On Shear-Generated Gravity Waves that Reach the Mesosphere. Part I: Wave Generation

Oliver Bühler* and Michael E. McIntyre

Centre for Atmospheric Science, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom

John F. Scinocca

Canadian Centre for Climate Modelling and Analysis, University of Victoria, Victoria, British Columbia, Canada

ABSTRACT

A long unresolved issue in nonorographic gravity wave generation is whether there is significant emission from Kelvin–Helmholtz (KH) shear instability in the lower stratosphere, for instance, just above tropopause jets. Such emission has often been suggested as significant for the angular momentum budget and hence for the wave-driven circulation of the middle atmosphere, most crucially in the summer mesosphere. An idealized model thought experiment is studied in which it is assumed that the KH shear instability rapidly mixes a thin layer, producing a ''pancake'' of three-dimensional clear-air turbulence, and emitting low-frequency inertia–gravity waves whose aspect ratio matches that of the turbulent layer and whose horizontal wavelength is large enough to avoid back-reflection and hence reach the summer mesosphere. The wave emission is modeled as a linear initial-value problem in which the rapid mixing of mass and momentum achieved by the turbulence is treated as instantaneous, and hence as determining the initial conditions. Care is taken to cast the problem into a form that permits well-conditioned numerical evaluation of the analytical solution, in both rotating and nonrotating cases, which behave very differently. Comparison with fully nonlinear numerical simulations in two dimensions of the same initial-value problem indicates that the linear theory is much better than might be expected on order-of-magnitude grounds. A companion paper (Part II) investigates the transmission of the emitted waves to the mesosphere subject to refraction and radiative damping.

1. Introduction

Gravity-wave-induced zonal momentum flux is believed on good evidence to account for most of the angular momentum transport into the mesosphere that is known to be crucial for many features of the observed global-scale circulation, especially in the summer mesosphere (e.g., Holton et al. 1995) and also in the upper stratosphere (e.g., Rosenlof 1996). Nonorographic sources need to be involved because orographic waves have critical layers near 20-km altitude (see Fig. 1) and cannot reach the mesosphere. One possible nonoro-

graphic source is Kelvin–Helmholtz (KH) shear instability above tropopause jets, forming pancake-shaped clear-air turbulence (CAT) layers with envelope scales of hundreds of meters in the vertical and tens of kilometers in the horizontal.

However, the horizontal wavelength of individual KH billows is typically far too short to reach the mesosphere or even to be emitted at all. As noted in the companion paper (Bühler and McIntyre 1999, hereafter Part II), waves long enough to reach the mesosphere, that is, long enough to avoid Doppler shifts up to the buoyancy frequency N and consequent back-reflection, need to have horizontal wavelength $\gtrsim 30$ km, implying that if KH emission is important, then it must be nonlinear emission on the envelope scale, somewhat as in the original suggestion by Fritts (1982, 1984) and Chimonas and Grant (1984).

Envelope-scale emission can occur either through nonlinear forcing of waves during the actively turbulent stage of the CAT layer, or during the collapse of the mixed or partially mixed layer that is left behind once the active turbulence has died away. A numerical study of the first mechanism based on nonlinear grouping

* Current affiliation: School of Mathematical and Computational Sciences, University of St. Andrews, The North Haugh, St. Andrews, United Kingdom.

Corresponding author address: Dr. Oliver Bühler, School of Mathematical and Computational Sciences, University of St. Andrews, The North Haugh, St. Andrews KY16 9SS, United Kingdom.
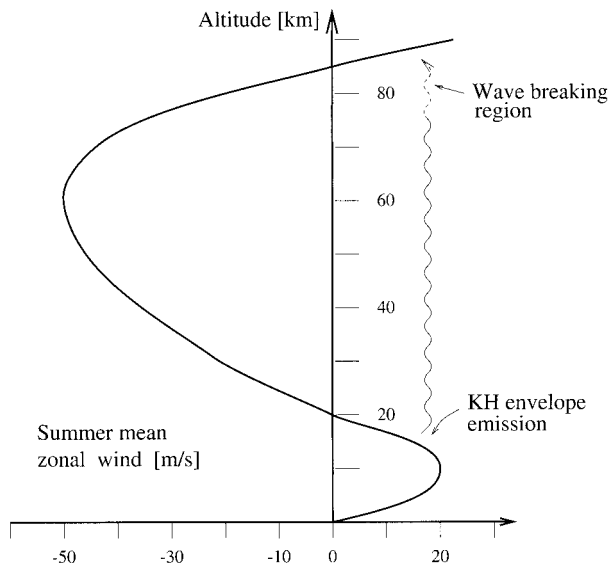E-mail: obuhler@mcs.st-and.ac.uk

FIG. 1. Typical summer mean zonal wind profile. The profile has a zero-wind line at 20-km altitude, which inhibits vertical propagation of orographic, zero-phase-speed waves.

events between individual KH billows has recently been undertaken by Scinocca and Ford (1999). The present paper focuses on the second mechanism for envelope-scale emission, namely, the mixed-layer collapse.

This is a formidable problem, and for tractability's sake a severe idealization of it is studied. The problem is broken down into three stages, as indicated in Fig. 2. Stage A is the initial mixing event, that is, the rapid formation of a CAT layer in which the background gradients of potential temperature $\theta(z)$ and mean zonal wind $u(z)$ are eroded, that is, weakened, by three-dimensional turbulent mixing. This provides the initial conditions for the linear wave emission problem studied in stage B. To allow the initial development of a slowly varying wave train, the mean shear must be neglected at this stage. The main subject of this paper is the analytic solution to this linear emission problem and a comparison of it with fully nonlinear two-dimensional numerical simulations of the same initial-value problem. The latter suggests that the linearization works surprisingly well, despite being far from strictly valid. In stage C

the subsequent development of the wave train is studied using a ray-tracing scheme that takes into account mean shear and wave dissipation by viscous diffusion and infrared radiative damping. This is the subject of the companion paper (Part II). The overall result is that only waves with positive zonal pseudomomentum can reach the summer mesosphere and hence exert a one-signed, ratchetlike eastward force there, and that the average of this force due to a realistic ensemble of stratospheric CAT layers cannot safely be neglected for the global angular momentum budget.

2. Stage A: Formation of KH-induced CAT layer

A comprehensive model for KH and KH-induced CAT in the lower stratosphere would necessarily be very complicated and would also involve subtle fluid-dynamical issues such as the detailed structure of shear-induced gravity wave breaking. Therefore, only one very simple source mechanism of KH is considered in detail here, which we believe is characteristic of many naturally occurring situations. It involves a thin vertical layer of increased static stability (i.e., high buoyancy frequency N) that is rendered KH unstable as the result of transient, that is, temporary, tilting by large-scale motions, producing KH instability in the manner of the Thorpe (1973) tilted-tank laboratory experiments in which the tank was first tilted and then brought back to the horizontal. The large-scale tilting motion could itself be due to, for instance, passing large-scale gravity waves or quasi-horizontal vortices, but it is not necessary to consider a specific model for it.

It is a well-established, if surprising, fact that such tilting motion makes layers with high N the preferred sites of KH instability, despite the fact that for fixed background shear u_z the background Richardson number $\mathrm{Ri} = (N/u_z)^2$ before the tilting is *larger,* and hence more stable, inside these layers. This is because the local shear that develops under large-scale tilting motion is not fixed but is proportional to the local N^2 in the "tilting" situation. In turn, the local Richardson number becomes proportional to N^{-2}; that is, it becomes least stable in
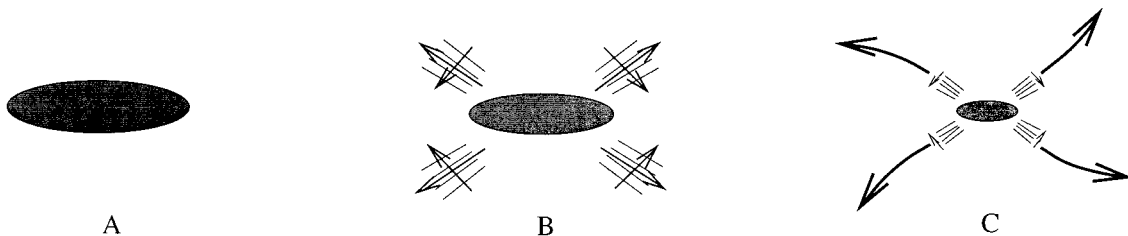

FIG. 2. Three stages of KH-induced wave problem seen in receding zoom: (a) rapid formation of a CAT layer (stage A), (b) initial development of wave train (neglecting mean shear) (stage B), and (c) subsequent development of wave train using ray tracing (taking negative mean shear into account) (stage C). Stage C is the subject of the companion paper (Part II). See text for details.
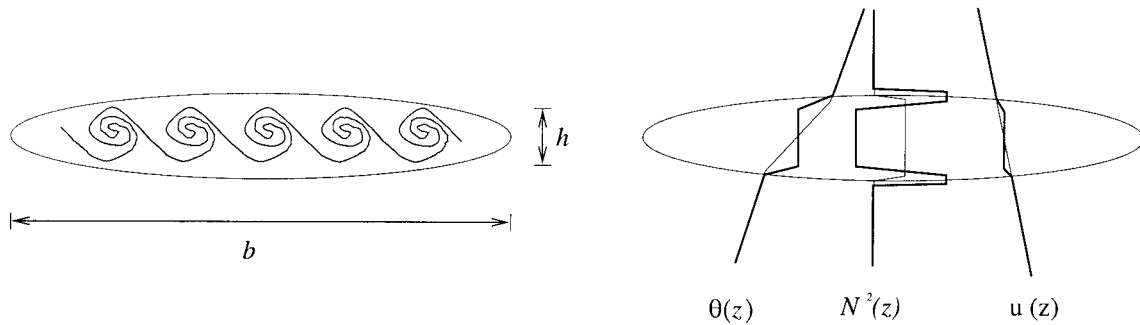
F<small>IG</small>. 3. Outline of typical CAT layer whose horizontal scale b is much larger than its vertical depth h. (The relative vertical depth h/b is exaggerated for clarity.) (left) KH billows rolling up in the initial stage of the event; (right) profiles of potential temperature $\theta(z)$, buoyancy frequency $N^2(z) \propto \theta_z$, and zonal velocity $u(z)$ before (thin lines) and after (bold lines) a strong turbulent mixing event. The layer itself is thought to be positioned in the lower summer stratosphere, i.e., above the tropopause jet, and hence inside a region of negative background zonal shear (cf. Fig. 1), as indicated.

regions of strong stratification.[1] This robust link between strong stratification and increased likelihood of KH is in accordance with many observations in the above-mentioned tilted-tank experiments, and field studies such as the study of KH billows on the ocean thermocline in Woods (1968); also, e.g., Phillips (1977) and Fritts and Rastogi (1985).

