} \left(m_\mathrm{2X}^{-2} \right) \right]$$
$$\times 2\sin(\Delta\phi/2)\hat{\mathbf{k}}_0, \qquad (11)$$

where (cf. above (3)) the superscript (1) signals that all but one dimension of spectral space has been integrated out.

The crossover wavenumber m_1X is determined efficiently by using a Newton-Raphson method to locate the zero-crossing point of the back-propagated chopping function minus the expression in the second line of (4). Then m_2X is obtained from (7).

The conditions typified by Fig. 2 assume that $m_\mathrm{min} < m_\mathrm{1X}$. If $m_\mathrm{min} \geq m_\mathrm{1X}$ then the required modification is

$$\rho(z_2)\mathbf{F}_\mathbf{p}^{(1)}(z_2, \phi_j) = \int_{m_{2\min}}^{\infty} \rho(z_2)\mathbf{F}_{\mathbf{p}S}^{(2)}(m, \phi_j)dm$$

$$= \left[\frac{\rho(z_2)C_\mathrm{L}m_\mathrm{LX}^2}{2} \left(m_{2\min}^{-2} \right) \right]$$
$$\times 2\sin(\Delta\phi/2)\hat{\mathbf{k}}_0, \qquad (12)$$

where $m_{2\min} = m_\mathrm{min}(z_2)$, defined by the appropriate Doppler shift, i.e., by substituting $m_1 = m_\mathrm{min}$ into (7).

Now consider the case of negative shear, with reference to Fig. 3. This is different because chopping need not occur at all, and when it does occur will generally give a less simple result. Figure 3 follows the same pattern as Fig. 2 and illustrates the case where there is no chopping: evolution is purely conservative. The crossover wavenumbers m_1X and m_2X are still defined, and computed by the same Newton-Raphson procedure, but, because negative shear raises the chopping function instead of lowering it, the Newton-Raphson procedure now yields a value of m_1X that falls to the right of m_LX, instead of to the left of m_LX, whereas, in reality, it should be identical to m_LX because the propagation is conservative. If chopping takes place, as will happen at sufficiently high al-

titudes, then the result will be a more complicated piecewise smooth curve, violating the assumptions used in deriving (11) and (12). Because of this, computing $\rho(z_2)\mathbf{F}_\mathbf{p}^{(1)}(z_2, \phi_j)$ from (11) or (12) when the shear is negative can yield a result that increases with altitude. This would correspond to waves that had previously broken being resurrected, which is clearly unphysical. We therefore insist, for all altitudes z, that $\rho(z)\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)$ stay constant or decrease monotonically as z increases. Therefore, we accept the value of $\rho(z)\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)$ computed using (11) and (12) (with z replacing z_2) only if it is less than the values obtained for smaller z. For the case shown in Fig. 3 this amounts to substituting $m_{1X} = m_{LX}$ together with the corresponding m_{2X}, obtained from (7), into (11). This is merely the simplest way of coping with the negative-shear case. We shall see that it may need some refinement; see Section 4.

2.4   Wave-induced forces per unit mass

As in WM96, we integrate-out the remaining spectral dimension, the direction ϕ in spectral space, to give, in discretized form,

$$\rho(z)\mathbf{F}_\mathbf{p}^{(0)}(z) = \rho(z) \sum_{j=0}^{n-1} \mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)$$

$$= \rho(z) \sum_{j=0}^{n-1} |\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)| \hat{\mathbf{k}}_\mathbf{0}. \quad (13)$$

Here n is the number of discrete azimuthal sectors. The corresponding wave-induced force per unit mass is ρ^{-1} times the one-dimensional divergence of $\rho(z)\mathbf{F}_\mathbf{p}^{(0)}(z)$, namely

$$\mathbf{G}(z) = -\rho(z)^{-1} \frac{\partial}{\partial z}[\rho(z)\mathbf{F}_\mathbf{p}^{(0)}(z)]. \quad (14)$$

The zonal and meridional components of $\mathbf{G}(z)$ can now be calculated. They are

$$\mathbf{G}(z) = (G_x(z), G_y(z)) = \sum_{j=0}^{n-1} |\mathbf{G}(z, \phi_j)|(\cos\phi_j, \sin\phi_j),$$

$$\quad (15)$$

where the contributions $G(z, \phi_j)$ to $G(z)$ from each sector are readily calculated by substituting $\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)$ for each sector into (14) in place of $\mathbf{F}_\mathbf{p}^{(0)}(z)$.

