J. Fluid Mech. (2004), *vol.* 521, *pp.* 343–352.
DOI: 10.1017/S0022112004002058 Printed in the United Kingdom

Optimal potential vorticity balance of geophysical flows

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

[1]Institut de Ciènces del Mar, CSIC, Barcelona, Spain
aviudez@cmima.csic.es

[2]School of Mathematics and Statistics, University of St Andrews, St Andrews, UK
dgd@mcs.st-and.ac.uk

(Received 30 June 2004 and in revised from 23 September 2004)

A method to decompose geophysical flows into a balanced flow (defined by its potential vorticity, PV) and an imbalanced component (inertia–gravity waves, IGWs) is introduced. The balanced flow, called the optimal potential vorticity (OPV) balance, is a solution of an IGW-permitting dynamics in which the amount of IGWs is minimal. The residual IGWs are those spontaneously generated by the vortical flow during the numerical integration in which the PV anomaly grows slowly over a sufficiently long ramp period toward a prescribed PV field. The OPV balanced flow is obtained, iteratively, in a cycle of backward and forward integrations where IGWs are removed and PV is restored in every loop. The method is applied to the flow of unsteady vortices in the three-dimensional baroclinic non-hydrostatic dynamics on the f-plane and to the single-layer shallow-water dynamics on the sphere. Both applications show that the iterative method converges strongly, after only a few iterations, to the balanced flow.

1. Introduction

Geophysical flows are to a great extent dominated by the planetary rotation and stable density stratification. Under these circumstances, flows exhibit a nearly 'balanced' state void of inertia–gravity waves (IGWs) and characterized by predominantly lower-frequency vortical motions. This hypothetical state of balance is thought to well-approximate the extra-tropical atmosphere and oceans over a wide range of scales (Hoskins, McIntyre & Robertson 1985; Ford, McIntyre & Norton 2000; and references therein).

A considerable effort has gone into defining and characterizing the properties of the balanced, or vortical, state of a flow (Leith 1980; Lorenz 1980; Medvedev 1999), together with those of the complementary wave state, but no definitive solution has yet been proposed for an unambiguous balanced-flow–wave decomposition of the total flow. Both in oceanic and atmospheric flows this decomposition is of major importance since short- and medium-range forecasting primarily aims to predict the balanced state. Unrealistic large-amplitude IGWs may be artificially generated as a result of a poorly balanced numerical model initialization, and these waves can severely deteriorate forecast accuracy.

In the following, we distinguish between the *balanced component* of a flow and a *balanced flow*. A balanced component of a flow refers to that component, at a given time t, that is free (in some sense) of IGWs. This component might have been

obtained by imposing relations between fields, e.g. as in hydrostatic and geostrophic balance, or through time-filtering of a time series of model data, centred at t, and may, or may not, be a solution of the model dynamics. A balanced flow is here defined as a time series of data which both constitutes a solution of the dynamical model and is free, or more precisely, minimally free, of IGWs. This is a new concept, which might more aptly be called a *minimally imbalanced flow*.

A first avenue to obtaining the balanced flow consists of restricting the dynamical model to a set of approximate equations inherently void of IGWs, as in the quasi-geostrophic model (see e.g. Daley 1991, Ch. 6). However, these approximations are valid only for small Froude numbers (i.e. weak flow speeds compared to IGW speeds), because they filter out part of the balanced motion as well, and are therefore not accurate enough to represent strongly ageostrophic phenomena such as intense mesoscale fronts and vortices. If high-accuracy solutions are sought, then an accurate set of equations has to be used. Such equations, in turn, are also able to capture the spontaneous generation of IGWs by the vortical flow (Lorenz & Krishnamurthy 1987; McWilliams & Yavneh 1988; Vanneste & Yavneh 2004; Vanneste 2004). It is for this reason that the balanced flow should have a small amount of wave activity corresponding to this spontaneous generation. This is what is meant by a balanced flow being *minimally* free of IGWs. One cannot model strongly ageostrophic phenomena accurately without permitting IGWs. Only in this way can one be sure that the balanced flow is not imposed by the dynamical equations but is a solution of the accurate dynamics.

