A balanced approach to modelling rotating stably stratified geophysical flows

By DAVID G. DRITSCHEL[1] AND ÁLVARO VIÚDEZ[2]

[1]School of Mathematics and Statistics, University of St Andrews, St Andrews, UK

[2]Institut de Ciènces del Mar, Passeig Maritim 37–49, 08003 Barcelona, Spain

We describe a new approach to modelling three-dimensional rotating stratified flows under the Boussinesq approximation. This approach is based on the explicit conservation of potential vorticity, and exploits the underlying leading-order geostrophic and hydrostatic balances inherent in these equations in the limit of small Froude and Rossby numbers. These balances are not imposed, but instead are used to motivate the use of a pair of new variables expressing the departure from geostrophic and hydrostatic balance. These new variables are the ageostrophic horizontal vorticity components, i.e. the vorticity not directly associated with the displacement of isopycnal surfaces. The use of potential vorticity and ageostrophic horizontal vorticity, rather than the usual primitive variables of velocity and density, reveals a deep mathematical structure and appears to have advantages numerically. This change of variables results in a diagnostic equation, of Monge–Ampère type, for one component of a vector potential $\boldsymbol{\varphi}$, and two Poisson equations for the other two components. The curl of $\boldsymbol{\varphi}$ gives the velocity field while the divergence of $\boldsymbol{\varphi}$ is proportional to the displacement of isopycnal surfaces. This diagnostic equation makes transparent the conditions for both static and inertial stability, and may change form from (spatially) elliptic to (spatially) hyperbolic even when the flow is statically and inertially stable. A numerical method based on these new variables is developed and used to examine the instability of a horizontal elliptical shear zone (modelling a jet streak). The basic-state flow is in exact geostrophic and hydrostatic balance. Given a small perturbation however, the shear zone destabilizes by rolling up into a street of vortices and radiating inertia–gravity waves.

1. Introduction

The atmosphere and oceans are frequently observed to be in a state of near 'balance'. This means that inertia–gravity waves, or the gravitational oscillations of density (or entropy) surfaces, do not often contribute significantly to the mean observed motions (important exceptions occur in the upper atmosphere where breaking gravity waves decelerate the mean flow, and elsewhere). Instead, the observed motions are to a great extent determined by a nearly materially conserved field, the potential vorticity, together with approximate 'balance relations' which hold instantaneously and filter the inertia–gravity waves (Hoskins, McIntyre & Robertson 1985; Ford, McIntyre & Norton 2000 and references therein).

The simplest balance relations are geostrophic and hydrostatic balance, in which respectively the horizontal acceleration (relative to the rotating Earth) and the vertical acceleration are neglected in the momentum equations. There are higher-order more

complex balance relations which appear capable of better representing the balanced component of the flow (cf. McIntyre & Norton 2000; Mohebalhojeh & Dritschel 2000, 2001). They are judged to be better because, for a given flow, they attribute a greater proportion of the flow to balanced motions. What is left over is called the 'imbalance' or (often misleadingly) the 'gravity waves'. Just what part of the imbalance is truly gravity waves is difficult to know precisely. Part of the problem is that this decomposition into balanced and imbalanced motions is flow dependent; hence, the balance relations must be tailored to the flow regime. No one set of balance relations is optimal for a given evolving flow.

In general, if a flow has both a small Froude number (ratio of horizontal vorticity ω_h to buoyancy frequency N) and a small Rossby number (ratio of relative vertical vorticity ζ to background planetary vorticity f), then one can expect high-order balance relations to hold and little interaction between the imbalanced and balanced parts of the flow. Under these conditions, the balanced part of the flow does not generate significant imbalance, i.e. through spontaneous gravity wave emission. Hence, if the imbalanced part of the flow is initially weak, it remains so. This is the basis for the initialization and data assimilation methods used in weather forecasting. Those methods apply balance relations to filter the initial imbalance, a significant component of which is considered to be due to data inaccuracies and model errors.

While the principle of balance can be applied strictly to design approximate models having no imbalance whatsoever, one need not go this far to exploit the underlying balanced structure in the equations of motion. Mohebalhojeh & Dritschel (2000, 2001) have shown, in the single-layer shallow-water equations, that one can simultaneously make this balanced structure explicit while retaining evolution equations for the imbalanced part of the flow. That is, balance is not imposed but, rather, it is used to motivate a choice for a pair of variables to represent the imbalanced part of the flow. The original 'primitive' variables are then recovered from a set of diagnostic relations. And here is the key point: these diagnostic relations reduce to balance relations when the time derivatives of the imbalance variables are set to zero. The underlying balance is explicit, and so is the imbalance. All of this is gained without approximating the equations of motion.

In practice, shallow-water numerical simulations based on this transformation are significantly more accurate than those using the original primitive variables (see Dritschel & Mohebalhojeh 2000), particularly when the Froude and Rossby numbers are small. The problem with the original variables is that they mix balanced and imbalanced motions, and, inevitably, numerical discretization errors lead to a false numerical generation of gravity waves. This in turn can lead to significant errors in the balanced part of the flow (Dritschel & Mohebalhojeh 2000).

In general, the balanced part of the flow depends only on the instantaneous distribution of potential vorticity. In the absence of frictional and diabatic processes, potential vorticity is a materially conserved field (this conservation is also known as Cauchy's vorticity formula). Under these conditions (assumed in this work), the potential vorticity is a natural prognostic variable for representing the balanced motion. It is not frequently used however, due to its nonlinear dependence on the original primitive variables. In the shallow-water equations, this is not a major limitation, although it can lead to nonlinear diagnostic equations for the primitive variables depending on the choice of imbalance variables (these are no worse however than the balance relations one must solve in balanced models, and indeed in some particularly useful cases they reduce to simple linear equations even when the corresponding balance relations are nonlinear).

In the three-dimensional rotating–stratified fluid equations examined in this work, the use of potential vorticity is more problematic than in the shallow-water equations. Inevitably, one is forced to solve a nonlinear diagnostic equation to recover the original variables no matter what choice is made for the imbalance variables. However, the situation is not as bleak as it appears, because a variety of choices for the imbalance variables leads to an appealing diagnostic equation, of the Monge–Ampère type, about which much is known. Here, we adopt a special choice of imbalance variables that represents the departure from the simplest leading-order geostrophic and hydrostatic balance. This does not lead to the best representation of the balanced part of the flow for small Froude and Rossby numbers, but there are practical reasons for not trying anything more elaborate. First, the Monge–Ampère structure would be lost and we would probably confront a more complicated, potentially less robust, nonlinear equation about which little or nothing is known. Second, the numerical cost of solving such an equation and computing the evolution of the imbalance variables would probably be far greater. The design of a numerical algorithm forces a compromise between mathematical exactness and efficiency.

The plan of the paper is as follows. In §2, the governing equations are reviewed and then transformed to use potential vorticity and the ageostrophic horizontal vorticity as prognostic variables. Several physical examples are discussed next in §3. These examples exhibit the spontaneous emission of inertia–gravity waves, albeit at surprisingly weak amplitudes even for moderately large Rossby and Froude numbers. In the vertical velocity field, this emission is of secondary importance, roughly five orders of magnitude smaller than the horizontal velocity field. Some conclusions and ideas for future work are given in §4. In an Appendix we outline the numerical algorithm developed to implement this theoretical approach.

2. The rotating–stratified fluid equations recast

2.1. *Primitive variables*

We consider the equations for a rotating stratified incompressible fluid whose density $\rho(\boldsymbol{x}, t)$ departs weakly from a background constant value ρ_0, as in the oceans. Under the so-called 'Oberbeck–Boussinesq' approximation (Oberbeck 1879; Boussinesq 1903), quadratic and higher-order corrections in $(\rho - \rho_0)/\rho_0$ are neglected. The difference $\rho - \rho_0$ is split into a mean linear part $\varrho_z z$, where $\varrho_z < 0$ is a constant, and an anomaly $\rho'(\boldsymbol{x}, t)$, or explicitly:

$$\rho'(\boldsymbol{x}, t) \equiv \rho(\boldsymbol{x}, t) - \varrho_z z - \rho_0. \tag{2.1}$$

Then, the conservation of mass, which becomes the material conservation of density ($\dot{\rho} = 0$) for an incompressible fluid, may be written

$$\dot{\rho}' + \varrho_z w = 0, \tag{2.2}$$

where $(\dot{\ }) = \mathrm{d}(\ )/\mathrm{d}t = (\ )_t + \boldsymbol{u} \cdot \nabla(\ )$ denotes the material time derivative (in the rotating frame), and w is the vertical component of the velocity.

Conservation of linear momentum, in a reference frame rotating with constant angular velocity $f/2$ around the vertical z-axis with respect to an inertial frame, becomes (under the approximation above)

$$\dot{\boldsymbol{u}} + f\boldsymbol{k} \times \boldsymbol{u} = -\alpha_0 \nabla \Phi - \alpha_0 g \rho' \boldsymbol{k}, \tag{2.3}$$

where $\boldsymbol{u}$ is the three-dimensional velocity field, $\Phi(\boldsymbol{x}, t)$ is the geopotential, g is gravity, $\alpha_0 \equiv \rho_0^{-1}$, and $\boldsymbol{k}$ denotes the vertical unit vector. One last equation is required to close

the system of equations: the isochoric condition

$$\nabla \cdot \boldsymbol{u} = 0. \tag{2.4}$$

Note that, as a consequence of (2.4), the full equations have only three independent time derivatives.

2.2. *The transformation*

Various transformations are commonly used to solve the equations above in a different way, for example the Craya–Herring decomposition (which expresses the velocity in terms of spatial derivatives of two scalar potentials, see e.g. Smith & Waleffe 2002 and references therein). However, we are not aware of any transformation having the form of a balance–imbalance decomposition – the Craya–Herring decomposition appears to be most useful for studying flows dominated by inertia–gravity waves, just the inverse of the situation considered in the present work. The transformation we use is described next.

2.3. *Buoyancy*

It is convenient to introduce the buoyancy field

$$b \equiv -\alpha_0 g \rho', \tag{2.5}$$

which appears multiplying $\boldsymbol{k}$ on the right-hand side of (2.3). From (2.2), we have

$$\dot{b} + N^2 w = 0, \tag{2.6}$$

where $N^2 \equiv -\alpha_0 g \varrho_z > 0$. The quantity N is the mean buoyancy frequency.

2.4. *Balance*

To motivate the choice of variables to be adopted, consider the equations that result from imposing geostrophic and hydrostatic balance:

$$f \boldsymbol{k} \times \boldsymbol{u}_h = -\alpha_0 \nabla_h \Phi, \tag{2.7a}$$

$$0 = -\alpha_0 \Phi_z + b. \tag{2.7b}$$

where a subscript h on any vector denotes the horizontal component of that vector, and ∇_h denotes the horizontal gradient operator (subscripts (x, y, z, t) denote partial derivatives). Now, we can eliminate Φ by taking a z-derivative of (2.7a) and substituting (2.7b) to give

$$-f v_z = -b_x, \tag{2.8a}$$

$$f u_z = -b_y, \tag{2.8b}$$

which are often referred to as the 'thermal-wind relations' (although here we are considering the ocean). Now, a very slight modification of the left-hand side of these equations yields

$$f \xi = -b_x, \tag{2.9a}$$

$$f \eta = -b_y, \tag{2.9b}$$

where $\xi \equiv w_y - v_z$ and $\eta \equiv u_z - w_x$ are the x- and y-components of the vorticity $\boldsymbol{\omega} = (\xi, \eta, \zeta)$. This is a very slight modification to the thermal wind relations not only because w is typically much smaller than u and v, but also because z derivatives are typically much larger than x- and y-derivatives.

2.5. *Imbalance*

We shall not impose these balance relations in what follows, but merely use them to motivate choosing the horizontal components of the vector

$$\mathscr{A} \equiv \boldsymbol{\omega}/f + \nabla b/f^2 \tag{2.10}$$

as the variables representing the imbalance. We stress that these variables only represent the true imbalance to leading order; they might be more appropriately called 'ageostrophic, non-hydrostatic vorticity', since these variables still contain (and may even be dominated by) balanced motions. When discussing the true imbalance below, we will refer to it as 'inertia–gravity waves' for clarity.

It is convenient at this stage to introduce the shorthand notation $\tilde{\chi} \equiv \chi/f$ for any quantity χ, and to express the buoyancy in terms of an isopycnal displacement $\mathscr{D} \equiv -b/N^2$ (note that $\dot{\mathscr{D}} = w$). Then we can rewrite (2.10) as

$$\mathscr{A} \equiv \tilde{\boldsymbol{\omega}} - c^2 \nabla \mathscr{D}, \tag{2.11}$$

where $c \equiv N/f$. We will also make use of the inverse of c denoted $\epsilon \equiv f/N$ (note that $\epsilon \sim 10^{-2}$ to 10^{-1} in the oceans).

Now let $\mathscr{A} = \nabla^2 \boldsymbol{\varphi}$, where $\boldsymbol{\varphi}$ is a vector potential. Then the divergence of (2.11) gives

$$\mathscr{D} = -\epsilon^2 \nabla \cdot \boldsymbol{\varphi}, \tag{2.12}$$

whereas the curl of (2.11) (using also (2.4)) gives

$$\tilde{\boldsymbol{u}} = -\nabla \times \boldsymbol{\varphi}. \tag{2.13}$$

Thus, from the potential $\boldsymbol{\varphi}$, we can recover all three components of the velocity field as well as the isopycnal displacement field.