2.5   Turbulent dissipation

Observed turbulent dissipation rates due to breaking mesospheric gravity waves (e.g., Lübken 1997) are an important observational constraint on theoretical models. Turbulent dissipation can be estimated to a first approximation as wave-energy flux convergences, i.e., neglecting work done by the turbulence against gravity (e.g., McIntyre, 1989). Computing wave-energy fluxes for the ultra-simple spectral parameterisation is carried out as follows. First the 2D wave-energy flux $\rho(z)F_E^{(2)}(z, m, \phi_j)$ is written in terms of the 2D pseudo-momentum flux $\rho(z)\mathbf{F}_\mathbf{p}^{(2)}(z, m, \phi_j)$:

$$\rho(z)F_E^{(2)}(z, m, \phi_j) = \rho(z)|\mathbf{F}_\mathbf{p}^{(2)}(z, m, \phi_j)| \frac{\hat{\omega}}{k_0}$$

$$= \rho(z)|\mathbf{F}_\mathbf{p}^{(2)}(z, m, \phi_j)| \frac{N}{m}, \quad (16)$$

where the second line is obtained using the approximate dispersion relation (1). Then the 1D wave-energy flux

$\rho(z)F_E^{(1)}(z, \phi_j)$ for the jth sector is computed by integrating over all m. For positive shear, the procedure is similar to that needed to obtain (11). Note that carrying out these integrations is not just a matter of scaling $|\rho(z)\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)|$, because the m-dependences of $\rho(z)F_E^{(2)}(z, m, \phi_j)$ and $\rho(z)\mathbf{F}_\mathbf{p}^{(2)}(z, m, \phi_j)$ are different. However, under the hydrostatic, non-rotating approximation the integrals are straightforward to evaluate. Details are omitted for brevity.

For negative shear, we again use the simplest way of coping. The relations (16) and (7), together with the condition $\rho(z)\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j) = $ constant, suggest taking

$$\rho(z)F_E^{(1)}(z, \phi_j) = \rho(z_L)F_E^{(1)}(z_L, \phi_j) \frac{m_{LX}}{N_L} \frac{N(z)}{m_{\text{peak}}(z)}, \quad (17)$$

assuming that $\rho(z)|\mathbf{F}_\mathbf{p}^{(2)}(z, m, \phi_j)|$ has a single peak at $m = m_{\text{peak}}(z)$. We then roughly estimate $m_{\text{peak}}(z)$ by identifying it with $m_{2X}(z)$ as discussed in connection with Fig. 3. This is not necessarily a good approximation—it probably underestimates the energy flux—but we defer attempts to improve it until the whole treatment of the negative-shear case is refined.

On the above basis, the energy dissipation rate is given from the one-dimensional divergence of the wave-energy flux

$$\rho(z)F_E^{(0)}(z) = \rho(z) \sum_{j=0}^{n-1} F_E(z, \phi_j), \quad (18)$$

together with a wave-mean interaction term giving a wave-induced energy dissipation rate per unit mass

$$\varepsilon(z) = -\frac{1}{\rho(z)} \frac{\partial}{\partial z}[\rho(z)F_E^{(0)}(z)] - \frac{\partial \mathbf{U}}{\partial z} \cdot \mathbf{F}_\mathbf{p}^{(0)}(z). \quad (19)$$

Here $\partial \mathbf{U}/\partial z$ is the wind shear vector. Contributions $\varepsilon(z, \phi_j)$ to $\varepsilon(z)$ from each sector are readily calculated by substituting $\mathbf{F}_\mathbf{p}^{(1)}(z, \phi_j)$ into (19) in the place of $\mathbf{F}_\mathbf{p}^{(0)}(z)$ and $F_E^{(1)}(z, \phi_j)$ in the place of $F_E^{(0)}(z)$.