A second avenue of research consists of diagnosing the balanced component of a flow at a given time by explicitly filtering the IGWs from a time series of a numerical model simulation (e.g. the digital filter initialization of Lynch & Huang 1992). This tacitly assumes that vortical motions exist entirely at low frequencies, which cannot be true in general (e.g. a slowly moving PV discontinuity or front generates high frequencies as it passes over fixed spatial locations). This approach is independent of the degree of approximation of the dynamical equations, and IGW-permitting models might be used for this purpose. However, the filtered, balanced component itself is not usually a solution of the dynamical equations, i.e. a time sequence of these balanced components would not constitute a balanced flow, and therefore some artificial IGW generation is expected when this balanced component is used to initialize a numerical simulation.

The general challenge, or third research avenue, is to obtain the balanced flow of an IGW-permitting dynamics directly. We demonstrate that this is indeed possible. Our approach exploits two characteristics of potential vorticity (PV): (i) its inability to transfer itself into IGWs, and (ii) its material conservation. The balanced flow solution, called here the *optimal PV* (OPV) balance, is obtained for two different IGW-permitting dynamics: a three-dimensional baroclinic model on the f-plane, and a two-dimensional rotating shallow-water model on the sphere. For both models we obtain the velocity and mass fields corresponding to a balanced flow, as defined above. The weak residual IGWs present are due to spontaneous generation.

2. The optimal PV balance

It is generally agreed that the PV field is the appropriate field to carry the complete information on the vortical motion, or more precisely, that the PV is the proper field to define it. This relation between PV and the vortical flow is particularly evident in the quasi-geostrophic model, since here all other fields can be obtained from a direct

inversion of the scalar PV. In IGW-permitting models, one might still associate PV with the vortical motion (see e.g. Wang & Zhang 2003 for a recent application to hurricane vortices), though now the velocity and mass fields cannot be obtained by a direct inversion of the PV field. In this work we introduce an approach to obtain the vortical flow, or the OPV balance, through PV inversion using not a direct, but an iterative procedure.

Let $\boldsymbol{x} = \boldsymbol{r}(\boldsymbol{X}, t)$ be the position of the fluid particle $\boldsymbol{X}$ at time t, and its inverse $\boldsymbol{X} = \boldsymbol{R}(\boldsymbol{x}, t)$ be the fluid particle at the position $\boldsymbol{x}$ at time t. For any function f in the spatial ('Eulerian') description we denote the corresponding function in the material ('Lagrangian') description $\tilde{f}$, i.e. $\tilde{f}(\boldsymbol{X}, t) = f(\boldsymbol{r}(\boldsymbol{X}, t), t)$, and $f(\boldsymbol{x}, t) = \tilde{f}(\boldsymbol{R}(\boldsymbol{x}, t), t)$.

In a series of recent studies (Viúdez & Dritschel 2003, 2004*a*,*b*; Dritschel & Viúdez 2003), it has been shown that a balanced state (with minimal IGWs) corresponding to a given *material* distribution of PV, $\tilde{\Pi}(\boldsymbol{X})$, can be obtained by an iterative procedure or time integration over a fictitious time τ. This procedure starts from zero velocity and density (mass) anomaly fields, having a reference PV field $\Pi_r(\boldsymbol{x})$ (which varies with latitude in spherical geometry). Then, the PV *anomaly* in the material description

$$\tilde{\varpi}(\boldsymbol{X}, t) \equiv \tilde{\Pi}(\boldsymbol{X}) - \tilde{\Pi}_r(\boldsymbol{X}, t)\,, \tag{2.1}$$

is increased slowly, by multiplying $\tilde{\varpi}$ by a smooth ramp function $T(\tau)$ which varies from 0 to 1 over the course of the integration. If we represent the PV by contours between which the PV is uniform, then this approach simply involves multiplying the PV jumps across the contours by $T(\tau)$ – see below.