Similar relations were used by Muraki, Snyder & Rotunno (1999, see their equation 27) in an attempt to develop an improved quasi-geostrophic model. They used the Helmholtz decomposition $(v, -u, \theta) = \nabla \Phi + \nabla \times \Psi$ where θ is potential temperature, and set $\boldsymbol{k} \cdot \Psi = 0$. The relations between the potentials in both approaches are therefore, scaling factors omitted, $\boldsymbol{\varphi}_h \to \boldsymbol{k} \times \Psi_h$ and $\phi \to \Phi$, where $\boldsymbol{\varphi}_h$ and ϕ are the horizontal and vertical parts of $\boldsymbol{\varphi}$.

2.6. *Potential vorticity*

For the variable representing the balance, we choose the potential vorticity (PV). The dimensionless PV is given by

$$\Pi \equiv \tilde{\boldsymbol{\omega}}^a \cdot \nabla Z, \tag{2.14}$$

where $\boldsymbol{\omega}^a$ denotes the absolute vorticity (including the background component $f\boldsymbol{k}$) and $Z(\boldsymbol{x}, t) = z - \mathscr{D}(\boldsymbol{x}, t)$ is the reference height of an isopycnal, or simply the density in disguise since

$$-N^2 Z(\boldsymbol{x}, t) = g\left[\alpha_0 \rho(\boldsymbol{x}, t) - 1\right]. \tag{2.15}$$

Hence we may expand (2.14) to give

$$\Pi = (\boldsymbol{k} + \tilde{\boldsymbol{\omega}}) \cdot (\boldsymbol{k} - \nabla \mathscr{D}) = 1 + \tilde{\zeta} - \mathscr{D}_z - \tilde{\boldsymbol{\omega}} \cdot \nabla \mathscr{D}. \tag{2.16}$$

We next express the PV completely in terms of the vector potential $\boldsymbol{\varphi} = \boldsymbol{\varphi}_h + \phi\boldsymbol{k}$ introduced above. From (2.11) and (2.12) we have

$$\tilde{\boldsymbol{\omega}} = \mathscr{A} - c^2 \nabla \mathscr{D} = \nabla^2 \boldsymbol{\varphi} - \nabla(\nabla \cdot \boldsymbol{\varphi}). \tag{2.17}$$

Hence, the dimensionless PV anomaly $\varpi \equiv \Pi - 1$ may be written in terms of $\boldsymbol{\varphi}$ as

$$\varpi = \mathscr{L}_{qg}\{\phi\} - (1 - \epsilon^2)\nabla_h \cdot \boldsymbol{\varphi}_{hz} + \epsilon^2[\nabla^2\boldsymbol{\varphi} - \nabla(\nabla \cdot \boldsymbol{\varphi})] \cdot \nabla(\nabla \cdot \boldsymbol{\varphi}), \qquad (2.18)$$

where the linear operator $\mathscr{L}_{qg} = \nabla_h^2 + \epsilon^2\partial_{zz}$ is the 'quasi-geostrophic' Laplacian operator.

2.7. *Inversion*

We intend to use $\mathscr{A}_h$ and ϖ as our prognostic variables, to make explicit use of PV conservation $\dot{\varpi} = 0$. But to do so we need to be able to recover the vector potential $\boldsymbol{\varphi}$, and thereby the primitive variables $\boldsymbol{u}$ and $b = -N^2\mathscr{D}$, from $\mathscr{A}_h$ and ϖ. The horizontal components of $\boldsymbol{\varphi}$ are found simply by inverting Laplace's operator:

$$\boldsymbol{\varphi}_h = \nabla^{-2}\mathscr{A}_h. \qquad (2.19)$$

On the other hand, the vertical component of $\boldsymbol{\varphi}$, namely ϕ, must be found from the definition of PV itself (2.18). This results in a double Monge–Ampère equation (the sum of a Monge–Ampère equation in xz and another in yz):

$$I_e\left(\phi_{zz}\nabla_h^2\phi - \phi_{xz}^2 - \phi_{yz}^2\right) + I_a\nabla_h^2\phi + 2I_b\phi_{xz} + 2I_b'\phi_{yz} + I_c\phi_{zz} + I_d = 0, \qquad (2.20)$$

a quadratically nonlinear equation having (space and time varying) coefficients

$$I_a = 1 + \epsilon^2\Theta_z, \quad I_c = \epsilon^2(1 - \Theta_z), \quad I_e = \epsilon^2, \qquad (2.21a,b,c)$$

$$I_b = \epsilon^2\left(\tfrac{1}{2}\nabla^2\varphi - \Theta_x\right), \quad I_b' = \epsilon^2\left(\tfrac{1}{2}\nabla^2\psi - \Theta_y\right), \qquad (2.21d,e)$$

$$I_d = \epsilon^2[\Theta_x\nabla^2\varphi + \Theta_y\nabla^2\psi - |\nabla\Theta|^2] - (1 - \epsilon^2)\Theta_z - \varpi, \qquad (2.21f)$$

where $\Theta \equiv \nabla_h \cdot \boldsymbol{\varphi}_h = \varphi_x + \psi_y$. A numerical procedure to solve this equation is outlined in the Appendix.

2.8. *Prognostic equations*

We next derive the prognostic equations satisfied by the imbalance variables $\mathscr{A}_h$. From (2.11), it is apparent that the evolution of $\mathscr{A}_h$ can be obtained from the vorticity equation (the curl of (2.3), with $-\alpha_0 g\rho' = b = -N^2\mathscr{D}$)

$$\dot{\boldsymbol{\omega}} = \tilde{\boldsymbol{\omega}} \cdot \nabla\boldsymbol{u} + \boldsymbol{u}_z + fc^2\boldsymbol{k} \times \nabla_h\mathscr{D}, \qquad (2.22)$$

combined with the gradient of $\dot{\mathscr{D}} = w$, i.e.

$$\frac{\mathrm{d}\nabla\mathscr{D}}{\mathrm{d}t} = \nabla w - \nabla\boldsymbol{u} \cdot \nabla\mathscr{D}. \qquad (2.23)$$

The result is

$$\dot{\mathscr{A}}_h = -f\boldsymbol{k} \times \mathscr{A}_h + (1 - c^2)\nabla_h w + \tilde{\boldsymbol{\omega}} \cdot \nabla\boldsymbol{u}_h + c^2\nabla_h\boldsymbol{u} \cdot \nabla\mathscr{D}, \qquad (2.24)$$

in which we have used the kinematic relation $\boldsymbol{u}_z = \nabla w - \boldsymbol{k} \times \boldsymbol{\omega}$ to express the group of terms $\boldsymbol{u}_z + fc^2\boldsymbol{k} \times \nabla\mathscr{D}$ in (2.22) as $\nabla w - f\boldsymbol{k} \times (\tilde{\boldsymbol{\omega}} - c^2\nabla\mathscr{D}) = \nabla w - f\boldsymbol{k} \times \mathscr{A}$.

The third prognostic equation is simply

$$\dot{\varpi} = 0, \qquad (2.25)$$

the conservation of PV.

3. Results

3.1. *Problem set-up*

Several numerical simulations are presented next using the Contour-Advective Semi-Lagrangian (CASL) algorithm outlined in the Appendix. Other simulations may be found in Viúdez & Dritschel (2003), who specifically study the vertical velocity field and how well it compares with that diagnosed from quasi-geostrophic balance for an ellipsoidal vortex. The parameters in each simulation are identical unless otherwise noted. We use a 64^3 grid together with 64 isopycnal surfaces, a frequency ratio $c \equiv N/f = 100$, a domain height $L_Z = 2\pi$, a domain width $L_X = L_Y = cL_Z$, mean buoyancy frequency $N = 2\pi$, a time step $\Delta t = 0.1$, a large scale length L (used for the node distribution on PV contours) equal to the diameter (given below) of the vortex simulated, and a 'surgical scale' δ equal to a twentieth of the horizontal grid scale.

The example presented below consists of a horizontal cylinder of anomalous PV, which induces two counter-flowing jets centred on the horizontal edges of the cylinder. This flow is part of a wide class of geostrophically balanced steady flows having no dependence on one horizontal spatial coordinate, say y, Such flows have $u = w = 0$, but v and $\mathscr{D}$ (or b) non-zero. As such, the PV is steady, with fluid moving in the y-direction only.

For these flows, $v = f\phi_x$ and $\mathscr{D} = -\epsilon^2\phi_z$, where $\phi(x, z)$ satisfies a simplified two-dimensional Monge–Ampère equation, that is (2.20) without y-derivatives, and with $I_a = 1$, $I_b = I_b' = 0$, $I_c = I_e = \epsilon^2$, and $I_d = -\varpi$. In fact, this equation is independent of ϵ if we scale z by ϵ:

$$\phi_{xx} + \phi_{zz} + \phi_{xx}\phi_{zz} - \phi_{xz}^2 = \varpi \tag{3.1}$$

(here z stands for Nz/f). Note that in this case, $\phi = \alpha_0\Phi/f^2$, where Φ is the geopotential in (2.3). We have developed a two-dimensional numerical method that solves this equation, given any distribution of ϖ, in the periodic computational domain. However, it is instructive to consider first an analytical solution $\phi(r)$ which is available to any radial distribution of potential vorticity $\varpi(r)$ in an unbounded fluid. Then, (3.1) reduces to

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}\phi}{\mathrm{d}r}\right) + \frac{1}{r}\frac{\mathrm{d}\phi}{\mathrm{d}r}\frac{\mathrm{d}^2\phi}{\mathrm{d}r^2} = \varpi \tag{3.2}$$

– in fact a first-order equation for $\mathrm{d}\phi/\mathrm{d}r$. This can be integrated directly for any distribution $\varpi(r)$ with the result

$$\frac{1}{r}\frac{\mathrm{d}\phi}{\mathrm{d}r} = \pm\sqrt{\tau(r)} - 1, \tag{3.3}$$

where

$$\begin{aligned}\tau &= 1 + \frac{2}{r^2}\int_a^r \varpi(r')r'\,\mathrm{d}r' \\ &= \frac{2}{r^2}\int_a^r \Pi(r')r'\,\mathrm{d}r', \end{aligned} \tag{3.4}$$

and where a is a constant of integration chosen to ensure $\tau(r) \geqslant 0$ for all r. For $\Pi \geqslant 0$ everywhere, $a = 0$, and the positive root in (3.3) ensures that the solution is both statically and inertially stable (the other root corresponds to an unstable solution). If $\Pi < 0$ over any range of r, then one may show that there are always angular sectors of static and inertial instability, whichever root is taken in (3.3). The integration

constant a must then be chosen so that the minimum of $K(r) = \int_0^r \Pi(r')r'\mathrm{d}r'$ occurs at $r = a$ (since then $\tau = 2(K(r) - K(a))/r^2 \geqslant 0$). As our focus is on statically stable flows, we will not consider this situation further.

Notably, for any compact distribution of $\varpi(r)$ or any sufficiently rapidly decaying distribution, $\mathrm{d}\phi/\mathrm{d}r \to \kappa/r$ as $r \to \infty$, where $\kappa = \int_a^\infty \varpi(r')r'\mathrm{d}r'$. That is, at large distances, the nonlinear terms in the Monge–Ampère equation are negligible compared to the linear terms.

For the special case of a uniform PV cylinder, $\varpi = \varpi_0$ for $r < r_e$ and $\varpi = 0$ for $r \geqslant r_e$, we have

$$\tau = \begin{cases} 1 + \varpi_0 & \text{for } r < r_e & (3.5a) \\ 1 + \varpi_0 r_e^2/r^2 & \text{for } r \geqslant r_e & (3.5b) \end{cases}$$

and in general τ is monotonic if ϖ is. The tangential velocity and displacement (also scaled by ϵ) are given by $v/f = xr^{-1}\mathrm{d}\phi/\mathrm{d}r$ and $\mathscr{D} = -zr^{-1}\mathrm{d}\phi/\mathrm{d}r$; in particular v is linear in x while $\mathscr{D}$ is linear in z within the cylinder. In the limit $1 + \varpi_0 \to 0$, the material lines $Z(x, z) = z - \mathscr{D}(x, z)$ and $X(x, z) = x + v(x, z)/f$ are expelled from the cylinder interior, leaving neutral stratification and neutral inertial stability there (Rellich's parameter $R = 0$ in this case as well, see the Appendix). The flow overall consists of two counter-propagating jets centred on $x = \pm r_e$, $z = 0$. Far from the cylinder, v and $\mathscr{D}$ decay like r^{-1}.

For flow initialization, the above solution is not suitable since the domain simulated is a periodic box of finite dimensions. The numerical solutions found by solving (3.1) in the periodic domain however differ little from the above when the size of the cylinder is sufficiently small compared to the domain dimensions, as in the examples discussed below. There is a further problem in using these two-dimensional numerical solutions, since any initial disturbance (required to allow three-dimensional evolution) will be initially unbalanced and will hence adjust by gravity wave emission. To reduce this effect, we instead follow the initialization procedure described in the Appendix, and start with a disturbed PV cylinder having initially zero PV anomaly. The anomaly is then artificially grown over 5 inertial periods to the desired final value by growing the jumps in ϖ across the PV contours.