3.   Comparative Tests for the July, 40°N, CIRA 1986 Summer Case

The July, 40°N, CIRA 1986 model atmosphere is shown in Fig. 4. We take $n = 4$ in (13)–(15) corresponding to four azimuths, eastward, northward, westward, and southward ($\phi_j = 0°, 90°, 180°,$ and $270°$), and to $\Delta\phi = 90°$. Spectra were launched from the zero of background zonal wind at an altitude of 19.2 km. The azimuths $\phi_j = 90°$ and $270°$ have no wind shear and so no Doppler shifts ($\mathbf{k_0}$ normal to $\mathbf{U}$). The results for $\phi_j = 90°$ and $270°$ will therefore be identical, so only the $\phi_j = 90°$ case will be shown. Two curves are shown in each panel of Figs. 5, 6, and 7. The solid curves are for the ultra-simple spectral parameterisation, and the dashed curves are for the full 3D power-spectral model. In each figure, panel (a) shows the $\phi_j = 180°$ case, panel (b) the $\phi_j = 0°$ case, and panel (c) the $90 \equiv 270°$ case. Panel (d), where present, shows information obtained by integrating over all azimuthal sectors.

Figure 5 shows the total pseudomomentum flux $\rho\mathbf{F}_\mathbf{p}^{(1)}$ for each azimuthal sector. Consider first the northward-propagating, no wind shear, case shown in panel (c). Waves break at all altitudes, as shown by the monotonic decrease in pseudomomentum flux. The two curves are seen almost to overlay
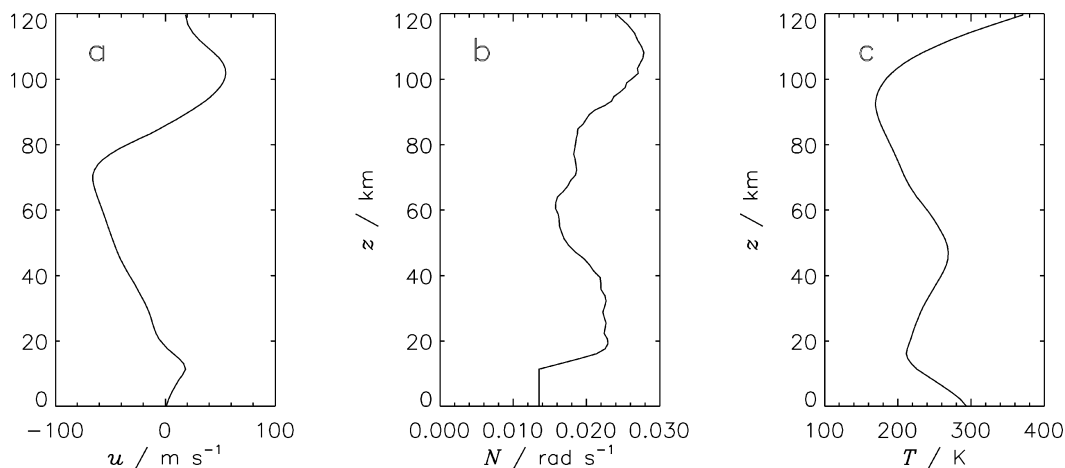
Fig. 4. July, 40°N, CIRA 1986 summer profiles used in the comparison between the ultra-simple spectral parameterisation and the full 3D power-spectral model: (a) zonal mean zonal wind, (b) zonal mean buoyancy frequency $N(z)$, and (c) zonal mean temperature $T(z)$.
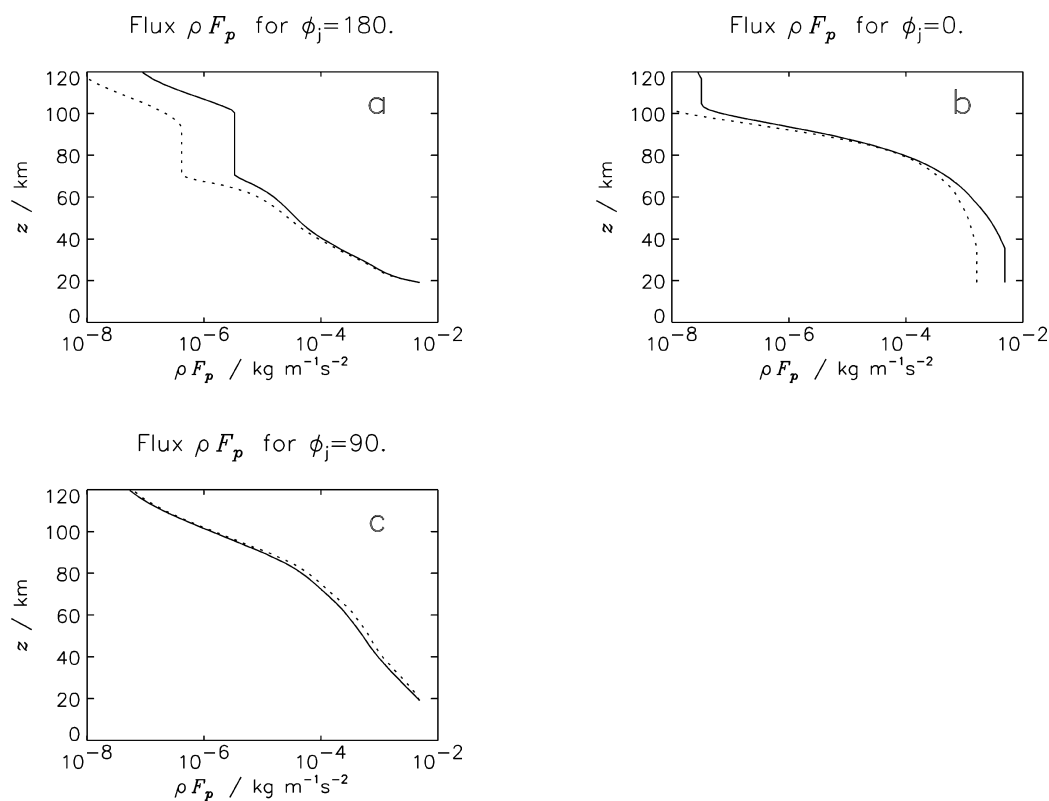