It is now assumed that the KH billows eventually break down into three-dimensional small-scale turbulence via a sequence of secondary instabilities, whose exact details are not important for the purpose of this paper. It is further assumed that the three-dimensional turbulence achieves rapid mixing of potential temperature and horizontal velocity, eroding the mean gradients in both fields. When the transient large-scale tilt is approximately reversed, a well-mixed layer of decaying turbulence remains in which the background gradients of $\theta(z)$ and $u(z)$ have been significantly eroded, this erosion being complete in the limiting case of perfect mixing that marks the outcome of a very strong mixing event. Although Ri values must be low during the tilt, we assume that they are relatively large afterward. Figure 3 shows a schematic of this mixed-layer scenario.

How well does this picture of the mixing process compare with field observations and numerical simulations? Observations such as those reported in Browning and Watkins (1970) show conspicuous evidence for the irreversible mixing of θ and the associated "rabbit-ear" structure of the N^2 profile across the layer, which corresponds to homogenized values of θ in the core and to all the stratification being pushed into the edges where θ has to match up with the background temperature profile. Observations of velocity profiles before and af-

ter CAT events are much more scarce and are also difficult to interpret because of the unknown importance of time-dependent background shear in the vicinity of the CAT layer. For instance, the observations of Browning and Watkins (1970) showed little change in the velocity profile before and after the CAT event apart from some degree of overall smoothing.

Direct numerical simulations of KH and its turbulent breakdown into CAT are, at present, barely beginning to achieve a degree of complexity that makes them useful for comparison with real atmospheric events (e.g., Scinocca 1995; and a series of papers including Fritts et al. 1996 and Werne and Fritts 1999). Such simulations are still severely limited in Reynolds number and by the degree of three-dimensionality that they can achieve. Atmospheric CAT Reynolds numbers are at least two orders of magnitude larger than those in the simulations, and hence atmospheric CAT can be expected to be much more vigorous than the simulated turbulence. Lack of full three-dimensionality can lead to qualitatively different long-time behavior such as long-lived two-dimensional vortex rolls that do not break down, contrary to what is conspicuously observed in the atmosphere.

Such long-lived rolls would imply persistent mean shear in the layer. However, once three-dimensional turbulence sets in vigorously, then robust fluid-dynamical energy arguments predict that the mean shear must be eroded to supply the energy for the three-dimensional eddies. These robust arguments underpin, for instance, any theory based on turbulent eddy diffusivity for the mean velocity, provided only that the natural variation of eddy diffusivity across the vertical depth of the layer (i.e., large diffusivity in the well-mixed strongly turbulent core, small diffusivity near the stratified weakly turbulent edges) is taken into account. This leads to the prediction that direct numerical simulations at higher resolution and larger Reynolds numbers will eventually corroborate the assumption made here that potential temperature θ and horizontal velocity u are mixed in essentially the same way in a CAT layer.

---

[1] For example, if Coriolis forces are negligible, then the local shear is proportional to $\gamma(t)N^2$, where $\gamma(t)$ is the time integral of the time-dependent tilting angle $\alpha(t)$ of the layer, and hence Ri $\propto (\gamma N)^{-2}$. In the opposite-extreme case of geostrophic balance, the formula Ri $\propto (\gamma N)^{-2}$ still holds but with $\gamma(t) = \alpha(t)/f$ where f is the Coriolis parameter.

3. Stage B: Analytic wave emission model

The general solution to the linear ''Rossby adjustment'' initial-value problem for a localized, compact initial disturbance embedded in an unbounded rotating Boussinesq system is derived. The compact initial conditions dissolve into two different flow components: propagating inertia–gravity waves and a residual flow in geostrophic balance. The solution is derived in detail for the most complicated three-dimensional rotating case, and, where appropriate, suitable modifications of the formulas to obtain other cases are indicated. Care is taken to obtain a form of the solution that is equally useful for calculating snapshots of the time evolution as well as for calculating the total asymptotic wave emission into different directions.

Sufficient detail is given for a numerical discretization of the solution using digital Fourier transforms (FTs). A nontrivial discretization difficulty arises for spectral modes with zero horizontal wavenumber, and appropriate limits have to be extracted carefully from the solution formulas to yield a robust discretization.

a. General solution to linear initial-value problem

The linearized f-plane Boussinesq equations in a frame of reference moving with the local mean velocity are

$$\mathbf{u}_t + \mathbf{f} \times \mathbf{u} + \nabla P = \sigma \hat{\mathbf{z}} \quad (1)$$

$$\sigma_t + N^2 w = 0 \quad (2)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (3)$$

A right-handed Cartesian coordinate system is used in which x and y are the horizontal coordinates, with x increasing eastward and y increasing northward, and in which $\hat{\mathbf{z}}$ is a vertical unit vector pointing upward in the direction of increase of the vertical coordinate z. The corresponding linearized velocity components are $\mathbf{u} = (u, v, w)$, $\mathbf{f} = f\hat{\mathbf{z}}$ is the Coriolis parameter, P is the pressure in excess of the hydrostatic pressure divided by the constant Boussinesq reference density, N is the constant buoyancy frequency, and σ is the buoyancy acceleration. In the usual way, the no-divergence condition (3), together with suitable boundary conditions at infinity, determines ∇P uniquely, which effectively reduces the number of independent fields by one. Hence the initial-value problem is complete by posing initial conditions for the velocities and the buoyancy acceleration only, that is, by posing

$$\mathbf{u}(x, y, z, 0) = \mathbf{u}_0(x, y, z) \quad \text{and}$$

$$\sigma(x, y, z, 0) = \sigma_0(x, y, z), \quad (4)$$

where the initial field $\mathbf{u}_0$ must satisfy (3). All initial fields are assumed to be localized in a finite region around the origin of the coordinate system, and they are also assumed to be proper disturbance fields in the sense that they must have zero mean value when integrated over the entire spatial domain.

As one would expect for a system with four time derivatives and one constraint (3), for each spectral wavenumber vector $\mathbf{k} = (k, l, m)$ in Fourier space the solution is described by the sum of three modes, namely, a steady mode in geostrophic balance and two time-dependent inertia-gravity-wave modes. The linear steady balanced mode for any given $\mathbf{k}$ can be extracted from the initial conditions in the usual way by ''inverting'' the potential vorticity (PV) of the initial fields, which is

$$Q \equiv v_x - u_y + \frac{f}{N^2}\sigma_z = v_{0x} - u_{0y} + \frac{f}{N^2}\sigma_{0z}. \quad (5)$$

The second equality follows from the identity $Q_t = 0$. Because Q is time-independent, it is clear that only the steady balanced mode can be significant for Q once all the inertia–gravity waves have propagated away. Therefore, for each $\mathbf{k}$ the steady balanced mode is found by setting all time derivatives to zero in (1)–(3) and by assuming that the PV for this $\mathbf{k}$ is entirely due to the steady balanced mode. Denoting the balanced fields by a superscript b, this results in

$$f(-u_z^b, v_z^b) = (\sigma_y^b, \sigma_x^b), \qquad w^b = 0,$$

$$u_x^b + v_y^b = 0 \quad (6)$$

$$\hat{Q} = i\left(k\hat{v}_0 - l\hat{u}_0 + \frac{f}{N^2}m\hat{\sigma}_0\right)$$

$$= i\left(k\hat{v}^b - l\hat{u}^b + \frac{f}{N^2}m\hat{\sigma}^b\right) \quad (7)$$

$$\Rightarrow \{\hat{u}^b, \hat{v}^b, \hat{w}^b, \hat{\sigma}^b\} = \{-l, k, 0, fm\}\frac{k\hat{v}_0 - l\hat{u}_0 + \frac{f}{N^2}m\hat{\sigma}_0}{k^2 + l^2 + m^2 f^2/N^2}. \quad (8)$$

The hat over symbols denotes spatial FTs of fields according to the definition of the n-dimensional spatial FT pair:

$$\hat{g}(\mathbf{k}) \equiv \text{FT}[g(\mathbf{x})] \equiv \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{-i\mathbf{k}\cdot\mathbf{x}}g(\mathbf{x}) \, dx_1 \cdots dx_n$$

$$g(\mathbf{x}) = \text{FT}^{-1}[\hat{g}(\mathbf{k})]$$

$$= \frac{1}{(2\pi)^n} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{+i\mathbf{k}\cdot\mathbf{x}}\hat{g}(\mathbf{k}) \, dk_1 \cdots dk_n, \quad (9)$$

where $g(\mathbf{x})$ is any function of $\mathbf{x}$.