The initial contours lie on flat isopycnal surfaces ($\mathscr{D} = 0$), and are chosen to represent the circular distribution of anomalous potential vorticity $\varpi(r) = \varpi_0(1 - r^2/r_e^2)$, where ϖ_0 is the maximum anomaly after the period of initialization, and here $r^2 = (x - x_0(y))^2 + z^2$, where $x_0(y)$ is the centreline displacement (in x) used to impart a transverse disturbance to the cylinder. Specifically, we take $x_0(y) = a_m \sin(2\pi m y/L_Y) + a_n \sin(2\pi n y/L_Y)$, with $m = 2$, $n = 3$, $a_m = 0.1$ and $a_n = -0.1$. The corresponding PV contours are curves of constant $x - x_0$ on each isopycnal $z = z_\ell$ (initially coincident with grid planes, i.e. $z_\ell = \bar{z}_\ell$). The values of $\hat{x} = x - x_0$ are determined by discretizing the distribution of $\varpi(r)$ on $z = z_\ell$ into a series of equal jumps $\Delta\varpi = \varpi_0/n$, with n an integer, except for the outermost jumps, which use $\Delta\varpi = \varpi_0/2n$. For a quadratic function of r, the optimal discretization is obtained by choosing the discrete radii as $r_k = r_e\sqrt{1 - (k/n)^2}$, with $k = 0, 1, \ldots, n-1$ (Dritschel 1998). This ensures that the circulation contained within $r = r_k$ is the same in the discrete and continuous profiles. From the r_k, the contour positions $\hat{x}_{k\ell}$ on each isopycnal surface are determined from $\hat{x}_{k\ell} = \sqrt{r_k^2 - z_\ell^2}$, for all $k \geqslant 0$ for which $|z_\ell| < r_k$. In general, the number of contours on each isopycnal surfaces varies, and there are no contours on surfaces for which $|z_\ell| \geqslant r_e$. Finally, the same discretization procedure can be used for an elliptical cylinder of $x : z$ aspect ratio λ if we multiply
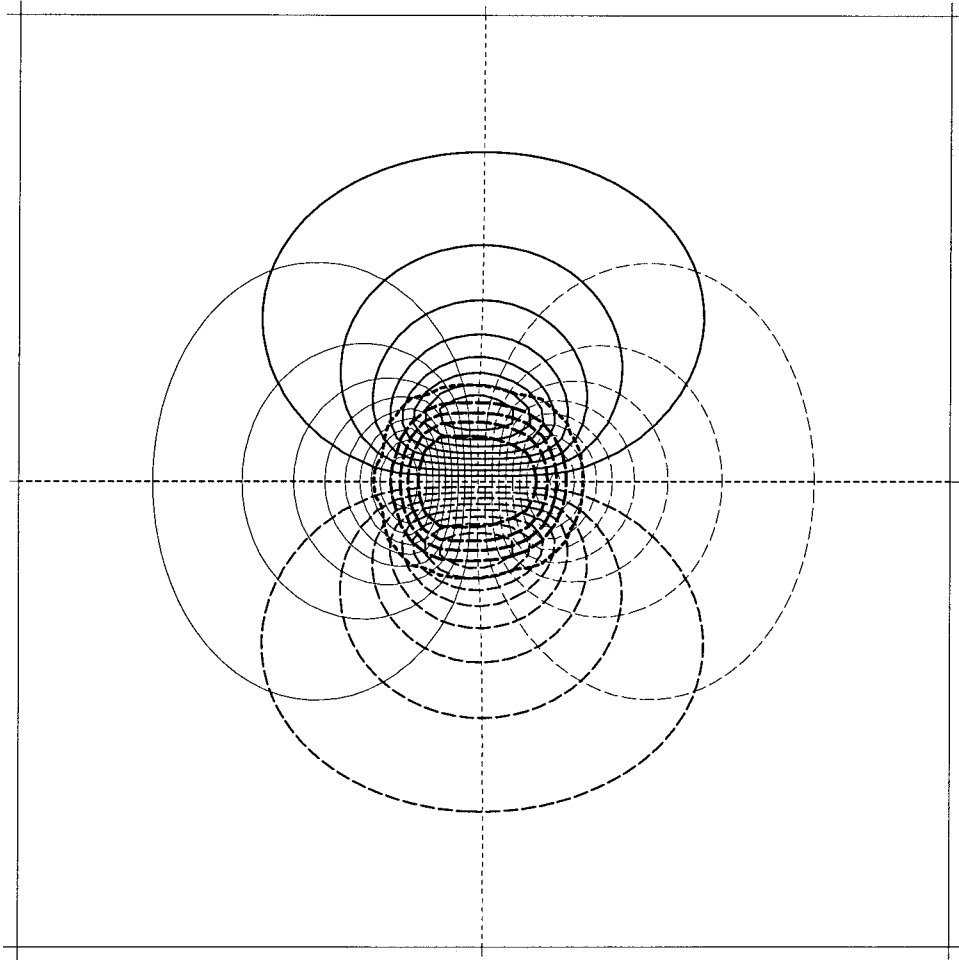
FIGURE 1. Vertical distribution at $y = 0$ (grid point $i_Y = 32$) and at $t = 5T_{ip}$ of $\mathscr{D}$ (medium thick lines, contour interval $\Delta = 10^{-2}$, extrema contour is ± 0.08), of v/f (thin lines, contour interval $\Delta = 1$, extrema contour is ± 9), and gridded PV contours (thick lines, $\Delta = 0.1$, minimum contour is -0.4). In this and similar figures the short-dashed horizontal and vertical lines in the plot are the zero contour lines of $\mathscr{D}$ and v/f, respectively; and dashed lines show negative values. Domain extent is $10^2 [-\pi, \pi] \times [-\pi, \pi]$.

the $\hat{x}_{k\ell}$ above by λ. (The circular cylinder here is in fact an ellipse of aspect ratio ϵ in the original coordinates.)

3.2. *Flow evolution*

The simulation presented next uses $n = 10$ PV levels and $\varpi_0 = -0.5$, corresponding to a moderately ageostrophic anticyclonic shear zone (other values of ϖ_0 have also been examined, and are compared below). The corresponding PV field ϖ for a cylinder of radius $r_e = 0.5c$, at the end of the initialization period ($t = 5T_{ip}$), is shown in figure 1, together with the distributions of v/f and $\mathscr{D}$, in the cross-section plane $y = 0$ (here and below z and $\mathscr{D}$ are scaled by $c = N/f$). Note that the PV contours displayed are found from the interpolated gridded field; the contours being explicitly advected lie on isopycnal surfaces oriented nearly perpendicular to the plane shown. The distribution of v/f corresponds to a double jet, with maxima located very near

the lateral edges of the PV cylinder. The distribution of $\mathscr{D}$ is similar to that of v/f, except rotated by $90°$ – isopycnals are displaced upwards above the centre of the cylinder and downwards below it. This pattern is consistent with near geostrophic balance.

At this time the Rossby number Ro (the extreme value of the dimensionless vertical vorticity $\tilde{\zeta}$) is approximately -0.34, while the Froude number Fr (the maximum value of $|\boldsymbol{\omega}_h|/\sqrt{N^2 + b_z}$) is approximately 0.16. In all cases examined here, $|Fr/Ro| \approx \frac{1}{2}$, and this ratio appears to be mainly controlled by the aspect ratio λ of the PV cylinder (here we have used $\lambda = 1$, in the rescaled coordinates). In general, we would expect $|Fr/Ro| \approx (2\lambda)^{-1}$. From the exact solution for a uniform PV anomaly, (3.3) and (3.5), one may show that $Ro = (1 + \varpi_0)^{-1/2} - 1$, and $\tilde{\zeta} = Ro$ on the edge of the cylinder at $x = \pm r_e$ and $z = 0$ (note that the sign of Ro is opposite to that of ϖ_0 in this case). For the parabolic profile used here, the extreme value of $\tilde{\zeta}$ occurs at the centre of the cylinder and $Ro = (1 + \varpi_0)^{1/2} - 1$ (this has the same sign as ϖ_0). For $\varpi_0 \ll 1$, $Ro \approx \frac{1}{2}\varpi_0$. Similarly, one may show that for both a uniform and a parabolic PV profile, the Froude number $Fr = |\Lambda - 1/\Lambda|$, where $\Lambda = [\tau(r_e)]^{1/4}$, and this is achieved on the outside edge of the cylinder at $x = r_e \cos\theta_c$, $z = r_e \sin\theta_c$, where $\tan\theta_c = \pm\Lambda$. Note that $\tau(r_e) = 1 + \varpi_0$ for a patch and $\tau(r_e) = 1 + \varpi_0/2$ for a parabolic PV distribution. For $\varpi_0 \ll 1$, $Fr \approx \frac{1}{4}|\varpi_0|$ for the parabolic distribution used here. These results are summarized in figure 2 for a parabolic PV distribution.

The PV evolution from $t = 0$ to $30T_{ip}$ on the central isopycnal $\ell = 32$ is illustrated in figure 3, along with a corresponding three-dimensional view in figure 4. As in two-dimensional barotropic flows (Dritschel 1989a), the PV rolls up into a street of vortices, here all anticyclonic, which subsequently merge into larger vortices (the domain periodicity, however, limits how far this can go). The flow remains strongly nonlinear, as seen from figure 5, which displays the time evolution of Ro_{min}, Ro_{max}, and Fr for this and two other simulations (for $\varpi_0 = -0.25$ and for $\varpi_0 = -0.75$). The vortices which first form exhibit a height-to-width aspect ratio of order unity (in scaled height coordinates), a value similar to that found in quasi-geostrophic turbulence (Reinaud, Dritschel & Koudella 2003), and consistent with the nearly constant ratio of $|Ro_{min}|/Fr$ exhibited in figure 5. On the basis of a wide range of simulations conducted, we believe that this similarity is not accidental – it indicates that the dynamics of nearly balanced rotating stratified flows at moderate Rossby numbers is qualitatively similar to the dynamics of quasi-geostrophic flows. Even quantitatively, the differences are often minor. In particular, the vertical velocity w differs by at most 10–20% from the vertical velocity w_{qg} diagnosed from quasi-geostrophic balance – see e.g. figure 6 for a comparison of w and w_{qg} at $t = 12T_{ip}$ on the horizontal plane $i_Z = 28$ where w is approximately largest, and figures 7 and 8 for comparisons in the (x, z)- and (y, z)-planes, respectively (for details on the computation of w_{qg}, see Viúdez & Dritschel 2003, who also found similar results for an ellipsoidal gyre). The vertical velocity pattern has an octupole structure associated with the ellipsoidal shape of the developing vortices. The vortex forming on the sides of the domain is more elongated than the one in the centre, and it has much larger vertical velocities.

The similarity between w and w_{qg} is quantified by fitting the linear relationship $w_{qg}(\boldsymbol{x}, t) = q(t) + s(t)w(\boldsymbol{x}, t)$ over all grid points at each time t. The slope $s(t)$, together with its probable uncertainty $\sigma_s(t)$, are plotted in figure 9 as a function of time for the simulation just described as well as a variety of others differing only in the PV anomaly ϖ_0 (the offset $q(t)$ is always found to be negligible). We find that $s(t) < 1$ generally, indicating that w_{qg} underestimates the true vertical velocity w associated
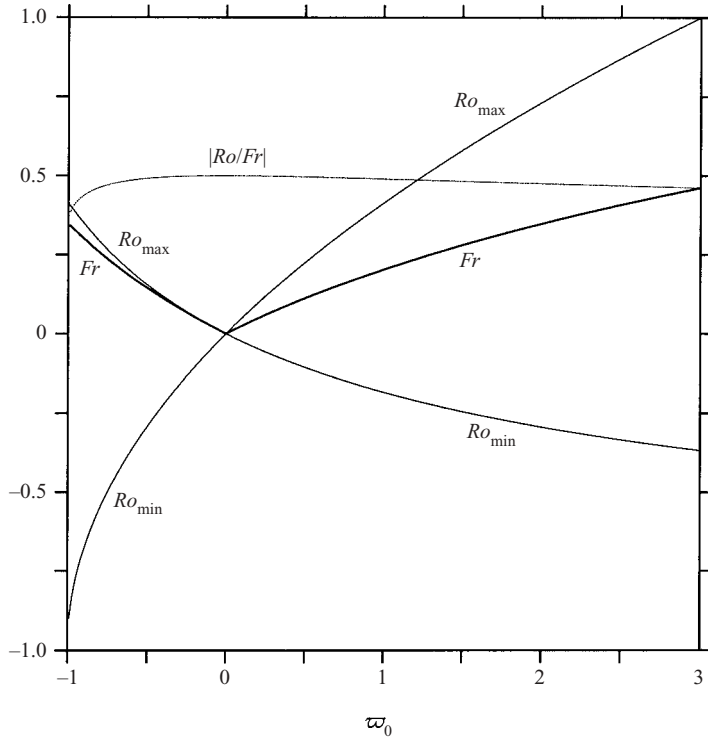
FIGURE 2. Minimum (most negative) and maximum (most positive) Rossby numbers Ro_{min}, Ro_{max} and Froude number Fr for the class of exact radial parabolic PV distributions, $\varpi(r) = \varpi_0(1 - r^2/r_e^2)$ for $r < r_e$ and $\varpi(r) = 0$ for $r \geqslant r_e$, as a function of the PV extremum ϖ_0. The overall Rossby number Ro is defined to be largest, in magnitude, of the two Rossby numbers, and it corresponds to the curve beginning at $(-1, -1)$, crossing $Ro = 0$ at $\varpi_0 = 0$, and extending to the upper right corner of the figure. This value of Ro is found at $r = 0$ and takes the form $Ro = Ro_1 = (1 + \varpi_0)^{1/2} - 1$. The second, lower-magnitude Rossby number is found at $x = \pm r_e$, $z = 0$ and takes the form $Ro = Ro_2 = (1 + \frac{1}{2}\varpi_0)^{-1/2} - 1$.

with negative PV anomalies ($s(t) > 1$ is found for positive PV anomalies, see Viúdez & Dritschel 2003 for further remarks). Moreover, on average s decreases with increasing magnitude of the PV anomaly, whereas σ_s increases with it. The increase in σ_s with $|\varpi_0|$ is not principally due to a breakdown in the linear relationship, but rather to an increased scatter about it. This scatter is associated with the presence of inertia–gravity waves in the field of w. Note however that even for $\varpi_0 = -0.75$ the level of scatter measured by σ_s is only about 0.1% of s.