Fig. 5. Comparison of pseudomomentum fluxes for the ultra-simple model (solid curves) and the full model (dotted curves); see text.

one another. Here the ultra-simple model does remarkably well.

Next, consider the westward-propagating case shown in panel (a). Here waves are initially in positive background wind shear, and the two curves are seen to have a similar altitude dependence, although here the discrepancy between the models builds up to an order of magnitude difference in pseudomomentum flux at 70 km altitude. Between altitudes of 75 km and 95 km, where the background wind shear for these waves is strongly negative, there is hardly any wave

breaking, for either model, and hardly any change in total pseudomomentum flux. Above 95 km, wave breaking starts to occur for the full model. However, because of the simplifications that we made in going from the full model to the ultra-simple model, wave breaking for the ultra-simple model only starts to occur above 100 km. The reason is that the pseudomomentum flux is overestimated by the ultra-simple model when the wind shear is negative, as is the case for these westward-propagating waves between about 70 km and 100 km. Above 100 km, the wind shear changes sign,
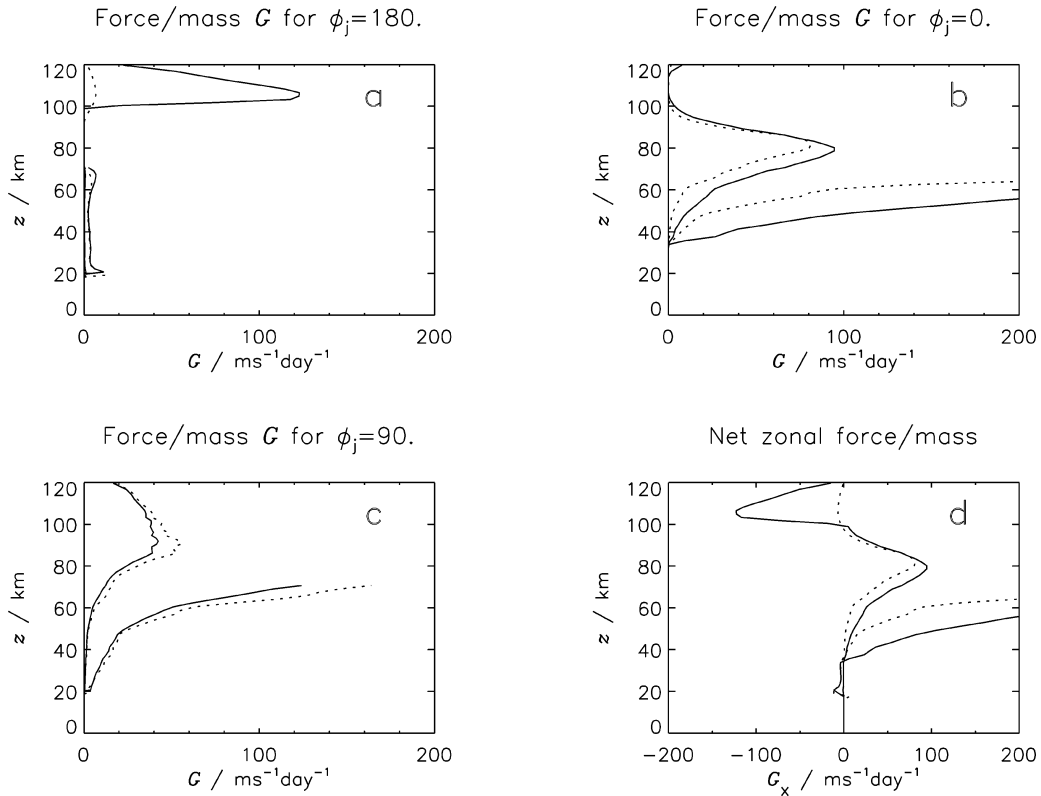
Fig. 6. Comparison of wave-induced forces G per unit mass for the ultra-simple model (solid curves) and the full model (dotted curves); see text. The forces below about 70 km are plotted twice, first with the scale shown and second with a scale running from 0 to 20 m s^{-1} in place of 0 to 200 m s^{-1}.