The balanced fields obtained by the PV inversion (8) are subtracted from the initial conditions to yield the relevant initial conditions for the remaining wave fields. Denoting wave fields with superscript w, that is,

$$\{u^w, v^w, w^w, \sigma^w\} \equiv \{u - u^b, v - v^b, w, \sigma - \sigma^b\}, \quad (10)$$

it then follows from (7) that there is no PV associated with the wave fields:

$$Q^w = v_x^w - u_y^w + \frac{f}{N^2}\sigma_z^w = 0. \qquad (11)$$

Many equivalent choices for expressing the spectral eigenvectors of the wave modes are possible, and it now proves convenient to introduce a particular complex amplitude function $\Phi(\mathbf{k})$ that allows treating all variables on an equal footing. In particular, the complex Φ neatly combines both real-valued wave modes, and $|\Phi|^2$ will turn out to be the relevant spectral density of wave energy. The definition of Φ is

$$\Phi(\mathbf{k}) \equiv -\frac{\sqrt{k_h^2 + m^2}}{k_h}\left(\hat{w}_0^w + i\hat{\omega}(\mathbf{k})\frac{\hat{\sigma}_0^w}{N^2}\right), \qquad (12)$$

where $k_h \equiv \sqrt{k^2 + l^2}$ is the magnitude of the horizontal wavenumber vector and $\hat{\omega}(\mathbf{k})$ is the intrinsic frequency given by the positive branch of the Boussinesq inertia-gravity wave dispersion relation:

$$\hat{\omega} \equiv \sqrt{f^2\frac{m^2}{k_h^2 + m^2} + N^2\frac{k_h^2}{k_h^2 + m^2}} \qquad \hat{\omega}(\mathbf{k}) = \hat{\omega}(-\mathbf{k})$$

$$N \geq \hat{\omega} \geq |f|. \qquad (13)$$

The indicated range of $\hat{\omega}$ holds if $N > |f|$.

With the aid of Φ the wave fields can be written as

$$u^w = \Re \mathrm{FT}^{-1}\left\{e^{-i\hat{\omega}t}\Phi(\mathbf{k})\left[\frac{k}{k_h} + i\frac{l}{k_h}\frac{f}{\hat{\omega}}\right]\frac{m}{\sqrt{k_h^2 + m^2}}\right\} \qquad (14)$$

$$v^w = \Re \mathrm{FT}^{-1}\left\{e^{-i\hat{\omega}t}\Phi(\mathbf{k})\left[\frac{l}{k_h} - i\frac{k}{k_h}\frac{f}{\hat{\omega}}\right]\frac{m}{\sqrt{k_h^2 + m^2}}\right\} \qquad (15)$$

$$w^w = \Re \mathrm{FT}^{-1}\left\{e^{-i\hat{\omega}t}\Phi(\mathbf{k})\frac{-k_h}{\sqrt{k_h^2 + m^2}}\right\} \qquad (16)$$

$$\sigma^w = \Re \mathrm{FT}^{-1}\left\{e^{-i\hat{\omega}t}\Phi(\mathbf{k})i\frac{N^2}{\hat{\omega}}\frac{k_h}{\sqrt{k_h^2 + m^2}}\right\}. \qquad (17)$$

Here $\Re$ denotes taking the real part and use has been made of the fact that the FT of a purely real or purely imaginary function $g(\mathbf{x})$ satisfies $\hat{g}(\mathbf{k}) = \pm\hat{g}^*(-\mathbf{k})$, respectively, where the star denotes taking the complex conjugate. For instance, combining (12) and (16) at $t = 0$ gives $w^w = \Re \mathrm{FT}^{-1}\{\hat{w}_0^w + i\hat{\omega}\hat{\sigma}_0^w/N^2\}$, and due to $\hat{\omega}(\mathbf{k}) = \hat{\omega}(-\mathbf{k})$ the second term results in a purely imaginary FT^{-1}, leaving only w_0^w as the real part.

The two-dimensional case is included by setting the respective wavenumber component to zero (e.g., $l \equiv 0$ if there is no y dependence) and using $n = 2$ in the definition of the FT pair in (9). Also, the nonrotating case is included by setting $f = 0$. The preceding analytic solution in terms of Φ can straightforwardly be evaluated numerically using digital FTs, except for modes with horizontal wavenumber $k_h = 0$. For these modes, it turns out to be vital for a robust and accurate numerical discretization that appropriate limiting values are care-

fully extracted from the preceding formulas, and this task is described in the appendix. Using these limiting values, it was sufficient to use 32 Fourier coefficients in each spatial direction to obtain convergence for the emission strength (i.e., convergence in the diagnostic $\mathcal{P}_*^l$ defined in section 3b), and it was sufficient to use 256 Fourier coefficients in each spatial direction to obtain very accurate snapshots of the time evolution of the wave field.

b. Asymptotic wave emission

The solution derived in the last section describes in detail how compact initial conditions unfold into a residual (steady) balanced flow and a propagating train of inertia-gravity waves. Asymptotically, that is, after sufficiently long time, this wave train can simply be characterized in terms of a local wavenumber and a suitable local wave amplitude. It is now demonstrated how the function $\Phi(\mathbf{k})$ defined in (12) can be used to calculate these local wave train characteristics. This is useful for two reasons: first, it gives a precise answer to the question of how much wave activity is radiated in different directions; second, it provides the relevant initial conditions for ray-tracing schemes in which the subsequent propagation of the wave train under the influence of mean shear, radiative damping, etc., can be studied (cf. the companion paper).

Use is made of the well-known fact (e.g., Whitham 1974, p. 371 ff; or Lighthill 1978, p. 351 ff) that if $s(\mathbf{x}, t)$ is a field with compact initial conditions and given in terms of a spectral amplitude $S(\mathbf{k})$ as

$$s = \Re \mathrm{FT}^{-1}\{e^{-i\hat{\omega}t}S(\mathbf{k})\}, \qquad (18)$$

and if $\mathcal{D}(t)$ is a *time-varying* region in n-dimensional $\mathbf{x}$ space, the boundary of which is moving with the local group velocity $\mathbf{c}_g = \mathbf{x}/t$, then

$$\int_{\mathcal{D}(t)} \overline{s^2}\, d\mathbf{x} = \frac{1}{2}\frac{1}{(2\pi)^n}\int_{\mathcal{K}} |S|^2\, d\mathbf{k} \qquad (19)$$

holds for sufficiently long times. Here $\overline{(\cdots)}$ denotes an average over the phase of the wave train, and $\mathcal{K}$ denotes the *fixed* region in $\mathbf{k}$ space corresponding pointwise to $\mathcal{D}(t)$ through the group-velocity condition:

$$(c_g)_i \equiv \frac{\partial\hat{\omega}}{\partial k_i}(\mathbf{k}) = \frac{x_i}{t}. \qquad (20)$$

Equation (19) holds with s^2 replaced by the square of any of the wave fields defined in (10), and this allows evaluating the wave energy per unit mass contained in a region $\mathcal{D}(t)$ by replacing the $\mathbf{x}$ integral over $\mathcal{D}(t)$ of the wave energy density

$$E \equiv \frac{1}{2}[\overline{|\mathbf{u}^w|^2} + \overline{(\sigma^w)^2}/N^2] \qquad (21)$$

by a $\mathbf{k}$ integral of a suitable spectral density over the

corresponding region $\mathcal{K}$. Combining (14)–(17), (13), (19), and (21),

$$\int_{\mathcal{D}(t)} E \, d\mathbf{x} = \frac{1}{2} \frac{1}{(2\pi)^n} \int_{\mathcal{K}} |\Phi|^2 \, d\mathbf{k} \qquad (22)$$

is found, which shows that $|\Phi|^2$ is the spectral wave energy density. Global conservation of wave energy is included as a special case by letting $\mathcal{D}$, and hence $\mathcal{K}$, cover the entire domain, that is,

$$\mathcal{E}^w \equiv \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} E \, dx_1 \cdots dx_n$$

$$= \frac{1}{2} \frac{1}{(2\pi)^n} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |\Phi|^2 \, dk_1 \cdots dk_n. \quad (23)$$

The total disturbance energy of the initial conditions, say, $\mathcal{E}_0$, can be written as the sum of $\mathcal{E}^w$ and the energy of the balanced flow $\mathcal{E}^b$. Specifically,

$$\mathcal{E}^b \equiv \frac{1}{2} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\overline{|\mathbf{u}^b|^2} + \overline{(\sigma^b)^2}/N^2) \, dx_1 \cdots dx_n,$$

and

$$\mathcal{E}^b + \mathcal{E}^w = \mathcal{E}_0. \qquad (24)$$

The derivation leading to (22) is unaffected if E is multiplied by a quantity that remains constant on group velocity rays. The local wavenumber vector $\mathbf{k}$ and the local frequency $\hat{\omega}(\mathbf{k})$ are such quantities, and this allows the simple evaluation of the x component of horizontal pseudomomentum per unit mass that is contained in $\mathcal{D}(t)$ as

$$\int_{\mathcal{D}(t)} \frac{k}{\hat{\omega}} E \, d\mathbf{x} = \frac{1}{2} \frac{1}{(2\pi)^n} \int_{\mathcal{K}} \frac{k}{\hat{\omega}} |\Phi|^2 \, d\mathbf{k}. \qquad (25)$$

General theoretical results from wave–mean interaction theory show that the vertical flux of horizontal pseudomomentum is equal at leading order to the wave-induced vertical flux of mean horizontal momentum (e.g., Andrews and McIntyre 1978a,b; McIntyre and Norton 1990). In other words, measuring the vertical flux of zonal pseudomomentum is equivalent to measuring the wave-induced flux of zonal momentum. Hence, pseudomomentum is the appropriate diagnostic for quantifying wave-induced momentum fluxes. It can be noted that the pseudomomentum density $kE/\hat{\omega}$ has a sign according to the sign of k and hence is not sign definite. This means that there is no obvious upper bound on the amount of pseudomomentum generated by given initial conditions.