3.3. *Inertia–gravity waves*

Balanced motions are clearly dominant in these examples. To obtain a better view of the wave motions, we would like to separate them from the balanced motions, but this is not straightforward. The separation depends on what balance is assumed – clearly quasi-geostrophic balance is inadequate at these Rossby numbers, since the difference $w - w_{qg}$ is still highly correlated with the PV field. No attempt has yet been made to apply higher-order balance relations, e.g. cyclo-geostrophic (that is, including the centripetal acceleration in the horizontal momentum balance), and instead we present an alternative way of assessing the importance of inertia–gravity waves.
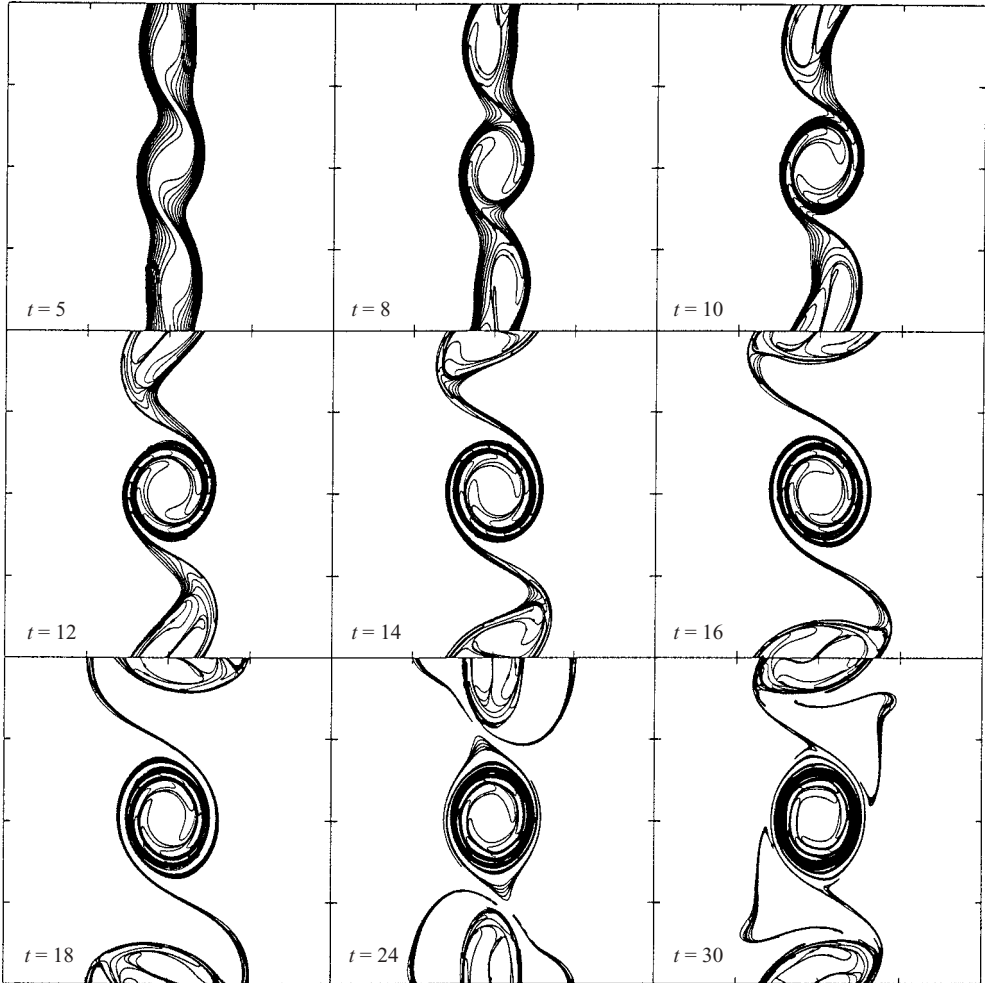
FIGURE 3. Time evolution of the PV contours lying on the middle isopycnal surface $\ell = 32$. Times are indicated in inertial periods. The points on the contours correspond to fluid particles initially lying on the x-axis.

Since such waves have frequencies ω lying in the range $f \lesssim \omega \lesssim N$ (the uncertainties being due to variations in both the local stratification (b) and the local rotation (ζ)), their importance may be estimated by the power within this frequency band compared with that at zero frequency (the time average). Here, power is measured in two ways: by the kinetic-energy frequency spectrum and by the squared vertical-velocity frequency spectrum, averaged over a coarse set of grid points. The points chosen lie on a regular $8 \times 8 \times 8$ grid, and for these points the kinetic energy density and vertical velocity are recorded each time step. The resulting frequency spectra are shown in figure 10 for the present simulation, and for two others having 50% smaller and 50% larger PV anomalies ($\varpi_0 = -0.25$ and -0.75), as labelled. Only the spectrum of w^2 shows a significant level of wave activity, predominantly in the form of near-inertial waves (frequencies close to f), but even here the power within the range $f \lesssim \omega \lesssim N$ is in all cases less than 1% of the total, and diminishing with decreasing ϖ_0, consistent with the spatial patterns of w shown previously in figures 6,
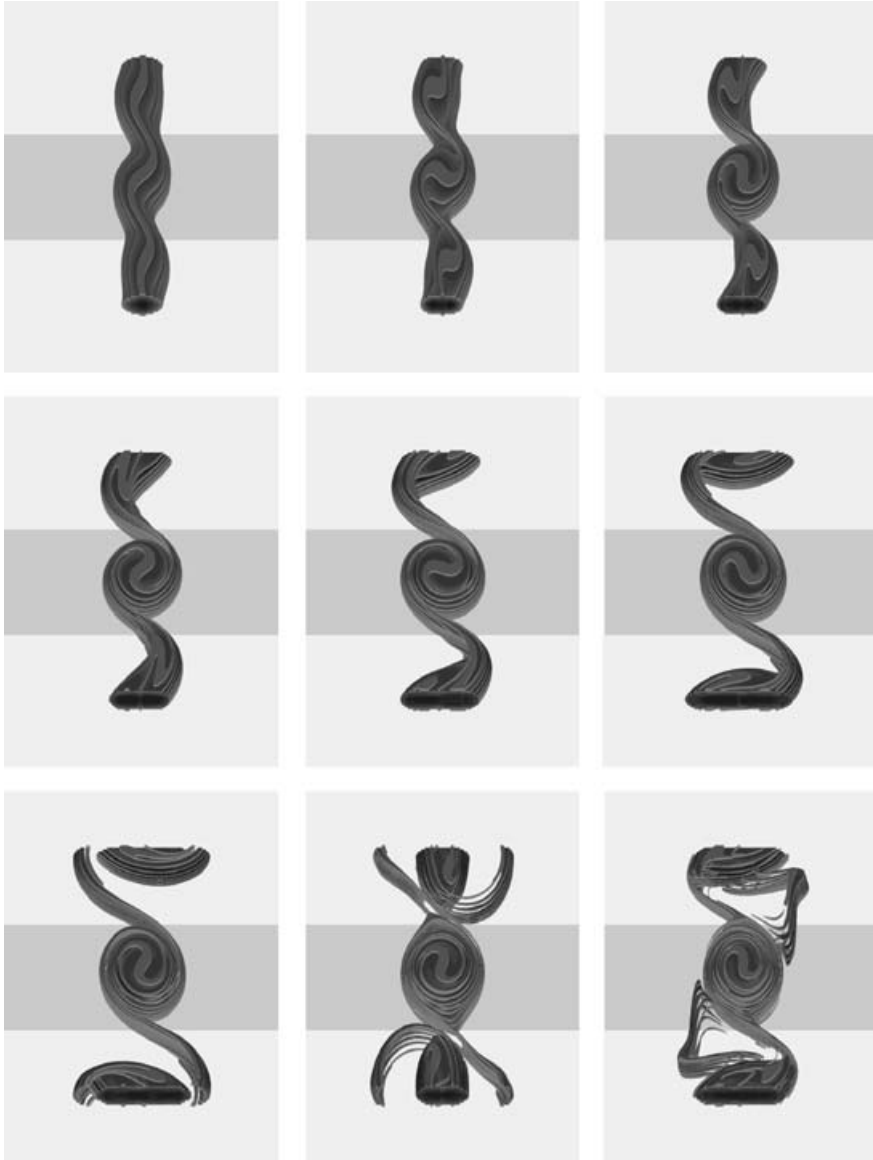
FIGURE 4. Three-dimensional evolution of the PV initially contained within the disturbed cylinder, at $t = 5$, 8, 10, 12, 14, 16, 18, 24 and $30T_{ip}$ (left to right, then downwards) – the same times shown in figure 3. The height coordinate is here stretched by $c = N/f = 100$. An orthographic perspective is used, from a point lying in the (y, z)-plane at an angle of $30°$ from the vertical. Only the vertical range containing anomalous PV is shown. The front and back faces of the domain are rendered light grey while the visible bottom is rendered a slightly darker shade of grey.

7 and 8 for $\varpi_0 = -0.5$. Most notably, these waves are virtually undetectable in the kinetic-energy frequency spectrum. Hence, PV advection is accomplished by a nearly balanced flow field. For all practical purposes, these flows are in a state of balance.

It is worth making the point that the kinetic-energy spectrum extends over all frequencies. There is no clear distinction between 'slow' and 'fast', except that the slow
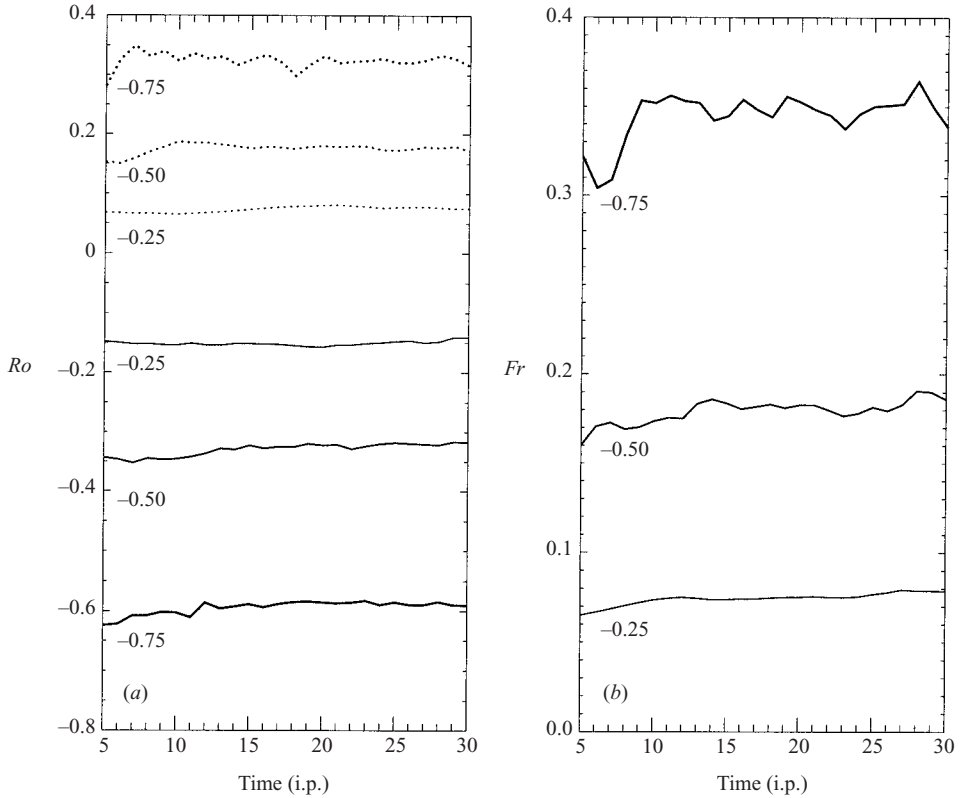
FIGURE 5. The time evolution of (*a*) the minimum and maximum Rossby number and (*b*) the Froude number, for three simulations differing only in the minimum PV anomaly ϖ_0, as labelled.
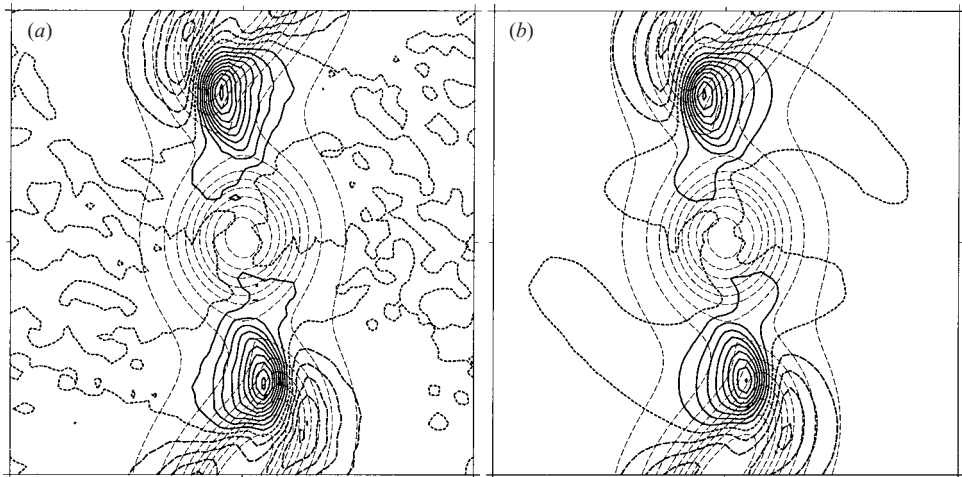