causing the ultra-simple model to deposit some of the excess pseudomomentum flux. Both cases again show significant wave breaking but an order of magnitude more flux is deposited in the ultra-simple model than in the full model.

Finally, consider the eastward-propagating case shown in panel (b). It is here that we come up against the limitation of the ultra-simple model that derives from its inability to represent the back-reflection of waves Doppler shifted up to $\hat{\omega} = N$. The full model, taking as it does the back-reflection into account, has cancelled about 70% of the upward pseudomomentum flux of the launch spectrum (dotted curve). The ultra-simple model has no means of doing this, as presently formulated, and so launches the full upward flux (solid curve). Higher up, there is a remarkable cancellation of errors, in which the ultra-simple model breaks the waves too strongly, bringing the flux back toward agreement with the full model, but at the cost of producing a spurious wave-induced force especially in the important altitude range of the upper stratosphere (see concluding remarks).

Over the altitude range 19.2 km to 30 km, there is hardly any change in total pseudomomentum flux for both the ultra-simple model and the full model, the background wind shear being sufficiently negative to avoid wave breaking. Above 90 km, the two curves diverge again and, at the highest altitudes, there is more than an order of magnitude difference in pseudomomentum flux between the two cases.

Figure 6 shows the wave-induced force per unit mass G for each azimuthal sector, together with the zonal component of wave induced force G_x. As the forces below about 70 km

can be very small they are plotted twice, the second set of curves scaled by a factor of 10. Consider first the northward-propagating, no wind shear, case shown in panel (c). The force G from the ultra-simple model is somewhat smaller than that from the full model over almost the entire altitude range, most obviously at the peak value around 90 km altitude where it is about 50% smaller. The falloff in G with altitude above about 90 km is a consequence of choosing $m_{min} = 2\pi/20$ km. G starts to fall off from a higher altitude if a smaller value of m_{min} is chosen, and starts to fall off from a lower altitude if a larger value of m_{min} is chosen. In choosing an appropriate m_{min}, we are guided by observations such as those of Lübken (1997). More details are given below.