This procedure is carried out at a fixed time t in a numerical simulation, with τ playing the role of variable time. It gives rise to a flow $\boldsymbol{x} = \boldsymbol{r}_T(\boldsymbol{X}, t, \tau)$, called a '$T$-flow', with inverse $\boldsymbol{X} = \boldsymbol{R}_T(\boldsymbol{x}, t, \tau)$, in which the total PV, $\tilde{\Pi}_T$, of every fluid particle increases as

$$\tilde{\Pi}_T(\boldsymbol{X}, t, \tau) = \tilde{\Pi}_r(\boldsymbol{X}, t) + \tilde{\varpi}(\boldsymbol{X}, t)T(\tau)\,. \tag{2.2}$$

Note that (2.2) does not imply that the *spatial* distribution of PV in the T-flow $\Pi_T(\boldsymbol{x}, t, \tau)$ equals $\Pi_r(\boldsymbol{x}) + \varpi(\boldsymbol{x}, t)T(\tau)$, i.e. at any time τ,

$$\begin{aligned}\Pi_T(\boldsymbol{x}, t, \tau) \equiv \tilde{\Pi}_T(\boldsymbol{R}_T(\boldsymbol{x}, t, \tau), t, \tau) &= \Pi_r(\boldsymbol{x}) + \tilde{\varpi}(\boldsymbol{R}_T(\boldsymbol{x}, t, \tau), t)T(\tau) \\ &\neq \Pi_r(\boldsymbol{x}) + \varpi(\boldsymbol{x}, t)T(\tau)\,.\end{aligned} \tag{2.3}$$

The quantity $\tilde{\varpi}(\boldsymbol{R}_T(\boldsymbol{x}, t, \tau), t)$ (a mixed distribution as explained in Viúdez 2003) is the PV anomaly in the actual flow $\boldsymbol{r}$ at time t of the fluid particle which in the T-flow is located at $\boldsymbol{x}$ at time τ.

An example of a simple ramp function is

$$T(\tau) \equiv \tfrac{1}{2}[1 - \cos(\pi\tau/\Delta_\tau)]\,, \tag{2.4}$$

which satisfies $T(0) = 0$, $T(\Delta_\tau) = 1$, and $T'(0) = T'(\Delta_\tau) = 0$, where a prime stands for $\partial/\partial\tau$ and Δ_τ is the initialization period, or length of the ramp function. Thus, at the end of the initialization period,

$$\tilde{\Pi}_T(\boldsymbol{X}, t, \Delta_\tau) = \tilde{\Pi}(\boldsymbol{X})\,, \quad \tilde{\Pi}'_T(\boldsymbol{X}, t, \Delta_\tau) = 0\,, \tag{2.5}$$

i.e. $\tilde{\Pi}_T$ is the prescribed PV and, consistent with the dynamical equations, it is materially conserved.

The fundamental drawback of the above procedure is that, although the material distributions of PV, $\tilde{\Pi}(\boldsymbol{X})$ and $\tilde{\Pi}_T(\boldsymbol{X}, t, \Delta_\tau)$, are identical, the corresponding spatial distributions $\Pi(\boldsymbol{x}, t)$ and $\Pi_T(\boldsymbol{x}, t, \Delta_\tau)$ differ since the fluid particles have been displaced during the course of the integration from their original positions