FIGURE 6. (*a*) Horizontal distribution of the vertical velocity w (thick lines, $\Delta = 10^{-5}$) and vertical displacement $\mathscr{D}$ (thin lines, $\Delta = 10^{-2}$) at $z = -\pi/8$ (grid point $i_Z = 28$) at $t = 12T_{\mathrm{ip}}$. Note that the vertical displacement is entirely negative, so that there are no thin solid lines. (*b*) Same as in (*a*) except the quasi-geostrophic vertical velocity w_{qg} is shown.
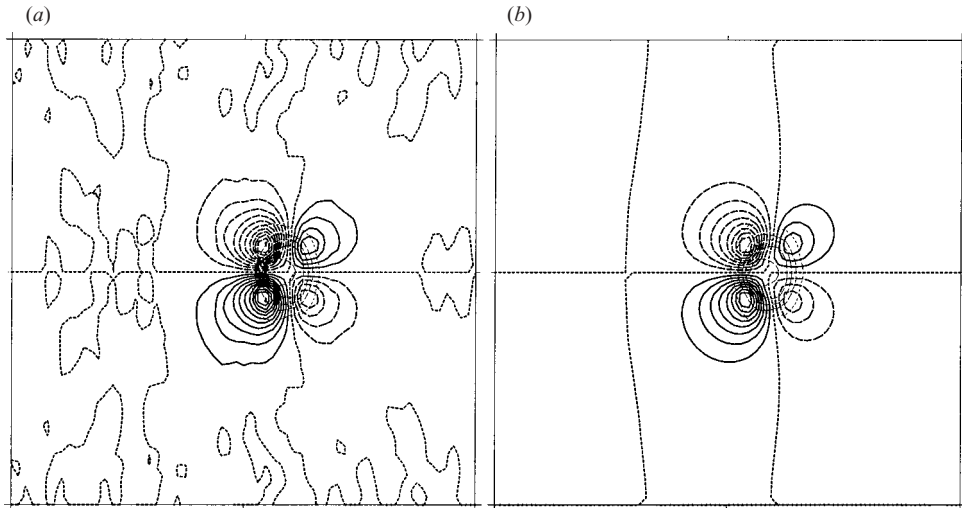
FIGURE 7. (*a*) Vertical distribution of the vertical velocity w (thick lines, $\Delta = 10^{-5}$) and gridded PV contours (thin lines, $\Delta = 0.1$) at $y = -5c\pi/8$ (grid point $i_Y = 12$) and at $t = 12T_{\mathrm{ip}}$. (*b*) Same as in (*a*) except the quasi-geostrophic vertical velocity w_{qg} is shown.
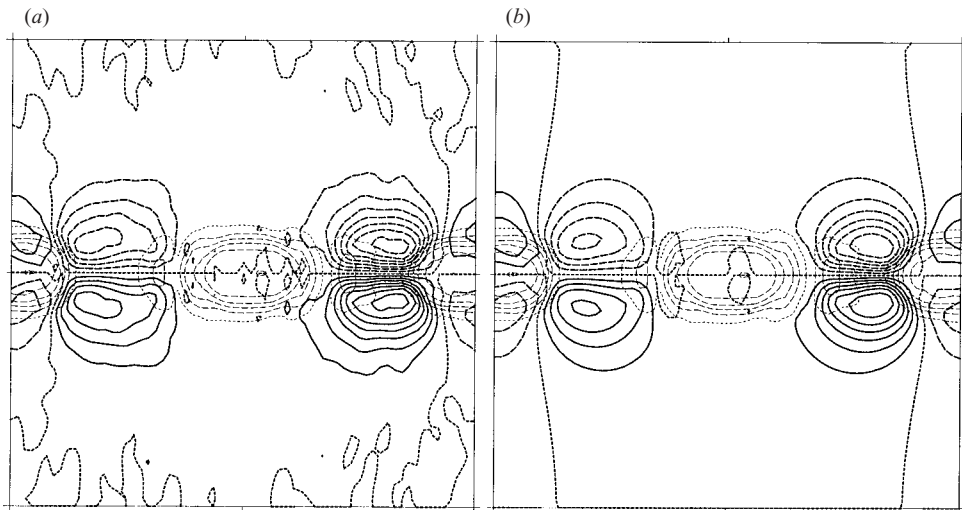

FIGURE 8. (*a*) Vertical distribution of the vertical velocity w (thick lines, $\Delta = 10^{-5}$) and gridded PV contours (thin lines, $\Delta = 0.1$) at $x = 0$ (grid point $i_X = 32$) and at $t = 12T_{\mathrm{ip}}$. (*b*) Same as in (*a*) except the quasi-geostrophic vertical velocity w_{qg} is shown.

motions clearly have more power. The frequencies $\omega > f$ here do not correspond to inertia–gravity waves, but rather to the balanced dynamics, i.e. to the part of the flow controlled by PV (this part is clearly more than that given by thermal-wind balance). The balanced motions do involve high frequencies, but they project remarkably weakly on inertia–gravity waves. What we have found in this example, and many others conducted, is that there is little significant inertia–gravity wave generation for Rossby numbers up to about 0.6 (in absolute value), and Froude numbers up to about 0.3, in flows starting from a nearly balanced state.
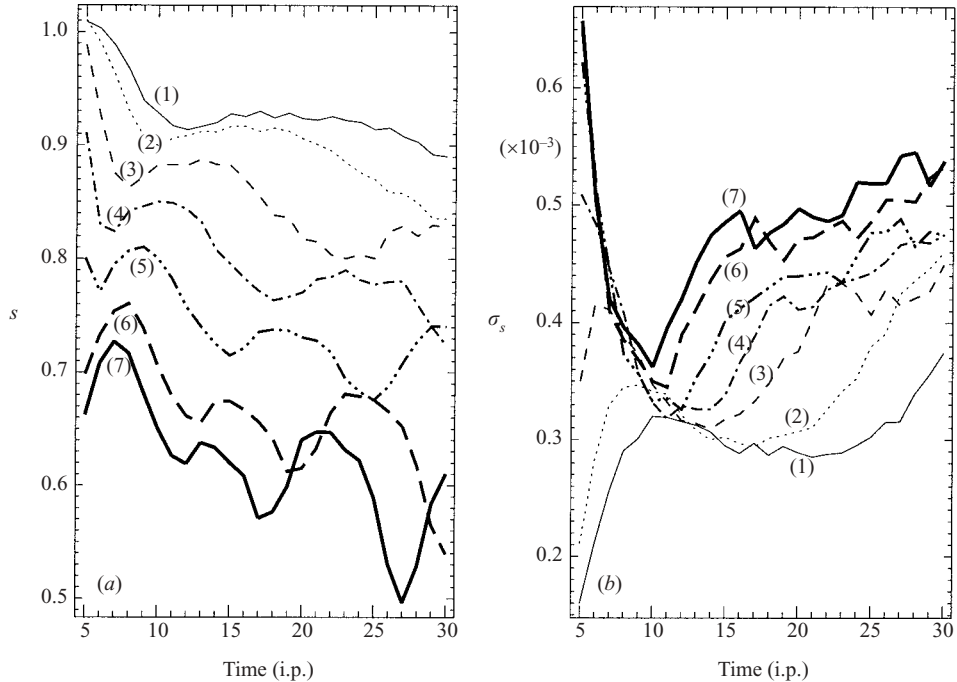
FIGURE 9. (*a*) The slope $s(t)$ and (*b*) the probable uncertainty $\sigma_s(t)$ in the estimate of s in the linear regression $w_{qg}(\boldsymbol{x}, t) = q(t) + s(t)w(\boldsymbol{x}, t)$, as a function of time and for different values of $\varpi_{max} = -0.25(1), -0.3(2), -0.4(3), -0.5(4), -0.6(5), -0.7(6)$ and $-0.75(7)$. The linear fit is computed, once each inertial period, using the w and w_{qg} data in the complete domain (64^3 data).

3.4. *Accuracy*

A few comments on the accuracy of this numerical solution are in order. The original grid, 64^3, may seem rather coarse for simulating geophysical flows at large Reynolds numbers. However, in previous work in the related shallow-water context (Dritschel, Polvani & Mohebalhojeh 1999), it was found that the CASL algorithm converged much faster than commonly used algorithms (such as pseudo-spectral and semi-Lagrangian), so that solutions on relatively coarse grids (such as used here) approximate well the solutions on much higher resolution grids.

Similar results have been found in the present context. We have computed the flow evolution, starting from the end of the initialization period ($t = 5T_{ip}$), using also a pseudo-spectral algorithm (described in the Appendix) for the basic example with $\varpi_0 = -0.5$. The vertical velocity and displacement fields for this experiment, at $t = 12T_{ip}$ on the horizontal plane $i_Z = 28$, are shown in figure 11, which is identical in format with figure 6 (and uses the same contour interval). Most notably, the vertical velocity in the pseudo-spectral simulation is only half of that in the CASL simulation. Moreover, the displacement field has weaker gradients and a broader scale overall. The quasi-geostrophic vertical velocity (*b*), however, compares with the actual vertical velocity (*a*), which indicates that the source of the problem is the PV field.

The PV field is not explicitly conserved in the pseudo-spectral algorithm, and the result is strong diffusion, as shown in figure 12(*a*) at $t = 12T_{ip}$ and $24T_{ip}$ in the middle isopycnal surface. (This level of diffusion is closely comparable to what was found previously in the shallow-water context at this resolution, see Dritschel
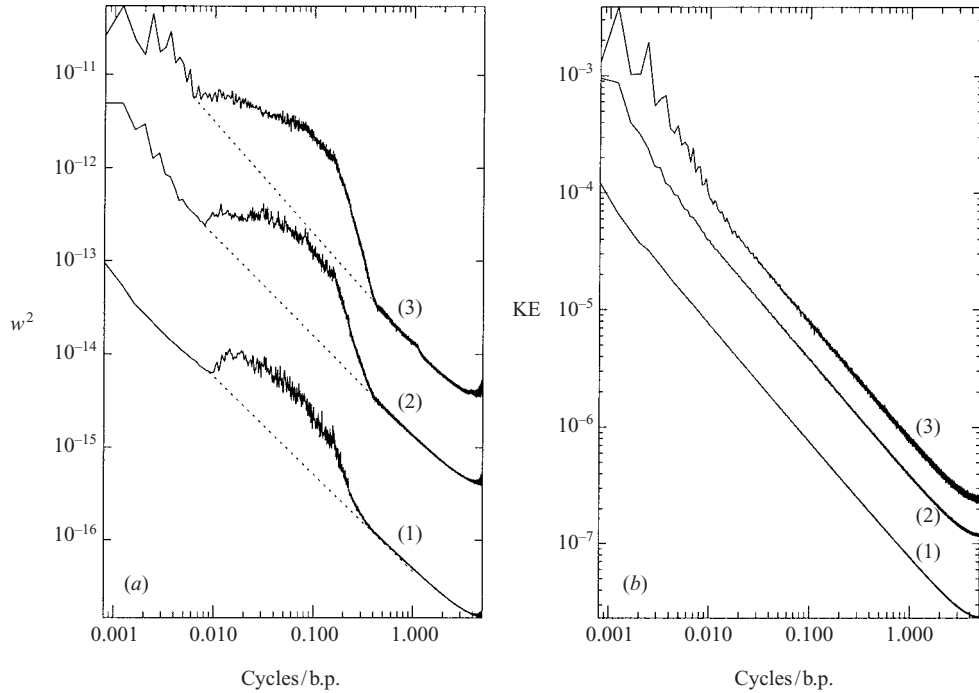
FIGURE 10. Averaged frequency spectra for (*a*) w^2 and (*b*) (twice) the specific kinetic energy $(u^2 + v^2 + w^2)$, and for the cases $\varpi_{min} = -0.25$ (1), -0.5 (2), and -0.75 (3). Horizontal axis is frequency in cycles/(buoyancy period). Every spectra is the average of 8^3 spectra each computed from the time series from $t = 5$ to $t = 30T_{ip}$ (thus comprising 25000×8^3 data points). In (*a*), an additional dotted line is plotted linearly interpolating between the spectral values at f and N.
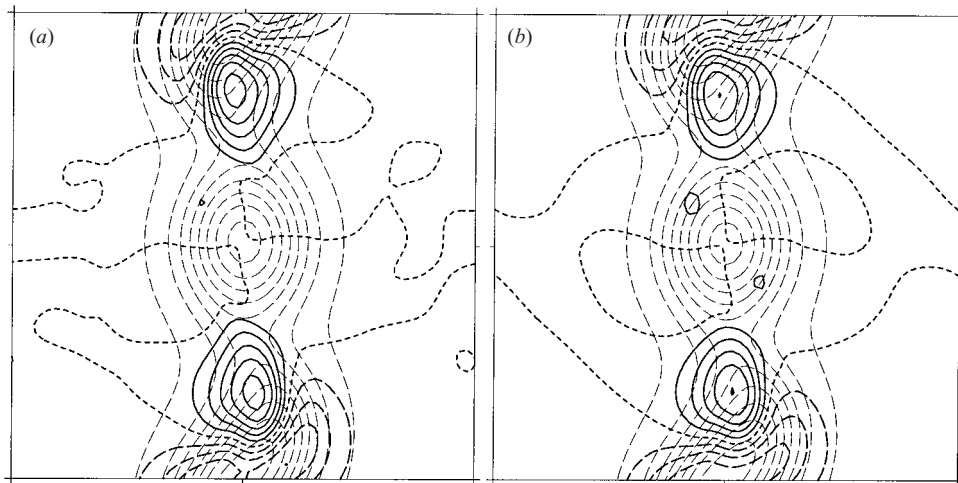