Panel (a) shows the westward-propagating case. The G values from these waves are seen to be small except at the highest altitudes. Between 100 km and 120 km altitude, the G in the ultra-simple model exceeds 120 m s^{-1}day^{-1}, while G for the full model is over an order of magnitude smaller. This large difference in G is a natural consequence of the large difference in the pseudomomentum fluxes of the spectra that we saw in Fig. 5. As already explained above, the simplifications that we made in going from the full model to the ultra-simple model result in the pseudomomentum flux being overestimated when the wind shear is negative. When the wind shear changes sign, some of this excess pseudomomentum flux is deposited and contributes to the large G values between 100 km and 120 km. Panel (b) shows the eastward-propagating case. The unrealistically high G values from the ultra-simple model in the upper stratosphere are
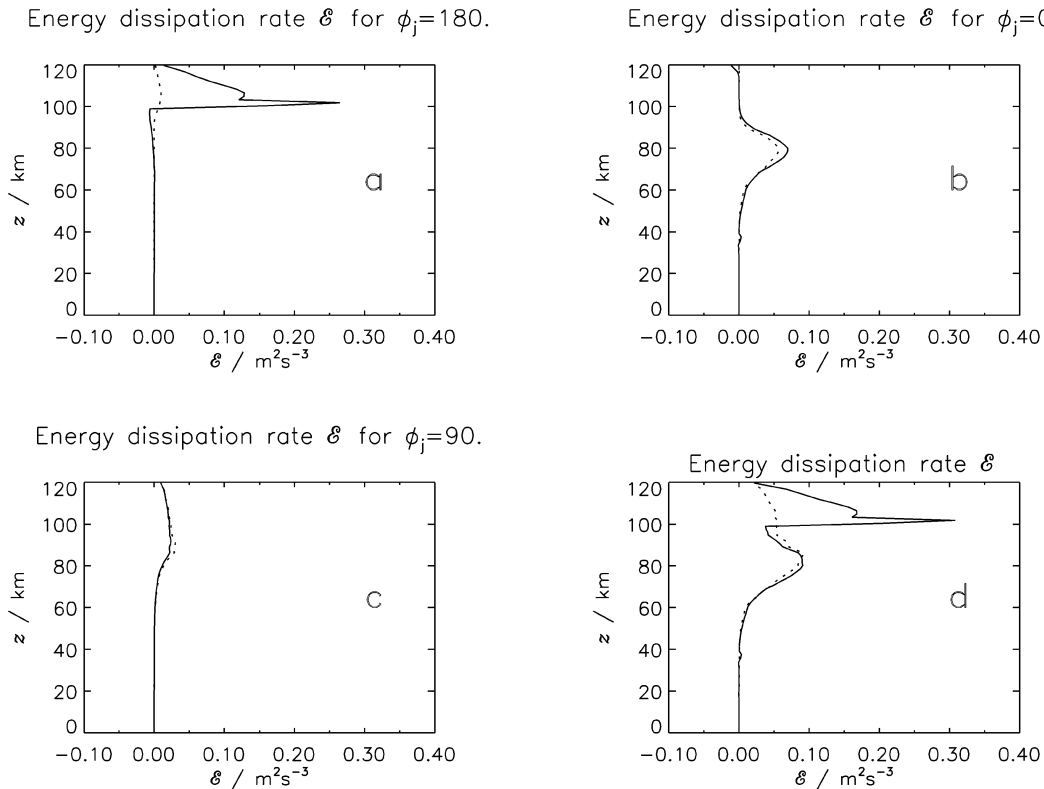
Fig. 7. Comparison of energy dissipation rates for the ultra-simple model (solid curves) and the full model (dotted curves); see text.

conspicuous.

The total wave-induced force **G**, summed over azimuths ϕ_j (which for this case is just the zonal component), is shown in panel (d). The peak values of G are 94 m s^{-1}day^{-1} eastward at an altitude of 80 km for the ultra-simple model, (as compared with 81 m s^{-1}day^{-1} eastward at an altitude of 82 km for the full model), and 123 m s^{-1}day^{-1} westward at an altitude of 107 km for the ultra-simple model, (as compared with 7 m s^{-1}day^{-1} westward at an altitude of 107 km for the full model).

Recalling the original results of WM96, we note particularly the very significant effect of the choice of m_{min}. The full model, with $m_{min} = 2\pi/20$ km, now gives much smaller forces at the highest altitudes than it did in the cases studied in WM96, with $m_{min} = 0$ and $s = 1$, which gave a peak G value of 350 m s^{-1}day^{-1} westward at an altitude of 110 km. As will be seen below, our choice of m_{min} can be justified to some extent by comparing energy dissipation rate curves from our two models with those of Lübken (1997).