Let $\mathcal{P}^I(t)$ denote the total amount of pseudomomentum found at time t upward and eastward of a compact source centered at the origin, that is,

$$\mathcal{P}^I(t) \equiv \int_{x>0} \int_{-\infty}^{+\infty} \int_{z>0} \frac{k}{\hat{\omega}} E \, dx \, dy \, dz; \qquad (26)$$

then (25) allows the simple evaluation of the limiting behavior of $\mathcal{P}^I(t)$ as $t \to \infty$ as

$$\lim_{t\to\infty} \mathcal{P}^I(t) = \frac{1}{2} \frac{1}{(2\pi)^3} \int_{k>0} \int_{-\infty}^{+\infty} \int_{m<0} \frac{k}{\hat{\omega}} |\Phi|^2 \, dk \, dl \, dm. \qquad (27)$$

Note that $k > 0$ selects waves with group velocity (and phase velocity) directed eastward, and that $m < 0$ selects waves with positive vertical group velocity. A nondimensional quantity $\mathcal{P}^I_*$ is now introduced as

$$\mathcal{P}^I_* \equiv \frac{HN}{\mathcal{E}_0} \lim_{t\to\infty} \mathcal{P}^I(t), \qquad (28)$$

where H is a length scale provided by the initial conditions. The quantity $\mathcal{P}^I_*$ is useful, as it can be easily evaluated for two- and three-dimensional versions of the linear solution, and it can also be approximately evaluated in a nonlinear simulation.

c. Mixed-layer initial conditions

Initial conditions are chosen that resemble in the simplest possible way the net mixing of mass and horizontal momentum in the CAT layer. The initial conditions consist of a relatively well-mixed core embedded in a smooth Gaussian envelope; see (31) and Fig. 4. In the core, the net mixing erodes the vertical background gradients of density and of the zonal velocity, which are N^2 and $-N/\sqrt{\text{Ri}}$, respectively, using the standard definition of the background Richardson number Ri. The sign of the background shear is chosen for consistency with Fig. 1 but turns out not to affect the emission process. The initial conditions thus chosen determine u_0 and σ_0, and the remaining velocity components v_0 and w_0 are chosen to satisfy the continuity equation (3), that is, $v_{0y} + w_{0z} = -u_{0x}$. In the three-dimensional case this does not determine v_0 and w_0 uniquely, and an additional condition must be specified. We envisage that the turbulent mixing acts on the mean velocities like a comparatively rapid zonal force-pair that erodes the zonal mean shear. It is straightforward to show (e.g., Batchelor 1967) that in the limit of an impulsive zonal force (i.e., a force that acts over a time interval $t \ll 1/N$), the (v_0, w_0) response consists of irrotational (i.e., $v_{0z} = w_{0y}$) two-dimensional motion in each y–z plane driven solely by a mass source–sink equal to $-u_{0x}$. This would lead to

$$\{\hat{v}_0, \hat{w}_0\} = -\frac{\{kl, \, km\}}{l^2 + m^2} \hat{u}_0 \qquad (29)$$

for the respective velocity FTs. This result cannot be used unmodified because KH mixing times do not usually satisfy $t \ll 1/N$. However, for the flat, pancakelike initial conditions with small aspect ratio h/b considered here, a modified version can be shown to hold for longer time intervals $t \ll b/(hN)$. Significant values of $\hat{u}_0$ are then found only in regions of spectral space satisfying
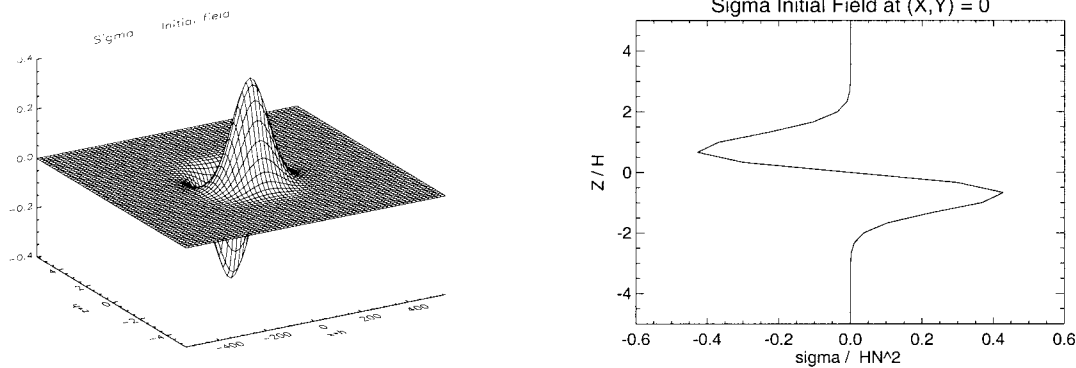
FIG. 4. (left) Initial σ disturbance over x–z plane at $y = 0$. Note that the direction into the paper corresponds to the vertical. (right) Initial σ disturbance over z at $(x, y) = 0$.

$k_h/|m| \sim h/b \ll 1$. This allows use of the hydrostatic approximation to (29), in which $O(k_h^2/m^2)$ terms are neglected. This leads finally to

$$\{\hat{v}_0, \hat{w}_0\} = \left\{0, -\frac{k}{m}\hat{u}_0\right\}. \tag{30}$$

The specific initial fields centered at the origin of the (x, y, z) coordinate system are given by

$$\{u_0, v_0, w_0, \sigma_0\} = \left\{+\alpha z\frac{N}{\sqrt{\text{Ri}}}, 0, -\alpha x\frac{H^2}{L^2}\frac{N}{\sqrt{\text{Ri}}}, -\beta z N^2\right\}$$
$$\times \exp\{-[(x^2 + y^2)/L^2 + z^2/H^2]\}. \tag{31}$$

The two-dimensional case is included by setting $y = 0$. The mixing parameters β and α can range from 0 to 1 and measure the extent to which the respective background gradients have been eroded. The limiting case of perfect mixing at the center of the core corresponds to a value of 1. The Gaussian envelope is determined by the vertical and horizontal half-width scales H and L. The corresponding initial disturbance energy is

$$\mathcal{E}_0 = H^4 N^2 \frac{\pi}{16}\frac{L}{H}\left[\frac{\alpha^2}{\text{Ri}}\left(1 + \frac{H^2}{L^2}\right) + \beta^2\right]$$

in two dimensions   and

$$\mathcal{E}_0 = H^5 N^2 \frac{\pi}{16}\sqrt{\frac{\pi}{2}}\frac{L^2}{H^2}\left[\frac{\alpha^2}{\text{Ri}}\left(1 + \frac{H^2}{L^2}\right) + \beta^2\right]$$

in three dimensions. (32)

The shape of σ_0 (which is the same as u_0 with a minus sign) is depicted in Fig. 4 for $\beta = 1$. The extent of the well-mixed core of the mixed region indicated by h and b in Fig. 3 must be related to envelope half-widths H and L. Assuming that $h/b = H/L$ and defining h as the vertical distance between the two extremes of the initial disturbance profile in Fig. 4, the relation

$$(h, b) = \sqrt{2}(H, L) \tag{33}$$

is obtained.

It is now convenient to nondimensionalize the problem and to use all its obvious symmetries in order to reduce the number of free parameters. The Boussinesq model has no intrinsic length scale, and hence a length scale of the initial conditions, say, H, can be used for that purpose. This leaves only the aspect ratio H/L to set. The choice

$$H/L = h/b = 1/100 \tag{34}$$

is compatible with many observations (e.g., Browning and Watkins 1970), and for simplicity's sake this value of H/L is used throughout.

The Boussinesq model has two intrinsic timescales, N and f. Using N to nondimensionalize the problem leaves open the choice of Prandtl's ratio f/N. The following values for Prandtl's ratio are used:

$$f/N = \{0, 0.005, 0.01\}, \tag{35}$$

which roughly span the stratospheric range of this parameter.

The mixing parameters β and α may vary independently in physical reality, but in the nondimensional linear problem only their ratio enters. This is because multiplying both parameters by the same number only multiplies all wave fields by that number, and in particular leaves $\mathcal{P}_*^l$, which is a nondimensional ratio of pseudomomentum and wave energy, unchanged. Furthermore, inspection of (31) and (32) shows that Ri enters only in the ratio $\alpha/\sqrt{\text{Ri}}$. Hence, α, β, and Ri can all be combined into a single parameter $\alpha/(\beta\sqrt{\text{Ri}})$. For simplicity, only the following few values for this parameter are considered in the following linear calculation:

$$\alpha/(\beta\sqrt{\text{Ri}}) = \{0, 0.5, 1\}. \tag{36}$$

Letting $\alpha/(\beta\sqrt{\text{Ri}})$ range from 0 to 1 brackets the extremes of relative importance of kinetic energy to available potential energy that is contained in the initial mixed layer. At one extreme, the kinetic energy is unimportant while at the other extreme there is equipar-

tition between both kinds of energy. Inclusion of the opposite extreme (i.e., dominating kinetic energy) was omitted for brevity and because we felt this case was less relevant to real CAT, for which it is natural to assume, influenced by the "tilted-tank" experiments, that density mixing is always involved at leading order. Including the case $\alpha/(\beta\sqrt{\mathrm{Ri}}) \to 0$ gives the important test case (for comparison with numerical simulations at least) of density mixing only, with zero background shear and temporary tilted-tank conditions.

In summary, the nondimensional diagnostic $\mathcal{P}_*^l$ in (28) depends only on $\alpha/(\beta\sqrt{\mathrm{Ri}})$, f/N, H/L, and the number of spatial dimensions. In comparison, the total pseudomomentum, as defined in (27), has some additional explicit dependence on the model parameters as follows. The total pseudomomentum has dimensions H^4N, (or H^3N in two dimensions). Hence, if the size of the mixed region is doubled while the values of f/N, H/L, Ri, and β, α are kept fixed, then the total amount of pseudomomentum is increased by a factor of 16 (or 8 in two dimensions), highlighting the sensitive dependence of the pseudomomentum emission on layer size. Also, if N is doubled under the same conditions, then the total pseudomomentum is doubled. Finally, the total pseudomomentum is proportional to the squares of the mixing parameters, that is, doubling both β and α quadruples the total pseudomomentum.

d. Model runs and upward–eastward pseudomomentum emission

The wave emission has been calculated in two and three dimensions and for the parameter combinations discussed above. It turns out that in all cases the upward–eastward pseudomomentum emitted due to joint velocity and density mixing is simply the sum of the upward–eastward pseudomomentum emitted due to either velocity or density mixing in isolation. In other words, the pseudomomentum fluxes due to velocity mixing and due to density mixing simply add. This is a nontrivial result, which depends on certain special (but quite reasonable) properties of the chosen initial conditions, as follows.