FIGURE 11. (*a*) Horizontal distribution of the vertical velocity w (thick lines, $\Delta = 10^{-5}$) and vertical displacement $\mathscr{D}$ (thin lines, $\Delta = 10^{-2}$) at $z = -\pi/8$ (grid point $i_Z = 28$) at $t = 12T_{ip}$ for the pseudo-spectral simulation having $\varpi_0 = -0.5$. (*b*) Same as in (*a*) except the quasi-geostrophic vertical velocity w_{qg} is shown.
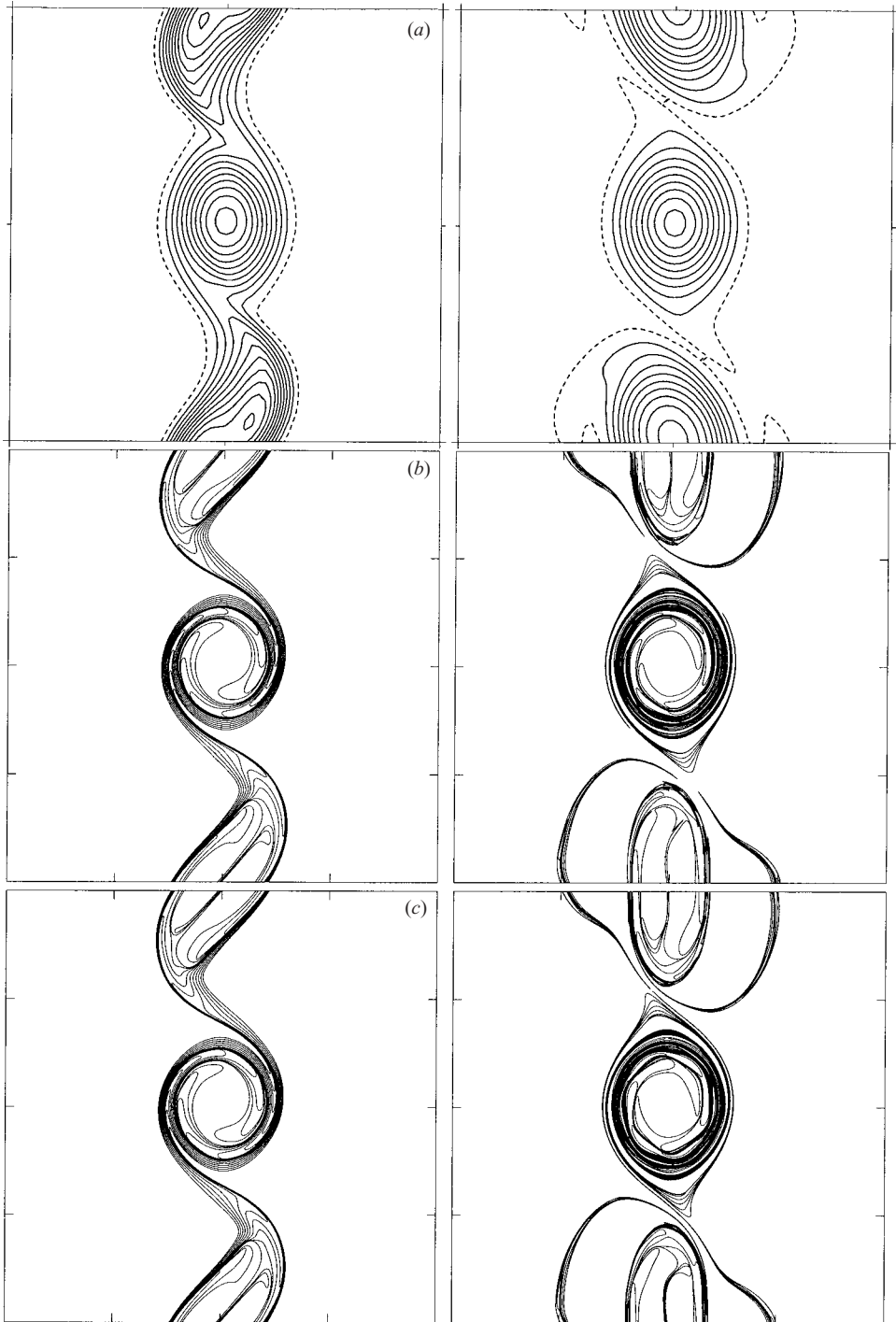
FIGURE 12. Comparison of the PV contours lying on the middle isopycnal surface for the 64^3 pseudo-spectral simulation (*a*), the 64^3 CASL simulation (*b*), and a 128^3 CASL simulation (*c*), at $t = 12T_{\mathrm{ip}}$ (left column) and $24T_{\mathrm{ip}}$ (right column).
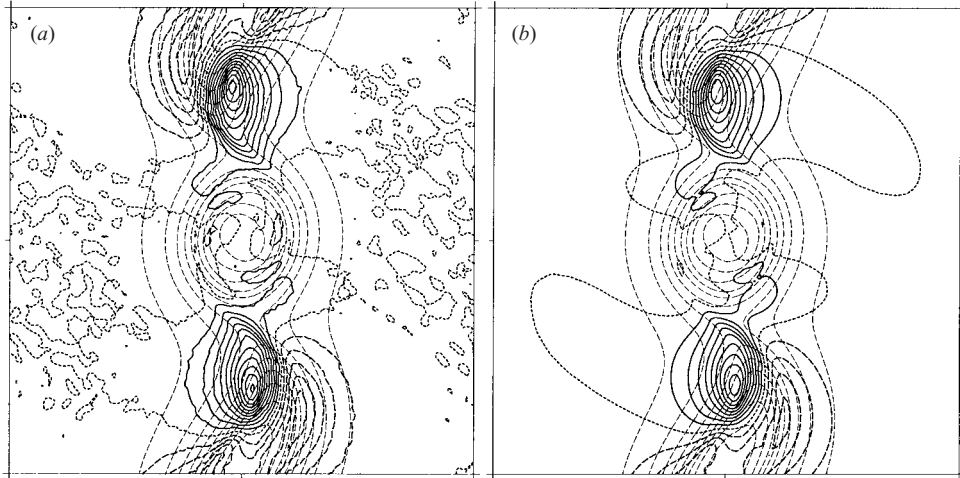
FIGURE 13. (*a*) Horizontal distribution of the vertical velocity w (thick lines, $\Delta = 10^{-5}$) and vertical displacement $\mathscr{D}$ (thin lines, $\Delta = 10^{-2}$) at $z = -\pi/8$ (grid point $i_Z = 56$) at $t = 12T_{\mathrm{ip}}$ for the 128^3 CASL simulation. (*b*) Same as in (*a*) except the quasi-geostrophic vertical velocity w_{qg} is shown. The same fields for the 64^3 simulation are shown in figure 6.

et al. 1999.) We also show for comparison the present CASL simulation results at the basic 64^3 resolution (figure 12*b*) as well as CASL simulation results at 128^3 resolution (figure 12*c*). This figure demonstrates the extremely rapid convergence of the CASL algorithm – 64^3 resolution is more than enough to capture the PV field accurately. On the other hand, the pseudo-spectral simulation results completely lack the fine-scale structure typical of PV advection. This poor representation of the PV field degrades the overall accuracy of the simulation, as we have seen already for the vertical velocity field in figure 11. In the CASL simulations, the vertical velocity field is insensitive to further increases in resolution – compare figure 13 for 128^3 with figure 6 for 64^3.

The general principle that we are stressing, noted previously in the shallow-water context, is that an accurate representation of the PV field is essential for an accurate representation of the dominant, balanced part of the fluid motion in the atmosphere and oceans. This accuracy can only be beneficial to modelling the imbalanced motions also. Here, we have gone a step further by separating balanced and imbalanced motions to first order (in the departure from hydrostatic and geostrophic balance), using appropriate numerical methods for each, and yet fully retaining the coupling between these motions.

4. Concluding remarks

We have developed a theory which distinguishes, at leading order, balanced and imbalanced motions in a rotating stably stratified fluid. The theory makes use of the simplest hydrostatic and geostrophic balance relations to identify an appropriate set of variables expressing the departure from this balance. The important point is that the theory does not impose balance, but rather it is used to help separate distinct kinds of fluid motion: the 'slow' potential-vorticity-controlled balanced motions and the 'fast' inertia–gravity waves. This has been found to be practically useful in a related shallow-water context (Mohebalhojeh & Dritschel 2000, 2001; Dritschel & Mohebalhojeh 2000), and is likely to prove useful in this context as well.

The resulting transformation of the equations of motion, to use potential vorticity together with a pair of variables representing the imbalance (here the ageostrophic horizontal vorticity), results in a series of diagnostic relations to recover the original 'primitive' variables, i.e. the velocity and density. In particular, the velocity and density fields are not explicitly evolved. The diagnostic equations are linear apart from the one associated with the definition of potential vorticity. The latter turns out to be a double Monge–Ampère equation, a quadratically nonlinear equation relating the vertical component of the vector potential to the potential vorticity and imbalance fields. It is shown that properties of this equation are intimately connected with inertial and static stability.

We have developed a new numerical method to solve these transformed equations and applied it to the evolution of a disturbed horizontal cylinder of potential vorticity having initial Rossby numbers ranging from -0.2 to -0.6. The cylinder destabilizes and rolls up into a series of three-dimensional vortices, yet despite the $O(1)$ Rossby numbers present in the flow, there is remarkably little inertia–gravity wave emission.

In future work, we would like to explore the instability of horizontal and vertical potential vorticity cylinders more thoroughly, comparing also with more conventional numerical approaches. In particular, we would like to quantify the level of wave emission as a function of Rossby number and other parameters characterizing the flow (e.g. the ratio f/N and the aspect ratio of the cylinder). This will entail developing 'radiation' (wave-absorbing) boundary conditions. We also intend to include solid boundaries and bottom topography.

Finally, to be able to better distinguish between balanced and imbalanced motions, methods need to be developed, perhaps analogous to the potential-vorticity inversion hierarchies examined by Mohebalhojeh & Dritschel (2000) in the shallow-water context. The results of the first example presented in this paper show clearly that there is a better balance available than the simplest hydrostatic–geostrophic one. For instance, one can improve on geostrophic balance by including the centripetal acceleration in the horizontal momentum balance, and so on. This may not only help to better distinguish balance and imbalance in a post-processing sense, but it may also lead to new more accurate numerical methods for simulating such flows. Indeed, it is sensible that numerical methods respect and exploit the underlying approximate relations between fields – accurate numerics may depend crucially on this often-hidden mathematical structure.

Appendix. The numerical algorithm

An algorithm has been developed for flow within an idealized geometry, namely a triply periodic box of dimensions $L_X \times L_Y \times L_Z$. Extensions to flows with boundaries are currently underway, and the boundary conditions required in such cases are discussed below. Since flows close to geostrophic and hydrostatic balance are typified by nearly isotropic PV structures in x, y, and cz, i.e. in coordinates vertically stretched by $c = N/f$ (see Reinaud *et al.* 2003 and references therein), we ensure that the grid resolution is similarly isotropic. (In previous works studying turbulence in rotating stratified flows (e.g. Smith & Waleffe 2002), an isotropic grid in x, y, and z was used, regardless of the value of N/f.)

A more conventional purely pseudo-spectral version of the algorithm evolving the full vector $\mathscr{A}$ has also been developed to demonstrate the advantages of explicit PV conservation in the present 'hybrid' CASL algorithm.

The basic structure of the algorithm consists of flow initialization, interpolation of the PV field to grid-point values, inversion to find the vector potential (and by differentiation the velocity and displacement fields), computation of the imbalance tendencies, and time integration. In addition, a number of optional diagnostic operations may be performed. In the subsections below, each of these steps is described in some detail, with an emphasis on the novel aspects of the algorithm. It turns out to be simplest to describe initialization after the other steps, so we turn first to the interpolation of the PV field.