Figure 7 shows the energy dissipation rate $\varepsilon(z, \phi_j)$ for each azimuthal sector, together with the energy dissipation rate $\varepsilon(z)$ (the sum over all azimuths). Consider first the northward-propagating, no wind shear, case shown in panel (c). Above about 25 km, the curves differ by no more than a third of the energy dissipation rate for the ultra-simple model. The westward-propagating case in panel (a) has a sharp spike at about 100 km for the ultra-simple model case and not for the full model case. The presence of this feature again shows the limitations of the simplifications we have made in going from the full model to the ultra-simple

model. The wave-energy flux (not shown) helps to explain this spike. At the top of the negative shear at about 100 km, the wave-energy flux for the ultra-simple model is over an order of magnitude larger than for the full model case. The change of sign of shear at 100 km altitude results in an immediate sharp drop in wave-energy flux over a single altitude increment and hence a sharp spike in the energy dissipation rate curve (approximately 40 times higher energy dissipation rate for the ultra-simple model than for the full model) that we see in Fig. 7, panel (a). In the eastward-propagating case shown in panel (b), the energy dissipation rate curves for the ultra-simple spectral parameterisation and the full 3D power-spectral model are very similar. Panel (d) shows $\varepsilon(z)$, the energy dissipation rate summed over all azimuthal directions. The obvious differences between the ultra-simple model and the full model are largely a consequence of the differences between the two models for the westward-propagating azimuth (panel (a)). If we look at the region below the sharp spike, the maximum value of the energy dissipation rate curve (0 · 108 m^2 s^{-3} = 108 mW kg^{-1} at 80 km) is similar to that shown in Lübken (1997) (156 mW kg^{-1} at 90 km) although its peak altitude is somewhat lower. Note, however that Lübken's results are taken at 70°N rather than 40°N, and that a zonally averaged climatalogical average atmosphere may be a poor approximation to reality at these altitudes. We therefore now compare energy dissipation rates $\varepsilon(z)$ for a range of values of m_{min} for the July, 70°N CIRA model atmosphere with Lübken's results.

Figure 8 shows the energy dissipation rate $\varepsilon(z)$ for a number of values of m_{min} with Lübken's energy dissipation rate
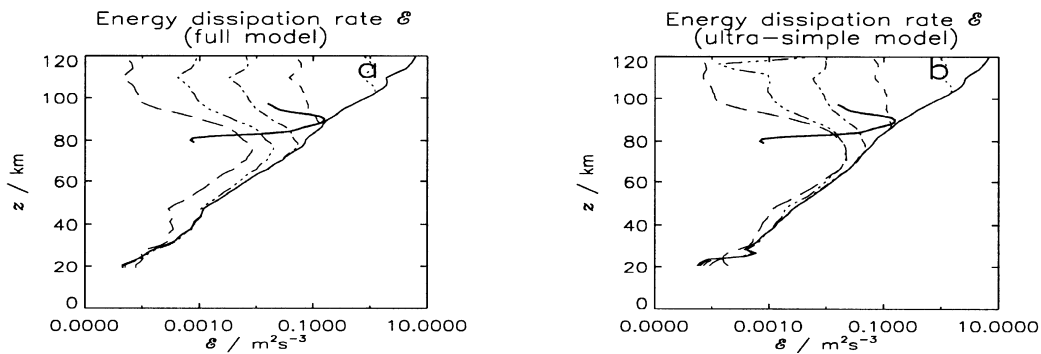
Fig. 8. Comparison of energy dissipation rates for various values of m_{min} with the observational energy dissipation rates of Lübken (1997). Panel (a) shows the full model case and panel (b) the ultra-simple model case. (Bold full curves are observational data. Other curves are model data: full curves are $m_{min} = 2\pi/(100 \text{ km})$, dotted curves are $m_{min} = 2\pi/(50 \text{ km})$, dashed curves are $m_{min} = 2\pi/(20 \text{ km})$, dot-dashed curves are $m_{min} = 2\pi/(10 \text{ km})$, triple dot-dashed curves are $m_{min} = 2\pi/(5 \text{ km})$, and long dashed curves are $m_{min} = 2\pi/(2 \text{ km})$.)

data (Lübken, 1997) superimposed. Panel (a) of Fig. 8 shows the full model case while panel (b) shows the ultra-simple model case. Our choice of 20 km for m_{min} is seen to yield a comparable peak energy dissipation rate and altitude to Lübken's result. Lübken's observational energy dissipation rate peak is much more localised in altitude than is the case for either of our models. This may in part be because a zonally averaged climatological atmosphere may be a poor approximation to reality at these altitudes. Tests with more realistic profiles are planned.