First, consider the spectral wave energy density $|\Phi|^2$, defined by (12), in the nonrotating two-dimensional case. In this case all balanced fields are 0. For the chosen initial conditions in (31) the property $u_0 \propto \sigma_0$ holds throughout the domain, and therefore (using $\hat{w}_0 = -k\hat{u}_0/m$) the complex FTs $\hat{u}_0$, $\hat{w}_0$, and $\hat{\sigma}_0$ all have the same phase in spectral space. This means that $|\Phi|^2$ in (12) is given for each spectral wavenumber vector $\mathbf{k}$ as a sum of squares of initial velocity and density FTs, without cross-correlation terms between these velocity and density FTs. In other words, $|\Phi(\mathbf{k})|^2$ is simply a (weighted) sum of α^2 and β^2.

A corresponding statement is trivially true for the rotating two-dimensional case as well as for the nonrotating three-dimensional case. In these cases the dis-

TABLE 1. Nondimensional wave energy $\mathcal{E}^w/\mathcal{E}_0$. Each column corresponds to a different value of f/N and each row corresponds to a different value of $\alpha/(\beta\sqrt{\mathrm{Ri}})$. Left and right entries correspond to two-dimensional and three-dimensional cases, respectively. The value of $\mathcal{E}^w/\mathcal{E}_0$ does not depend on the signs of α and β. In the nonrotating two-dimensional case the PV is identically zero and, hence, $\mathcal{E}^w = \mathcal{E}_0$ there. The differences between two-dimensional and three-dimensional values are in many cases significant.

	f/N					
α	0		0.005		0.01	
$(\beta\sqrt{\mathrm{Ri}})$	2D	3D	2D	3D	2D	3D
0	1.00	1.00	0.44	0.64	0.25	0.40
½	1.00	0.90	0.56	0.65	0.40	0.48
1	1.00	0.76	0.72	0.66	0.62	0.60

turbance PV is due only to a single term, which is either $f\sigma_z/N^2$ or $-u_y$, respectively. Using the expression for $\hat{\sigma}^b$ in (8), we see that the wave field $\hat{\sigma}_0^w = \hat{\sigma}_0 - \hat{\sigma}^b$ in (12) is then simply proportional to $\hat{\sigma}_0$ (in the $f = 0$ case simply because $\sigma^b \equiv 0$), with a real-valued factor of proportionality that is either less than or equal to unity, in the respective cases. Again, this implies that there are no velocity–density cross-correlation terms in $|\Phi(\mathbf{k})|^2$.

On the other hand, in the rotating three-dimensional case the PV depends on a linear combination of velocity and density contributions, and this leads to an expression for $\hat{\sigma}_0^w$ in (12) that depends on both α and β. Consequently, $|\Phi(\mathbf{k})|^2$ now includes a cross-correlation term $\propto -2\alpha\beta lmf$. Nevertheless, for initial conditions that factorize in the y coordinate (which includes those considered here) the contribution of this cross-term to an integral of $|\Phi|^2$ over all y wavenumbers l must vanish. This is because the cross-correlation term in (12) is then necessarily an odd function of l, as can be easily shown.

All of the above remains true if $|\Phi|^2$ is replaced by the spectral pseudomomentum density $|\Phi|^2 k/\hat{\omega}$. Both the total wave energy $\mathcal{E}^w$ and the total pseudomomentum emitted upward–eastward involve an integral over all l, and hence both are simply additive over velocity and density mixing in all cases. One consequence of this is that $\mathcal{E}^w/\mathcal{E}_0$ as well as $\mathcal{P}_*^l$ in (28) can only depend on $|\alpha/(\beta\sqrt{\mathrm{Ri}})|$, implying, for instance, that the signs of the initial conditions in (31) cannot affect the net upward–eastward pseudomomentum emission.

Consider now the numerical results for the ratio $\mathcal{E}^w/\mathcal{E}_0 = (\mathcal{E}_0 - \mathcal{E}^b)/\mathcal{E}_0$ in Table 1, which describes how $\mathcal{E}_0$ is split between energy of the propagating gravity waves and energy of the balanced residual state. Basically, this ratio is diminished as the PV content of the initial conditions is increased, and vice versa. Hence, there is a general decrease of $\mathcal{E}^w/\mathcal{E}_0$ as f/N is increased. In the two-dimensional cases $\mathcal{E}^w/\mathcal{E}_0$ increases as the velocity mixing strength α is increased, which is because velocity does not contribute to the PV in these cases.

The shape of the pseudomomentum density $|\Phi|^2 k/\hat{\omega}$ differs in an interesting way from that of the energy
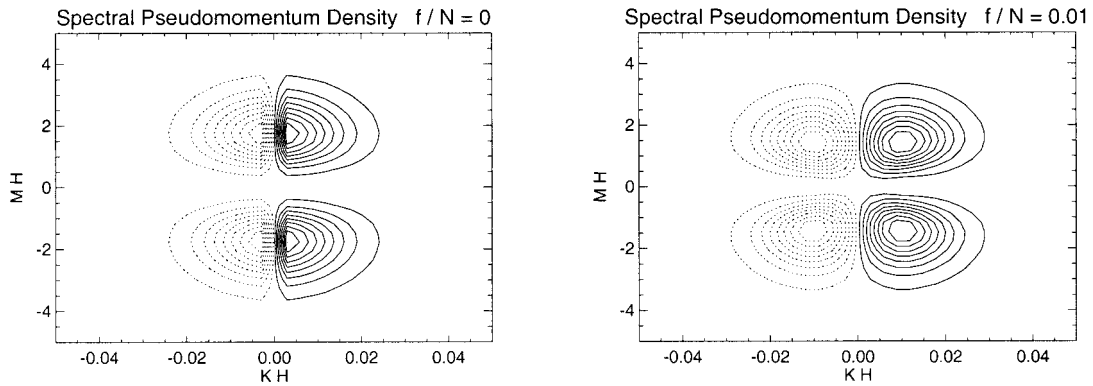
FIG. 5. Spectral pseudomomentum density. Solid contours denote positive values, dotted contours denote negative values: (left) without background rotation; (right) with background rotation equal to $f = N/100$. In both cases $\alpha = \beta$.

density $|\Phi|^2$. A typical shape of $|\Phi|^2 k/\hat{\omega}$ is illustrated in Fig. 5, where the two-dimensional $|\Phi|^2 k/\hat{\omega}$ is contoured as a function of k and m for two cases, $f = 0$ on the left and $f = N/100$ on the right. The total pseudomomentum emitted upward–eastward is equal to the integral over the lower-right quadrant, in which $k \geq 0$ and $m \leq 0$. Note that the pseudomomentum density takes both positive and negative values, which could not occur for the everywhere nonnegative energy density. The plots illustrate that the total pseudomomentum radiated upward (or downward) is 0, that is, initial conditions such as (31) emit a wave train with an eastward–westward symmetry. There is also a conspicuous difference between the rotating and nonrotating cases, namely, that in the rotating case (on the right) there is a well-defined maximum of the spectral pseudomomentum density in the interior of each quadrant at

$$|kH| \cong 0.01 \qquad |mH| \cong 1.5, \qquad (37)$$

whereas in the nonrotating case there is a jump discontinuity along $k = 0$ that connects extreme values on either side of the discontinuity. This is due to the factor $k/\hat{\omega}$, which in the nonrotating case jumps at $k = 0$, and

in the rotating case goes to zero at $k = 0$, as illustrated in Fig. 6 and further discussed in the appendix.

Numerical results for $\mathcal{P}^l_*$ are collected in Table 2. It can be noted that $\mathcal{P}^l_*$ always decreases as f/N increases, and that, somewhat surprisingly, the differences between two-dimensional and three-dimensional values of $\mathcal{P}^l_*$ are small, except in the case $f = 0$. This can be compared with the values for $\mathcal{E}^w/\mathcal{E}_0$ in Table 1, in which these differences are in many cases significant. This appears to be due to the factor $k/\hat{\omega}$ in the spectral pseudomomentum density, which counteracts to some extent the dependence of the spectral wave energy density on the number of spatial dimensions.

In summary, for values of the Prandtl ratio f/N comparable with the aspect ratio H/L, there is little difference in $\mathcal{P}^l_*$ between two-dimensional and three-dimensional cases. This allows the use of two-dimensional emission models in these cases. If, for example, $f/N \approx H/L$, then, depending on the relative strength of the density and the velocity mixing, the values of $\mathcal{P}^l_*$ range through

$$f/N \approx H/L \Rightarrow \mathcal{P}^l_* = [0.05 \ldots 0.10]. \qquad (38)$$

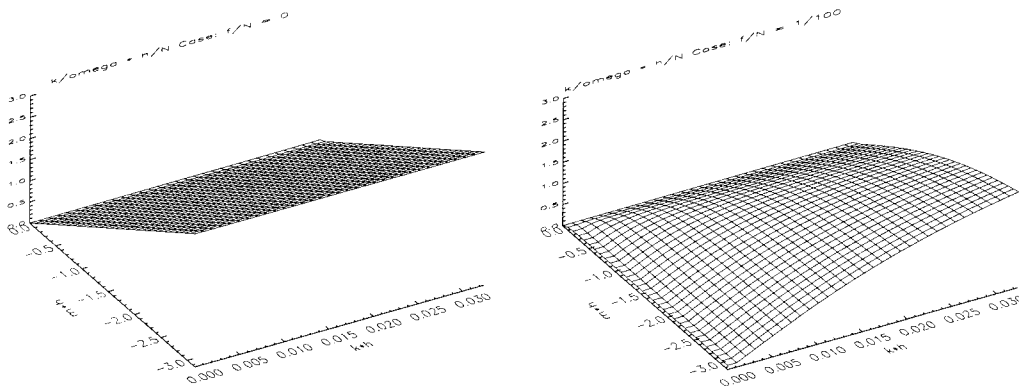
FIG. 6. The factor $k/\hat{w}$ over the quadrant ($k \geq 0$, $m \leq 0$) of spectral space: (left) without background rotation; (right) with background rotation equal to $f = N/100$.