A.1. *Potential vorticity interpolation*

The advection of potential vorticity, $\dot{\varpi} = 0$, is carried out explicitly in the algorithm by tracking contours of ϖ on a number n_ρ of isopycnal (constant-density) surfaces, following Dritschel & Ambaum (1997) (hereafter referred to as DA97), who introduced this method for the quasi-geostrophic equations. This leads to a significant improvement in PV conservation over grid-based methods (cf. Dritschel *et al.* 1999). Across each contour, ϖ jumps by a given value $\Delta\varpi$ (which may vary from contour to contour), and so the full PV field is not known in advance. It needs to be generated by first filling in the grid points between the contours on an isopycnal surface with the appropriate value of PV – this is done by a 'fast-fill' procedure described in DA97. In that procedure, a grid a power of 2 finer in x and y is used to fill in the values, which are subsequently averaged to the original (coarser) grid. Normally, a grid 2 or 4 times finer in each direction is used (here we use 4). The use of a fine grid allows a more accurate account of fine-scale PV for little extra computational cost (cf. DA97).

The novelty here is that a further vertical interpolation is required to find the PV on a regular Cartesian grid, for compatability with the remaining grid-based parts of the algorithm. This involves two successive interpolations, first to find the height of every isopycnal surface from the gridded displacement field, then to find the PV at grid points lying generally between these surfaces. Both interpolations are taken to be linear for simplicity. The following is done for each (x, y) grid point:

First interpolation

Compute the reference isopycnal height $\bar{Z}_i$ (proportional to density as given in (2.15)) of each vertical grid point $\bar{z}_i$, $i = 1, 2, \ldots n_Z$, from $\bar{Z}_i = \bar{z}_i - \mathscr{D}_i$.

Find the grid points $(i, i+1)$ straddling isopycnal Z_ℓ, $\ell = 1, 2, \ldots n_\rho$, from $\bar{Z}_i \leqslant Z_\ell < \bar{Z}_{i+1}$. The values of Z_ℓ are given by the initialization procedure, e.g. they may be equally spaced to uniformly represent the density.

Compute the actual height z_ℓ of the isopycnal ℓ by linear interpolation, assuming $\mathscr{D}$ varies linearly between adjacent grid points:

$$z_\ell = \bar{z}_i + \Delta\bar{z}\frac{Z_\ell - \bar{Z}_i}{\bar{Z}_{i+1} - \bar{Z}_i} \tag{A 1}$$

where $\Delta\bar{z}$ is the vertical grid spacing (constant).

Second interpolation

Find the isopycnals $(\ell, \ell+1)$ straddling grid point $\bar{z}_i$, $i = 1, 2, \ldots n_Z$, from $Z_\ell \leqslant \bar{z}_i < Z_{\ell+1}$.

Linearly interpolate the PV field between these isopycnals to find the gridded values:

$$\overline{\varpi}_i = \varpi_\ell + (\varpi_{\ell+1} - \varpi_\ell)\frac{\bar{z}_i - Z_\ell}{Z_{\ell+1} - Z_\ell}, \tag{A 2}$$

where ϖ_ℓ is the PV on isopycnal ℓ.

A.2. *Inversion*

Inversion, or the recovery of the original primitive variables from the prognostic set $(\mathscr{A}_h, \varpi)$, is accomplished by solving (2.19) and (2.20) for $\boldsymbol{\varphi}$, then substituting this into (2.13) and (2.12) to calculate $\boldsymbol{u}$ and $\mathscr{D}$ (and if needed $b = -N^2\mathscr{D}$). The pressure is not required (since $\nabla \cdot \boldsymbol{u} = 0$), but if desired it can be obtained by taking the divergence of (2.3) and inverting the Laplacian operator.

In a triply periodic geometry, inversion does not explicitly require boundary conditions on $\boldsymbol{\varphi}$, so long as all fields are considered to be periodic (i.e. by expanding them in Fourier series, as is done here). Otherwise, explicit boundary conditions must be imposed. In particular, at a solid wall the normal velocity must vanish, which involves a condition on the first-order derivatives of $\boldsymbol{\varphi}$, namely $\hat{\boldsymbol{n}} \cdot (\nabla \times \boldsymbol{\varphi}) = 0$, where $\hat{\boldsymbol{n}}$ is the normal vector at the wall. Details are left for future work.

For the periodic boundaries used in the present work, the horizontal potential $\boldsymbol{\varphi}_h$ is computed in spectral space, after a fast Fourier transform, from (2.19) by dividing the spectral components of $-\mathscr{A}_h$ by the total wavenumber squared. All derivatives, like those needed to recover $\boldsymbol{u}$ and $\mathscr{D}$ from $\boldsymbol{\varphi}$, as well as those appearing in the prognostic equations (2.24) for $\mathscr{A}$, are also computed in spectral space (by wavenumber multiplication).

Having found $\boldsymbol{\varphi}_h$, it remains to compute ϕ from the double Monge–Ampère equation (2.20). This is a nonlinear equation with no direct solution method. We have developed a method based on iteration, in which the linear constant-coefficient terms in ϕ are separated from the remaining terms, which are regarded as the 'source', i.e.

$$\mathscr{L}_{\mathrm{qg}}\{\phi\} = S\{\boldsymbol{\varphi}\}, \tag{A 3}$$

where $\mathscr{L}_{\mathrm{qg}}$ is the quasi-geostrophic Laplacian operator defined just following (2.18). We use the previous guess for ϕ to compute S and invert $\mathscr{L}_{\mathrm{qg}}$ in spectral space (by division) to find a new guess for ϕ. The new guess is accepted as the solution if it differs pointwise by less than a prescribed tolerance, here 10^{-7}, from the previous guess. In practice, this procedure converges exponentially fast and in just a few iterations except for extreme situations when Π is significantly less than zero or much larger than unity.

The lack of convergence in these situations appears to be associated with rapid changes in the isopycnal displacement $\mathscr{D}$ and, when $\Pi < 0$, inertial and eventually static instability (see below). Since the numerical algorithm advects PV on isopycnal surfaces, this is sensible only when these surfaces are stably orientated. When $\Pi \gg 1$, isopycnals are strongly bunched in the vertical, leading to strong spatial variations in $\mathscr{D}$ (since, by definition, the mean value of Π is always 1). In this case, higher resolution may be required to achieve convergence of the iterative procedure.

The double Monge–Ampère equation can change type, from elliptic to hyperbolic, over any spatial subdomain when the quantity

$$R \equiv I_a I_c - I_b^2 - I_b'^2 - I_e I_d \tag{A 4}$$

– referred to as Rellich's parameter – changes from positive to negative (see e.g. Courant & Hilbert 1962 or Bakelman 1994). This condition is obtained by linearizing

the original equation about a given solution, and then classifying the resulting linear equation as elliptic or hyperbolic in the standard way. Substituting the expressions for I_a, etc., from (2.21), one may show that

$$Rc^2 = \Pi - \epsilon^2 |\mathscr{A}_h|^2 / 4 \qquad (A\,5)$$

– a remarkably simple combination of the three prognostic variables used in the algorithm. In particular, the imbalance comes in quadratically, and reduces R. One can also see that $\Pi < 0$ implies $R < 0$. In flows not departing greatly from hydrostatic and geostrophic balance, $\mathscr{A}_h = O(1)$ or smaller, while $\epsilon^2 \ll 1$, implying $Rc^2 \approx \Pi > 0$ under these circumstances.

In general, $R > 0$ implies both inertial and static stability, $\tilde{\zeta}^a = 1 + \tilde{\zeta} > 0$ and $Z_z = 1 - \mathscr{D}_z > 0$, respectively (cf. Holton 1992; Knox 1997). This follows by rewriting Rellich's parameter making use of the original double Monge–Ampère equation (2.20):

$$R = [I_e(\phi_{xx} + \phi_{yy}) + I_c][I_e\phi_{zz} + I_a] - (I_e\phi_{xz} - I_b)^2 - (I_e\phi_{yz} - I_b')^2; \qquad (A\,6)$$

hence, if $R > 0$, we must have that the first two square-bracketed terms are either both positive or both negative. When they are both positive, they give simply the conditions for inertial and static stability. This is one of the two solutions to the double Monge–Ampère equation; the other is characterized by both inertial and static instability. These solutions represent fundamentally different physical situations, and an evolving flow cannot simply switch from one to the other. The stable solution is the one always found by the iterative procedure.

In principle, it is possible for the algorithm to deal with situations when $R < 0$ and $\Pi < 0$. We have been able to simulate the beginning stages of inertial instability in a disturbed vertical column having $\Pi = -0.1$ at its centre. The iterative procedure for ϕ converges even though $R < 0$ in a small region near the vortex core, from the end of the initialization period at $t = 5T_{\mathrm{ip}}$ until approximately $t = 7.2T_{\mathrm{ip}}$. The horizontal vorticity grows exponentially fast and violates the CFL criterion around this time. A variable time step might have allowed us to go a little further, but the growth of horizontal vorticity would probably soon lead to shear instability (Richardson numbers $Ri = 1/Fr^2 < \frac{1}{4}$) and overturning. Notably, the same numerical instability occurs, in fact a little sooner, in a pseudo-spectral code (described below) which finds $\boldsymbol{\varphi}$ directly from $\boldsymbol{\varphi} = \nabla^{-2}\mathscr{A}$. It appears necessary to substantially increase the diffusion or use proper molecular-type diffusion to simulate such flows.

Finally, the iterative procedure developed is based on an elliptic solution procedure, whereas flows with $R < 0$ are generally of mixed hyperbolic/elliptic type (in space, not in space–time as is frequently found in other applications). Based on a variety of exact solutions to the linearized equation when it is of mixed type, we believe that there is always a solution. However, in general one may require different solution procedures in the hyperbolic and elliptic subdomains, e.g. the method of characteristics in the hyperbolic ones and successive over-relaxation in the elliptic ones, and tie them together using appropriate boundary conditions along $R = 0$. Certainly this would be a challenging exercise, but perhaps one having only limited utility.

A.3. *Time integration*

The advection of PV is carried out explicitly by solving

$$\frac{\mathrm{d}x_\ell}{\mathrm{d}t} = u_\ell(\boldsymbol{x}_\ell, t), \qquad \frac{\mathrm{d}y_\ell}{\mathrm{d}t} = v_\ell(\boldsymbol{x}_\ell, t) \qquad (A\,7)$$

for each point (x_ℓ, y_ℓ) lying on the PV contours on each isopycnal surface ℓ. The contours are represented in practice by a finite variable number of nodes, which are redistributed at periodic intervals, as discussed in the following subsection. The horizontal velocity used above is found by interpolation, first by linear interpolation of (u, v) from the regularly spaced vertical grid points to the isopycnal heights z_ℓ of every horizontal grid point, then by a bi-linear interpolation of this isopycnal horizontal velocity field to the node positions (x_ℓ, y_ℓ) (as in the original quasi-geostrophic algorithm developed by DA97). In effect, this is a tri-linear interpolation of the horizontal velocity field from the regular three-dimensional grid to an arbitrary point (x_ℓ, y_ℓ, z_ℓ). Higher-order interpolation methods could be used, but there is no evidence to date that this is required for extra accuracy (cf. Dritschel *et al.* 1999).

Once (u_ℓ, v_ℓ) is found in this way, the new node positions are found by a standard third-order three-time-level Adams–Bashforth integration procedure (full details may be found in Dritschel *et al.* 1999). This is highly accurate given that PV contours dominantly exhibit low-frequency motions, and that a small time step is used for evolving the imbalance fields $\mathscr{A}_h$. These (grid-point) fields are evolved by an explicit leap-frog scheme, together with a weak Robert–Asselin time filter (identical to that in Dritschel *et al.* 1999) to avoid the decoupling of even and odd time levels. Explicitly,

$$\mathscr{A}_h^{n+1} = \mathscr{A}_h^{n-1} + 2\Delta t \mathscr{A}_{ht}^n, \qquad (A\,8)$$

where n refers to the time level and Δt is the time step ($t = n\Delta t$). The tendencies $\mathscr{A}_{ht}$ are computed from (2.24), after moving the advective rate of change $\boldsymbol{u} \cdot \nabla \mathscr{A}_h$ to the right-hand side. (As already noted, all nonlinear products required are computed on the grid, while all spatial derivatives are carried out spectrally.)

The use of an explicit leap-frog scheme is accurate and stable only for sufficiently small time steps. The highest frequency motions arise from inertia–gravity waves, which range in frequency roughly between f and N. Accordingly, to resolve the highest frequency motions, we choose $N\Delta t < 1$. In practice, we always choose $N = 2\pi$ (giving a buoyancy period of 1), and $\Delta t = 0.1$. Previous experience with the two-dimensional analogue of this algorithm (Viúdez & Dritschel 2003) has shown that the gain in accuracy from using yet smaller time steps is insignificant.