4. Concluding Remarks

The 2D power-spectral model at its current level of simplicity has three substantial shortcomings, which we hope to overcome by work in progress. The first is the neglect of rotation in the dispersion relation (1), one of whose effects is the underestimation of wave-breaking as $\hat{\omega}$ is Doppler-shifted down toward f, in comparison with the wave breaking predicted by the full 3D power-spectral model. The second shortcoming is the lack of back-reflection when $\hat{\omega}$ reaches N; see especially Fig. 5(b).

The third shortcoming is the underestimation of breaking after propagation in negative shear. The problem is due to the true spectral shape taking a very different form from the standard shape assumed in the 2D model. In particular, when a significant altitude range of negative shear is followed by an altitude range of positive shear, there is an underestimation of breaking in the negative-shear region and an overestimation in the positive-shear region. This is one cause of the sharp peak seen in the westward wave-induced force at the top of Fig. 6(a), and in dissipation rate at the top of Fig. 7(a).

The discrepancies between 2D and 3D models are too large for comfort. There is also the problem that (for both models) the force in the summertime upper stratosphere has the wrong sign. A westward force appears critical to the upper branch of the poleward summertime stratospheric circulation (Rosenlof, 1996), and hence to the transport of water vapor from there up into the mesosphere, and hence to the formation of noctilucent and polar mesospheric clouds.

References

Broutman, D., C. Macaskill, M. E. McIntyre, and J. W. Rottman, On Doppler-spreading models of internal waves, *Geophys. Res. Lett.*, **24**, 2813–2816, 1997.

Dewan, E. M. and R. E. Good, Saturation and the "universal" spectrum for vertical profiles of horizontal scalar winds in the atmosphere, *J. Geophys. Res.*, **91**, 2742–2748, 1986.

Eckermann, S. D., Influence of wave propagation on the Doppler spreading of atmospheric gravity waves, *J. Atmos. Sci.*, **54**, 2554–2573, 1997.

Fritts, D. C. and W. Lu, Spectral estimates of gravity wave energy and momentum fluxes II: Parameterisation of wave forcing and variability, *J. Atmos. Sci.*, **50**, 3695–3713, 1993.

Fritts, D. C. and T. E. VanZandt, Spectral estimates of gravity wave energy and momentum fluxes. Part I: Energy dissipation, acceleration, and constraints, *J. Atmos. Sci.*, **50**, 3685–3694, 1993.

Hines, C. O., The saturation of gravity waves in the middle atmosphere. Part I: Critique of linear-instability theory, *J. Atmos. Sci.*, **48**, 1348–1359, 1991.

Hines, C. O., Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. 1. Basic formulation, *J. Atmos. Sol.-Terr. Phys.*, **59**(4), 371–386, 1997a.

Hines, C. O., Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. 2. Broad and quasi monochromatic spectra, and implementation, *J. Atmos. Sol.-Terr. Phys.*, **59**(4), 387–400, 1997b.

Lübken, F.-J., Seasonal variation of turbulent energy dissipation rates at high latitudes as determined by in situ measurements of neutral density fluctuations, *J. Geophys. Res.*, **102**(D12), 13441–13456, 1997.

McIntyre, M. E., On the dynamics and transport near the polar mesopause in summer, *J. Geophys. Res.*, **94**, 14617–14628, 1989.

Rosenlof, K., Summer hemisphere differences in temperature and transport in the lower stratosphere, *J. Geophys. Res.*, **101**, 19129–19136, 1996.

Smith, S. A., D. C. Fritts, and T. E. VanZandt, Evidence of a saturated spectrum of atmospheric gravity waves, *J. Atmos. Sci.*, **44**, 1404–1410, 1987.

Warner, C. D. and M. E. McIntyre, On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere, *J. Atmos. Sci.*, **53**, 3213–3235, 1996.

Warner, C. D. and M. E. McIntyre, Gravity wave spectral models and the shapes of gravity wave spectra at low vertical wavenumbers, in *Gravity Wave Processes: Their Parameterization in Global Climate Models*, edited by Kevin Hamilton, pp. 217–226, Springer-Verlag, Heidelberg, 1997.

C. D. Warner (e-mail: C.D.Warner@damtp.cam.ac.uk) and M. E. McIntyre