TABLE 2. Nondimensional pseudomomentum $\mathcal{P}_*^I$; cf. Table 1 for details. The values for the nonrotating two-dimensional case depend only very weakly on $\alpha/(\beta\sqrt{\mathrm{Ri}})$ and are equal to the precision shown here. The value of $\mathcal{P}_*^I$ does not depend on the signs of α and β.

$\dfrac{\alpha}{(\beta\sqrt{\mathrm{Ri}})}$	f/N
	\multicolumn{2}{c}{0}	\multicolumn{2}{c}{0.005}	\multicolumn{2}{c}{0.01}
	2D	3D	2D	3D	2D	3D
0	0.40	0.25	0.12	0.12	0.05	0.06
½	0.40	0.24	0.14	0.13	0.07	0.07
1	0.40	0.21	0.17	0.13	0.10	0.09

If, on the other hand, $f = 0$, then the three-dimensional results must be used because the two-dimensional model greatly overpredicts the emission strength. The range of $\mathcal{P}_*^I$ for $f = 0$ is

$$f/N = 0 \Rightarrow \mathcal{P}_*^I = [0.21 \ldots 0.25]. \qquad (39)$$

e. Sensitivity to small-scale details in the envelope structure

The Gaussian envelope structure of the initial conditions in (31) was chosen for its simplicity, and it is important to check whether the pseudomomentum emission could be sensitive to small-scale details of the envelope structure. Clearly, such small-scale details would be very hard to predict because they are likely to be, to some degree, dependent on the details of the actively turbulent phase of the CAT layer. The emitted wave energy $\mathcal{E}^w$ is always bounded by $\mathcal{E}_0$ and is hence insensitive to small-scale changes. However, as noted before, there is no such obvious bound for the emitted pseudomomentum $\mathcal{P}^I$. It is the pseudomomentum that gives the wave-induced momentum flux, and hence its sensitivity needs to be considered.

The behavior of $k/\hat{\omega}$, which multiplies the spectral energy density to obtain the spectral pseudomomentum density, is investigated for this purpose. Sensitivity to small-scale detail is indicated if $k/\hat{\omega}$ can grow without bound as the typical wavenumber vector $\mathbf{k} = (k, l, m)$ takes on large absolute values. Rotating and nonrotating cases need to be considered separately, and $f = 0$ is considered first. Waves with zonal wavenumber k larger than $\approx 2\pi/30$ km will be back-reflected by the easterly jet before they reach the summer mesosphere (as shown in detail in Part II); and hence only bounded values of k need to be considered. Large absolute values of meridional wavenumber l mean that $\hat{\omega} \to N$ in $k/\hat{\omega}$, but this leads only to bounded values provided that k remains bounded. Hence, only waves with large absolute values of vertical wavenumber m need to considered.

In this case, $k/\hat{\omega} \to |m|k/(k_h N)$, which exhibits unbounded growth as $|m|$ grows (cf. left panel in Fig. 6). This implies sensitivity of the pseudomomentum emission to vertical small-scale details, that is, to details of the vertical envelope shape. However, these large-$|m|$ waves necessarily have low vertical group velocities,

and hence spend a long time near their launch altitude where they are subject to dissipative processes such as radiative damping. Radiative damping acts very effectively on large-$|m|$ waves (*provided* that background rotation is absent; this crucial condition is investigated in detail in Part II); and hence these waves will be dissipated near their launch altitude and not reach the mesosphere.

The above points concerning the horizontal wavenumbers k and l apply equally in the rotating case $f \neq 0$. In the rotating limit of large $|m|$ the factor $k/\hat{\omega} \to k/f$, which is again bounded as long as k is bounded; and hence there is no sensitivity of the pseudomomentum emission to vertical small-scale details in the rotating case. In summary, it can be concluded that the amount of pseudomomentum that reaches the summer mesosphere is probably quite insensitive to the small-scale details of the initial conditions.

4. Nonlinear numerical simulations

Linear theory is not formally valid in the early stage of the emission, when the wave train first develops from the mixed region. This is true especially in the limiting case of perfect mixing, in which isentropes acquire vertical slopes. For this reason, independent fully nonlinear simulations of the emission stage have been performed to test the accuracy of the linear theory. The simulations were two-dimensional, that is, their domain was restricted to the x–z plane, but a three-dimensional velocity field $\mathbf{u} = (u, v, w)$ was considered. The flow fields are then y independent but, in the presence of nonzero background rotation f, the velocity component v in the y direction is nontrivially coupled to the other components.

This allows, at reasonable computational cost, some form of dynamically significant nonlinear vortex stretching to take place, in addition to allowing nonlinear effects caused by the interplay of advection and buoyancy forces in the x–z plane. The latter effects may actually be stronger in two than in three dimensions, due to the comparatively less rapid amplitude decrease of the dispersing flow fields in two dimensions.

Various simulations with and without background rotation and with or without a balanced flow component have been performed, and good quantitative agreement with the linear theory could be demonstrated. Although these simulations provide a good test of the linear predictions, there remains one caveat, namely, that mean shear has again been neglected.

a. Numerical model

The numerical model used was the Clark model (Clark 1977). It is a finite-difference gridpoint model, which in these simulations solved the nonlinear two-dimensional Boussinesq equations with background ro-
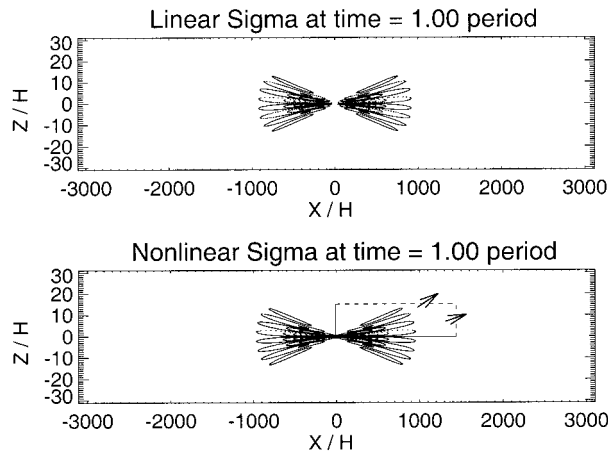
FIG. 7. Contour plots of linear and nonlinear σ field after one period of time: $t = 200\pi/N$. Note the stretched vertical coordinate. The broken lines show the control sides across which the upward–eastward pseudomomentum flux is calculated. This is case B of the cases summarized below.

tation with a three-dimensional velocity field $\mathbf{u} = (u, v, w)$, that is,

$$\mathbf{u}_t + (\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u} + f\hat{\mathbf{z}} \times \mathbf{u} + \boldsymbol{\nabla}P = \sigma\hat{\mathbf{z}} + \nu\nabla^2\mathbf{u} \quad (40)$$

$$\sigma_t + (\mathbf{u} \cdot \boldsymbol{\nabla})\sigma + N^2w = \nu\nabla^2\sigma \quad (41)$$

$$\boldsymbol{\nabla} \cdot \mathbf{u} = 0, \quad (42)$$

where all fields are y independent; $\hat{\mathbf{z}}$ is the upward unit vector; $\boldsymbol{\nabla} = (\partial/\partial x, 0, \partial/\partial z)$; and $N, f,$ and ν are constants.

The vertical scale of the initial conditions H and the buoyancy frequency N were chosen as length and (inverse) timescales. The Reynolds number H^2N/ν was very large, of the order of 10^6, and hence the simulations were nearly inviscid. The resolution was always 400×400 grid points, and the (x, z) domain size was always $(62.155, 6215.5)H$. Outer boundary conditions were chosen as radiation conditions. This was achieved by applying a no-stress condition at the boundary, and by smoothly increasing Rayleigh damping in all model fields toward the outer boundaries of the computational domain. The simulations showed no sign of spurious back-reflection from the boundary.

b. Pseudomomentum diagnostics

Consider Fig. 7, which shows a snapshot at a given time t of the nonlinear σ field in a particular simulation (case B in the summary below). Note that for comparison the corresponding linear solution using 256×256 spectral modes is also shown, indicating the good qualitative agreement in this case.

Now, the pseudomomentum radiation upward–eastward that has occurred up to time t is denoted by $P_{\rm rad}(t)$, and it can be estimated as the sum of two parts as follows. The first part is the pseudomomentum content at time t in the dashed–solid control rectangle located in

the first quadrant of Fig. 7. The second part is the time-integrated (from the initial time up to time t) pseudo-momentum flux across the two dashed control sides of this rectangle.

Hence, given a suitably defined pseudomomentum density per unit mass p and pseudomomentum flux vector $\mathbf{F}$, the diagnostic $P_{\rm rad}(t)$ is calculated as

$$P_{\rm rad}(t) \equiv \iint {\rm p}\, dx\, dz + \int_0^t \left(\int \mathbf{F} \cdot \mathbf{n}\, ds \right) dt, \quad (43)$$

where the area integral is extended over the dashed–solid control rectangle in Fig. 7, and the line integral, in which $\mathbf{n}$ is the outward unit normal vector, is extended over the two dashed sides, which form *part* of the boundary of the control rectangle. This flux integration is further symbolized by the two arrows in Fig. 7, which highlight the fact that the flux integration leaves out the solid edges of the control rectangle. Clearly, in cases where linear theory is quantitative applicable, $P_{\rm rad}(t) \approx \mathcal{P}^I(t)$, where $\mathcal{P}^I(t)$ is defined by the two-dimensional version of (26).