The numerical integration of nonlinear partial differential equations, spatially discretized on a fixed grid or by a finite set of Fourier coefficients, normally requires some form of diffusion, physical or numerical, for stability. For the present application, molecular viscosity is grossly insufficient to provide numerical stability for the physical scales of interest in the oceans that are normally modelled (tens of metres to thousands of kilometres). Instead, an artificially inflated viscosity or an alternative form of numerical diffusion must be used, e.g. hyperdiffusion. Ideally, one would like to use some kind of turbulent diffusion, but the general form of this diffusion (if it is diffusion) remains obscure generally, and particularly for rotating stratified flows. Instead, researchers have used forms which are easy to implement, which dissipate minimally, and which have minor side effects. These are somewhat conflicting requirements, and there is an art to balancing them. We have made a common compromise by using bi-harmonic hyperdiffusion, which exhibits less serious Gibbs-fringing and overshooting than higher-order forms (cf. Mariotti, Legas & Dritschel 1994 and Jiménez 1994), while also reducing the damping of the principal scales of interest relative to molecular-type harmonic diffusion. Specifically, we add the term $\nu\mathscr{L}_{\mathrm{qg}}^2\{\mathscr{A}_h\}$ to the right-hand side of (2.24); we use $\mathscr{L}_{\mathrm{qg}}$, rather than ∇^2, since $\epsilon\partial_z \sim \nabla_h$ at least in flows having weak to moderate imbalance (such as observed oceanic vortices, cf. Paillet *et al.* 2002). The hyperviscosity coefficient ν is chosen

by specifying the damping rate of the largest wavenumber in spectral space. In the examples presented here, and in another work using this algorithm (Viúdez & Dritschel 2003), we have been able to use a very small damping rate of 1 per inertial period $T_{ip} \equiv 2\pi/f$. Note that $T_{ip} = cT_{bp}$, where the buoyancy period $T_{bp} \equiv 2\pi/N$ (= 1 here), and $c = N/f \gg 1$. That is, the damping rate per buoyancy period is very small, even though the imbalance fields $\mathscr{A}_h$ contain, in part, inertia–gravity waves of frequencies comparable to N.

Finally, to reduce aliasing errors arising from the computation of nonlinear products of fields in (2.24), a steep spectral filter of the form $F(k) = \exp(-c(k/k_{max})^{10})$, with c chosen so that $F(k_{max}) = 10^{-14}$, is applied when computing derivatives (here k refers to the x-, y- or z-wavenumber, as the case may be). This filter was first introduced by Broutman *et al.* (1997) as a low-cost replacement for de-aliasing, and was employed in the two-dimensional analogue of the present algorithm by Viúdez & Dritschel (2002). Aliasing errors otherwise lead to the growth of erroneous grid-scale noise, which eventually becomes overwhelming. In the examples presented in Viúdez & Dritschel (2003) and here, the effects of this filter and of hyperviscosity have not been found to be significant, most probably because the imbalance fields $\mathscr{A}_h$ do not appear to exhibit a strong scale cascade.

A.4. *Contour regularization*

Periodically, the nodes lying on the PV contours need to be redistributed in order to maintain an adequate resolution along the contours. Here, this is done every 5 time steps (half a buoyancy period), but it can be done as infrequently as a twentieth of the characteristic 'eddy turn-around time' $T_{eddy} \equiv 4\pi/|\varpi|_{max}$, for flows that remain close to balance (i.e. when the velocity field is dominantly attributable to ϖ). Typically, PV exhibits a strong scale cascade, via the generation and stretching of filaments, and it is practically impossible to follow these thinning structures for long times. It is also not sensible to follow these structures well below the computational grid scale, since the velocity at the contour nodes is found by interpolation from the grid. In fact, a series of studies has shown that filaments as thin as a tenth of the grid scale still generally exhibit consistent dynamical behaviour (see Waugh & Plumb 1994; DA97, and Methven & Hoskins 1998). The key point is that the velocity field is often significantly smoother than the PV field, so the former can be represented more coarsely than the latter. The representation here of the PV by contours allows one to take advantage of this property, leading to significant gains in accuracy over grid-based numerical methods (Dritschel *et al.* 1999).

Thin filaments are removed in the algorithm by 'surgery' (Dritschel 1989*b*; DA97), a procedure which essentially eliminates filaments, or parts of filaments, thinner than a prescribed surgical scale δ. This dissipation is essential for limiting the growth of contour complexity, and it is relatively benign compared to hyperviscosity and other grid-based dissipation (see e.g. Mariotti *et al.* 1994; Yao, Dritschel & Zabusky 1995; and Dritschel *et al.* 1999). Here, as in Viúdez & Dritschel (2002), we take $\delta = \Delta_h/5m_g$, where Δ_h is the horizontal grid length and m_g is the fine-grid to coarse-grid ratio used in converting PV contours to gridded values on each isopycnal surface. Since subgrid-scale PV contributes to the velocity field, we take the surgical scale to be one-fifth of this subgrid scale (Δ_h/m_g), even though this may be smaller than one-tenth of the original grid scale (as when $m_g = 4$, as used here). This implies a slightly more accurate contour representation overall, since for consistency the maximum node separation $\mu L = \sqrt{2\delta L}$ depends on δ as well as a prescribed large-scale length L. The density of nodes on a contour roughly scales with the square-root of local

curvature divided by μ, but is never less than $(\mu L)^{-1}$ – full details are available in DA97.

A.5. *Initialization procedure*

Our main interest is to study flows starting from a state of near balance in order to better understand both how balance is maintained and when it may (partially) break down, e.g. through spontaneous adjustment emission. This problem underpins both theoretical and practical studies of atmospheric and oceanic dynamics, and yet it remains elusive, in part because of the difficulty in disentangling balanced motions from inertia–gravity waves but also because of poor numerical convergence (Ford *et al.* 2000; Dritschel & Mohebalhojeh 2000). The present numerical method has been developed in order to circumvent these difficulties, by greatly improving the representation of PV, and by better distinguishing balanced from imbalanced motions. We believe this approach is sensible theoretically, and preliminary results (Viúdez & Dritschel 2003) are encouraging.

If we start with a given PV distribution, the fields $\mathscr{A}_h$ must be found which in some measure minimize the initial generation of inertia–gravity waves by adjustment. We have tried various simple balancing methods, e.g. based on quasi-geostrophic balance, but none were found to be satisfactory – in particular, weak fields like vertical velocity were found to be dominated by inertia–gravity waves resulting from the initial imbalance. Instead, we have developed an alternative procedure which does not require specifying balance relations. We start with a zero PV anomaly everywhere but specify PV contours, then we slowly grow the PV anomaly within these contours (by increasing the PV jumps across the contours) over a period long compared with both the buoyancy and inertial periods, here $5T_{ip}$, then leave the PV constant. This very simple procedure has been found to generate virtually no inertia–gravity waves, making it possible to easily detect the balanced component of the vertical velocity and other fields.

The function used to grow the PV is given by

$$W(t) \equiv \tfrac{1}{2}(1 - \cos(\pi t/t_i)), \qquad (A\,9)$$

where $t_i = 5T_{ip}$. Note that W is smooth at $t = 0$ and at $t = t_i$ ($W(0) = 0$, $W(t_i) = 1$, $W_t(0) = W_t(t_i) = 0$). This smoothness is important for reducing the generation of inertia–gravity waves – a linear ramp function does not work nearly so well.

Note that this procedure changes the mass between isopycnal surfaces, since as the PV grows, the isopycnal surfaces deflect, spreading apart where the PV anomaly is positive, and bunching together where it is negative. The fractional added mass between isopycnal surfaces, however, is very small (less than 0.1%) in the examples studied so far. The procedure also does not keep the PV contours in their original positions, so the state of the flow after the period of initialization is not directly under control. Nevertheless, this is unimportant for the present purposes, since we are not trying to reproduce a specific initial condition (e.g. an oceanic observation) but rather exploring how flows in general keep or lose their balance.

A.6. *Pseudo-spectral algorithm*

The pseudo-spectral version of the algorithm developed to compare with the CASL algorithm above uses the same grid-based procedures but of course does not use any of the procedures involving the contours or the PV. Hence, there is no nonlinear equation to invert, but only a Poisson equation for all three components of the vector potential, $\nabla^2 \boldsymbol{\varphi} = \mathscr{A}$, which is inverted spectrally. All of the parameter settings are

the same except for the hyperviscous damping rate, which had to be significantly increased to maintain numerical stability: a value of 200 per inertial period was necessary – a value 100 times greater than needed in the CASL algorithm. This extra diffusion is needed to cope with the strong cascade of PV typical of nonlinear vortex interactions.

In comparisons between the algorithms, we used the fields from the CASL algorithm, at the end of the initialization period, as the initial conditions for the pseudo-spectral algorithm. This permits the closest possible comparison between the two algorithms at subsequent times.

REFERENCES

BAKELMAN, I. 1994 *Nonlinear Geometric Elliptic Equations.* Springer.

BOUSSINESQ, J. 1903 *Theorie Analytique de la Chaleur.* Gauthier-Villars.

BROUTMAN, D., MACASKILL, C., MCINTYRE, M. E. & ROTTMAN, J. 1997 On Doppler-spreading models of internal waves. *Geophys. Res. Lett.* **24**, 2813–2816.

COURANT, R. & HILBERT, D. 1962 *Methods of Mathematical Physics*, Vol. II. Interscience.

DRITSCHEL, D. G. 1989a On the stabilization of a two-dimensional vortex strip by adverse shear. *J. Fluid Mech.* **206**, 193–221.

DRITSCHEL, D. G. 1989b Contour dynamics and contour surgery: numerical algorithms for extended, high-resolution modelling of vortex dynamics in two-dimensional, inviscid, incompressible flows. *Computer Phys. Rep.* **10**, 77–146.

DRITSCHEL, D. G. 1998 On the persistence of non-axisymmetric vortices in inviscid two-dimensional flows. *J. Fluid Mech.* **371**, 141–155.

DRITSCHEL, D. G. & AMBAUM, M. H. P. 1997 A contour-advective semi-Lagrangian algorithm for the simulation of fine-scale conservative fields. *Q. J. R. Met. Soc.* **123**, 1097–1130 (referred to herein as DA97).

DRITSCHEL, D. G. & MOHEBALHOJEH, A. R. 2000 The contour-advective semi-Lagrangian algorithm: keeping the balance. *ECMWF Workshop Proc. on Developments in Numerical Methods for very High Resolution Global Models*, also available at http://www-vortex.mcs.st-andrews.ac.uk/~arm/ecmwf00/index.html.

DRITSCHEL, D. G., POLVANI, L. M. & MOHEBALHOJEH, A. R. 1999 The contour-advective semi-Lagrangian algorithm for the shallow-water equations. *Mon. Wea. Rev.* **127**(7), 1151–1165.

FORD, R., MCINTYRE, M. E. & NORTON, W. A. 2000 Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**, 1236–1254.

HOLTON, J. R. 1992 *An Introduction to Dynamic Meteorology.* Academic.

HOSKINS, B. J., MCINTYRE, M. E. & ROBERTSON, A. W. 1985 On the use and significance of isentropic potential-vorticity maps. *Q. J. R. Met. Soc.* **111**, 877–946.

JIMÉNEZ, J. 1994 Hyperviscous vortices. *J. Fluid Mech.* **279**, 169–176.

KNOX, J. A. 1997 Generalized nonlinear balance criteria and inertial stability. *J. Atmos. Sci.* **54**, 967–985.

MCINTYRE, M. E. & NORTON, W. A. 2000 Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.* **57**, 1214–1235, and Corrigendum **58**, 949.

MARIOTTI, A., LEGRAS, B. & DRITSCHEL, D. G. 1994 Vortex stripping and the erosion of coherent structures in two-dimensional flows. *Phys. Fluids* **6**, 3954–3962.

METHVEN, J. & HOSKINS, B. J. 1998 Spirals in potential vorticity. Part I: Measures of structure. *J. Atmos. Sci.* **55**, 2053–2066.

MOHEBALHOJEH, A. R. & DRITSCHEL, D. G. 2000 On the representation of gravity waves in numerical models of the shallow-water equations. *Q. J. R. Met. Soc.* **126**, 669–688.

MOHEBALHOJEH, A. R. & DRITSCHEL, D. G. 2001 Hierarchies of balance conditions for the f-plane shallow-water equations. *J. Atmos. Sci.* **58**, 2411–2426.

MURAKI, D. J., SNYDER, C. & ROTUNNO, R. 1999 The next-order corrections to quasigeostrophic theory. *J. Atmos. Sci.* **56**, 1547–1560.

OBERBECK, A. 1879 Über die wärnekeutung der flüssigkeiten bei berücksichtigung der strömungen infolge von temperaturdifferenzen. *Ann. Phys. Chem.* **7**, 271–292.

PAILLET, J., LE CANN, B., CARTON, X., MOREL, Y. & SERPETTE, A. 2002 Dynamics and evolution of a northern meddy. *J. Phys. Oceanogr.* **32**, 55–79.

REINAUD, J., DRITSCHEL, D. G. & KOUDELLA, C. R. 2003 The shape of vortices in quasi-geostrophic turbulence. *J. Fluid Mech.* **474**, 175–191.

SMITH, L. M. & WALEFFE, F. 2002 Generation of slow large scales in forced rotating stratified turbulence. *J. Fluid Mech.* **451**, 145–168.

VIÚDEZ, A. & DRITSCHEL, D. G. 2002 An explicit potential vorticity conserving approach to modelling nonlinear internal gravity waves. *J. Fluid Mech.* **458**, 75–101.

VIÚDEZ, A. & DRITSCHEL, D. G. 2003 Vertical velocity in mesoscale geophysical flows. *J. Fluid Mech.* **483** 199–223.

WAUGH, D. W. & PLUMB, R. A. 1994 Contour advection with surgery: a technique for investigating finescale structure in tracer transport. *J. Atmos. Sci.* **51**, 530–540.

YAO, H.-B., DRITSCHEL, D. G. & ZABUSKY, N. J. 1995 High-gradient phenomena in 2D vortex interactions. *Phys. Fluids* **7**, 539–548.