There is, however, a problem that requires some discussion. Suitable nonlinear definitions for the pseudomomentum density p and its flux $\mathbf{F}$ in terms of the Eulerian flow variables σ and $\mathbf{u}$ are readily available in the nonrotating case $f = 0$ (e.g., Scinocca and Shepherd 1992). Here it is understood that "suitably defined" implies not only that the pseudomomentum is conserved, but also that the vertical pseudomomentum flux is equal to the leading-order wave-induced flux of zonal momentum in the usual way. However, it turns out that such a suitable Eulerian definition of pseudomomentum is not available in the rotating case $f \neq 0$. This fact seems to be related to the essential need, in the rotating case, for additional Lagrangian particle displacement information in the sideways (y) direction in order to obtain a suitable pseudomomentum definition. This point is currently investigated further by one of us (O. Bühler, 1999, unpublished manuscript).

Here, a simple far-field approximation is used that avoids introducing additional Lagrangian fields. Consider the following set of equations, which are derived by straightforward manipulations of the nonlinear Boussinesq equations (40)–(42):

$$\mathrm{p}_0 \equiv \frac{\sigma}{N^2}(u_z - w_x) \quad (44)$$

$$F_x = \mathrm{p}_0 u + \frac{1}{2}\left(\frac{\sigma^2}{N^2} + u^2 - w^2 - v^2 - \frac{v^2\sigma_z}{N^2} \right) \quad (45)$$

$$F_z = \mathrm{p}_0 w + uw - \frac{f}{N^2}\sigma v + \frac{v^2}{2}\frac{\sigma_x}{N^2} \quad (46)$$

$$Q \equiv (f\hat{\mathbf{z}} + \boldsymbol{\nabla} \times \mathbf{u}) \cdot (\hat{\mathbf{z}} + \boldsymbol{\nabla}\sigma/N^2) - f \quad (47)$$

$$= v_x + \frac{f}{N^2}\sigma_z + \frac{1}{N^2}(v_x\sigma_z - v_z\sigma_x) \quad (48)$$

TABLE 3. Parameters used in the nonlinear simulations. The velocity mixing rates in cases C, D, and F are chosen such that the minimum local Richardson number in the center of the mixed region is ¼. The notional background Richardson number Ri = 1. Only case E has nonzero potential vorticity.

Case	Velocity mixing α	Density mixing β	f/N
A	0.0	0.5	0
B	0.0	1.0	0
C	2.0	0.0	0
D	1.41	0.5	0
E	0.0	0.5	0.01
F	2.0	0.0	0.01

$$\frac{\partial \mathrm{p}_0}{\partial t} + \nabla \cdot \mathbf{F} + \nu Q = 0, \qquad (49)$$

Here p_0 is the Eulerian pseudomomentum density of the nonrotating case $f = 0$ for uniform background fields (cf. Scinocca and Shepherd 1992), and Q is the PV in excess of the background PV divided by N^2, and hence $Q_t + (\mathbf{u} \cdot \nabla)Q = 0$.

Now, in all the numerical cases studied, p_0 and the underlined parts of $\mathbf{F}$ in (45) and (46) have been used in (43) in order to calculate the diagnostic $P_{\mathrm{rad}}(t)$. It is straightforward to show that in the small-amplitude WKB limit both p_0 and the underlined parts of $\mathbf{F}$ reduce to the standard WKB expressions $kE/\hat{\omega}$ and $\mathbf{c}_g kE/\hat{\omega}$ for pseudomomentum density and flux, where $\mathbf{c}_g$ is the group velocity. Importantly, this is true in both rotating and nonrotating cases. Hence, provided that the dashed control edges in Fig. 7 lie in the far field where the flow has already dispersed into a small-amplitude slowly varying wave train, the underlined terms can be used as an approximation to $\mathbf{F}$.

Near the origin, on the other hand, p_0 is not strictly conserved if $\nu Q \neq 0$ [cf. (49)], and the use of p_0 as the pseudomomentum density is hence only approximate. Still, the balanced part of p_0 is clearly zero by inspection (because both u^b and w^b are zero even in nonlinear balance); and hence the approximation becomes better with increasing time t. This is because more and more waves will by then have propagated away from the initial seat of the disturbance, where Q might be nonzero. This is corroborated by case E of the numerical simulations.

c. Run parameters and results

Six nonlinear runs have been performed. The same initial conditions as in the linear theory were used, that is, the initial fields were given by (31) with $y = 0$, $H/L = 0.01$, and with varying mixing and rotation parameter settings that are collected in Table 3. The notional background Richardson number has been set to Ri = 1 for simplicity. However, as in the linear theory, no mean shear has actually been present in the simulations.

The density mixing parameter β ranged between 0 and 1, and the velocity mixing parameter α was either

0 or chosen such that the minimum local Richardson number in the center of the mixed region [which is given there by $(1 - \beta)/\alpha^2$] was equal to a quarter. This was done in order to achieve maximal initial disturbance amplitudes while avoiding unstable initial conditions. Only cases E and F had nonzero background rotation, and only case E has nonzero potential vorticity.

The results for $P_{\mathrm{rad}}(t)$ are collected in Fig. 8. The corresponding linear estimates from the previous section are in all cases indicated by a broken line. In all cases with zero Q, $P_{\mathrm{rad}}(t)$ quickly converges to an asymptotic value close to the linear estimate. On the other hand, in case E the convergence is markedly slower and also involves stronger oscillations. It has been checked that these strong oscillations are due to slowly decaying correlations between the balanced, steady part of σ and the wavelike, unsteady part of $u_z - w_x$. These correlations impact on p_0 as defined in (44), and hence on the first part of (43). It has also been checked that these oscillations occur equally when diagnosing a time-dependent linear solution.

The agreement between the linear and nonlinear estimates for the emitted pseudomomentum has been remarkably good, suggesting an error margin of perhaps 20%. This provides a good check on the use of linear theory for the emission stage, if, as noted before, the effect of mean shear can indeed be neglected.

5. Concluding remarks

A simple linear model for the emission of inertia–gravity waves by shear-generated mixed layers of CAT in the lower stratosphere has been formulated. The solution to this mixed-layer linear initial-value problem in two and three dimensions has been derived, with due care over numerical details in order to achieve a robust discretization of it in both the rotating and the nonrotating cases, the latter being the relevant case in equatorial regions.

The total pseudomomentum emission upward–eastward from a mixed layer has then been computed over a range of parameters (as summarized in Table 2), and these results have been been partially cross-checked by independent nonlinear numerical simulations that showed very good agreement with the linear theory. The formulated model therefore appears to capture most of the relevant features of the emission problem, with the possible exception of strong nonlinear effects associated with strong mean shear on the scale of the mixed region. It is conceivable that nonlinear effects due to strong shear might significantly disrupt the formation of the inertia-gravity wave train, with unclear implications for the resulting pseudomomentum fluxes. Hence, the impact of strong shear on mixed-layer emission remains a fruitful topic of further research, which will however require techniques completely different from the ones used here.

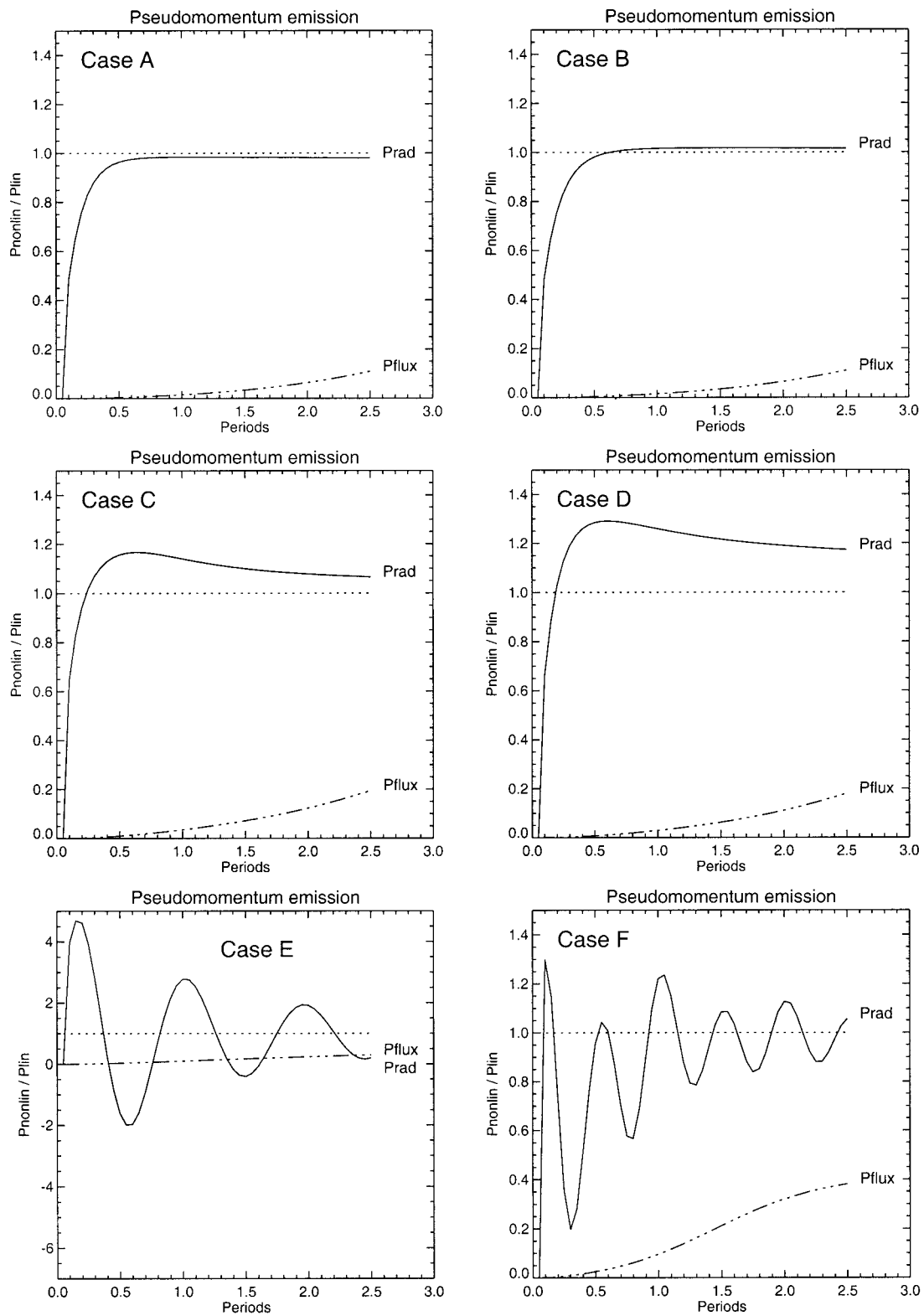The companion paper (Part II) takes up the question

FIG. 8. Nonlinear pseudomomentum diagnostics. The broken lines represent the linear estimates. Note the much larger vertical scale for case E. Each period corresponds to a time unit of $200\pi/N$. The diagnostic "P flux" corresponds to the second, flux-related part of $P_{rad}(t)$ in its definition (43).

of vertical propagation of the emitted wave spectrum to the mesosphere.

APPENDIX

Discretization Formulas for Modes with $k_h = 0$

As $k_h \to 0$ the appropriate limits for the balanced fields in (8), the amplitude function Φ in (12), and the wave fields in (14)–(17) must be extracted carefully. If $k_h \to 0$ and $m \to 0$ simultaneously, then Φ, as well as the balanced and the wave fields, goes to 0 because all initial fields are assumed to have zero mean value. Therefore, only the limit $k_h \to 0$ and $m \neq 0$ needs to be considered. It turns out that rotating and nonrotating cases must be considered separately, and that the rotating case is easier to deal with than the nonrotating case.

In the rotating case (i.e., $f \neq 0$) the balanced fields in (8) tend to

$$\{\hat{u}^b, \hat{v}^b, \hat{w}^b, \hat{\sigma}^b\} \to \{0, 0, 0, \hat{\sigma}_0\} \qquad \text{(A1)}$$

as $k_h \to 0$ and $m \neq 0$. The amplitude function Φ in (12) tends to [using $\nabla \cdot \mathbf{u} = 0$ and (11)]

$$\Phi \to \text{sgn}(m)\left[\hat{u}_0^w\left(\frac{k}{k_h} - i\frac{l}{k_h}\right) + \hat{v}_0^w\left(\frac{l}{k_h} + i\frac{k}{k_h}\right)\right], \quad \text{(A2)}$$

where $\text{sgn}(m) \equiv \pm 1$ if $m > 0$ or $m < 0$, respectively. This shows that the limit of Φ as $k_h \to 0$ is path-dependent; that is, it depends on the limiting values of k/k_h and l/k_h, and therefore Φ is discontinuous at this limit. However, $|\Phi|$ as well as all the combinations appearing as integrands in (14)–(17), are in fact *independent* of the limiting values of k/k_h and l/k_h. Hence, any choice of limiting values will lead to the same spectral wave energy density $|\Phi|^2$ and wave fields. For definiteness, the convention

$$(k/k_h, l/k_h) \to (1, 0) \Rightarrow \Phi \to \text{sgn}(m)(\hat{u}_0^w + i\hat{v}_0^w) \quad \text{(A3)}$$

is adopted for the numerical discretization.

In the nonrotating case (i.e., $f = 0$) the balanced fields tend to

$$\{\hat{u}^b, \hat{v}^b, \hat{w}^b, \hat{\sigma}^b\} \to \left\{\frac{l^2}{k_h^2}\hat{u}_0 - \frac{kl}{k_h^2}\hat{v}_0, \frac{k^2}{k_h^2}\hat{v}_0 - \frac{kl}{k_h^2}\hat{u}_0, 0, 0\right\}, \quad \text{(A4)}$$

that is, $\hat{u}^b$ and $\hat{v}^b$ are discontinuous in this limit and

$\hat{\sigma}^b \to 0$, which can be compared with the completely different limits in the rotating case. The function Φ tends to

$$\Phi \to \text{sgn}(m)\left(\hat{u}_0^w\frac{k}{k_h} + \hat{v}_0^w\frac{l}{k_h}\right) - i\frac{\hat{\sigma}_0^w}{N}. \qquad \text{(A5)}$$

Now both Φ and $|\Phi|$ are discontinuous in this limit. Using (A5) it can be checked that the sum of the balanced fields in (A4) and the wave fields in (14)–(17) is continuous in this limit. This means that snapshots of the balanced plus wave fields will not be affected by the choice of limiting value for k/k_h and l/k_h. The same convention as in (A3) is adopted, and hence

$$\{\hat{u}^b, \hat{v}^b, \hat{w}^b, \hat{\sigma}^b\} \to \{0, \hat{v}_0, 0, 0\} \quad \text{and} \quad \text{(A6)}$$

$$\Phi \to \text{sgn}(m)(\hat{u}_0^w - i\hat{\sigma}_0^w) \qquad \text{(A7)}$$

are obtained.

The genuine discontinuity as $k_h \to 0$ of the spectral wave energy density $|\Phi|^2$, while not affecting the validity of the analytic solution, remains an obstacle for calculating $\mathcal{P}_*^l$ accurately. In the numerical discretization $|\Phi|^2$ has to be evaluated (and integrated) pointwise in spectral space; and hence the suitability of the convention (A7) must be checked. In the two-dimensional nonrotating case the discontinuity of $|\Phi|$ affects the pseudomomentum estimate $\mathcal{P}_*^l$ significantly because $|\Phi|$ reaches its maximum for fixed m at $k = 0$, as was illustrated in Figs. 5 and 6. The limiting value of Φ at $k = 0$ in (A7) is then consistent with the limit $k \to 0+$, that is, with $k = 0$ being approached through positive values of k. Hence, the convention (A7) allows the desired pseudomomentum estimate for eastward-moving waves, which have $k > 0$, to integrate correctly. In the three-dimensional nonrotating case the discontinuity of $|\Phi|$ turns out not to affect $\mathcal{P}_*^l$ significantly. This is because the locus of the discontinuity is now only a line in three-dimensional spectral space, as opposed to a line in two-dimensional spectral space in the two-dimensional case. Upon integration the relative importance of the values of Φ on that line compared to the importance of the values of Φ throughout the spectral domain is very small. The convention (A7) therefore provides a simple limit that is accurate in both two-dimensional and three-dimensional cases.

REFERENCES

Andrews, D. G., and M. E. McIntyre, 1978a: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.,* **89,** 609–646.

——, and ——, 1978b: On wave-action and its relatives. *J. Fluid Mech.,* **89,** 647–664; Corrigendum, **95,** 796; also **106,** 331.

Batchelor, G. K., 1967: *An Introduction to Fluid Dynamics.* Cambridge University Press, 615 pp.

Browning, K. A., and C. D. Watkins, 1970: Observations of clear air turbulence by high power radar. *Nature,* **227,** 260–263.

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.

Chimonas, G., and J. R. Grant, 1984: Shear excitation of gravity waves. Part II: Upscale scattering from Kelvin–Helmholtz waves. *J. Atmos. Sci.,* **41,** 2278–2288.

Clark, T. L., 1977: A small scale numerical model using a terrain following coordinate system. *J. Comput. Phys.,* **24,** 186–215.

Fritts, D. C., 1982: Shear excitation of atmospheric gravity waves. *J. Atmos. Sci.,* **39,** 1936–1952.

——, 1984: Shear excitation of atmospheric gravity waves. Part II: Nonlinear radiation from a free shear layer. *J. Atmos. Sci.,* **41,** 524–537.

——, and P. K. Rastogi, 1985: Convective and dynamical instabilities due to gravity wave motions in the lower and middle atmosphere: Theory and observations. *Radio Sci.,* **20,** 1247–1277.

——, T. L. Palmer, O. Andreassen, and I. Lie, 1996: Evolution and breakdown of Kelvin–Helmholtz billows in stratified compressible flows. Part I: Comparison of two- and three-dimensional flows. *J. Atmos. Sci.,* **53,** 3173–3191.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister, 1995: Stratosphere–troposphere exchange. *Rev. Geophys.,* **33,** 403–439.

Lighthill, M. J., 1978: *Waves in Fluids.* Cambridge University Press, 504 pp.

McIntyre, M. E., and W. A. Norton, 1990: Dissipative wave–mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.,* **212,** 403–435; Corrigendum, **220,** 693.

Phillips, O. M., 1977: *The Dynamics of the Upper Ocean.* 2d ed. Cambridge University Press, 336 pp.

Rosenlof, K., 1996: Summer hemisphere differences in temperature and transport in the lower stratosphere. *J. Geophys. Res.,* **101,** 19 129–19 136.

Scinocca, J. F., 1995: The mixing of mass and momentum by Kelvin–Helmholtz billows. *J. Atmos. Sci.,* **52,** 2509–2530.

——, and T. G. Shepherd, 1992: Nonlinear wave-activity conservation laws and Hamiltonian structure for the two-dimensional anelastic equations. *J. Atmos. Sci.,* **49,** 3–25.

——, and R. Ford, 1999: The nonlinear forcing of large-scale internal gravity waves by stratified shear instability. *J. Atmos. Sci.,* in press.

Thorpe, S. A., 1973: Experiments on instability and turbulence in a stratified shear flow. *J. Fluid Mech.,* **61,** 731–752.

Werne, J., and D. C. Fritts, 1999: Stratified shear turbulence: Evolution and statistics. *Geophys. Res. Lett.,* **26,** 439–442.

Whitham, G. B., 1974: *Linear and Nonlinear Waves.* Wiley Interscience, 620 pp.

Woods, J. D., 1968: Wave-induced shear instability in the summer thermocline. *J. Fluid Mech.,* **32,** 791–800.