$\boldsymbol{r}_{\mathsf{f}_i}(\boldsymbol{X},t,0)\mapsto\boldsymbol{r}_{\mathsf{b}_i}(\boldsymbol{X},t,0)$ $\Pi_{\mathsf{f}_i}(\boldsymbol{x},t,0)=\Pi_r(\boldsymbol{x})$ $\boldsymbol{\varphi}_{\mathsf{f}_i}(\boldsymbol{x},t,0)\mapsto\boldsymbol{0}$	forward $\Rightarrow$	$\boldsymbol{r}_{\mathsf{f}_i}(\boldsymbol{X},t,\Delta_\tau)$ $\Pi_{\mathsf{f}_i}(\boldsymbol{x},t,\Delta_\tau)$ $\boldsymbol{\varphi}_{\mathsf{f}_i}(\boldsymbol{x},t,\Delta_\tau)$
remove IGWs $\Uparrow$		$\Downarrow$ restore PV, $i\to i+1$
$\boldsymbol{r}_{\mathsf{b}_i}(\boldsymbol{X},t,0)$ $\Pi_{\mathsf{b}_i}(\boldsymbol{x},t,0)=\Pi_r(\boldsymbol{x})$ $\boldsymbol{\varphi}_{\mathsf{b}_i}(\boldsymbol{x},t,0)$	backward $\Leftarrow$	$\boldsymbol{r}_{\mathsf{b}_i}(\boldsymbol{X},t,\Delta_\tau)\mapsto\boldsymbol{r}(\boldsymbol{X},t)$ $\Pi_{\mathsf{b}_i}(\boldsymbol{x},t,\Delta_\tau)\mapsto\Pi(\boldsymbol{x},t)$ $\boldsymbol{\varphi}_{\mathsf{b}_i}(\boldsymbol{x},t,\Delta_\tau)\mapsto\boldsymbol{\varphi}_{\mathsf{f}_{i-1}}(\boldsymbol{x},t,\Delta_\tau)$

TABLE 1. Schematic of the OPV balance cycle.

$\boldsymbol{r}_T(\boldsymbol{X},t,0)=\boldsymbol{r}(\boldsymbol{X},t)$ to $\boldsymbol{r}_T(\boldsymbol{X},t,\Delta_\tau)\neq\boldsymbol{r}(\boldsymbol{X},t)$. This implies that

$$\tilde{\Pi}_T(\boldsymbol{R}_T(\boldsymbol{x},t,\Delta_\tau),t,\Delta_\tau)=\Pi_T(\boldsymbol{x},t,\Delta_\tau)\neq\Pi(\boldsymbol{x},t)\,. \tag{2.6}$$

This problem can be overcome if, instead of carrying out the τ-integration starting from $\boldsymbol{r}_T(\boldsymbol{X},t,0)=\boldsymbol{r}(\boldsymbol{X},t)$, it is started from a previous initial state of the T-flow, called here the *base configuration*, $\boldsymbol{r}_T(\boldsymbol{X},t,0)$, defined so that at the end of the τ-integration the fluid particles are advected exactly to the position they occupy in the actual flow at time t, that is,

$$\boldsymbol{r}_T(\boldsymbol{X},t,\Delta_\tau)=\boldsymbol{r}(\boldsymbol{X},t)\,. \tag{2.7}$$

If this is achieved then by (2.3) a balanced flow, dependent only on $\Pi(\boldsymbol{x},t)$ and Δ_τ, is obtained at the correct diagnostic time t at the end of the T-flow, $\tau=\Delta_\tau$.

The base configuration $\boldsymbol{r}_T(\boldsymbol{X},t,0)$ is obtained in an iterative way by means of successive backward and forward integrations of the T-flow. Let $\boldsymbol{r}_{\mathsf{b}_i}$ and $\boldsymbol{r}_{\mathsf{f}_i}$ denote the ith backward and ith forward τ-integration of the T-flow. The position of the fluid particles in the actual flow at time t, $\boldsymbol{r}(\boldsymbol{X},t)$, and the material distribution of PV, $\tilde{\Pi}(\boldsymbol{X})$, provide the spatial PV distribution of the actual flow $\Pi(\boldsymbol{x},t)$. Let $\boldsymbol{\varphi}$ be a vector whose components define in some way the flow variables which completely characterize the state of the flow. For example, in incompressible flows $\boldsymbol{\varphi}$ may be the vector potential of the velocity field $\boldsymbol{u}\propto\nabla\times\boldsymbol{\varphi}$ as well as the source of the density anomaly $\rho'\propto\nabla\cdot\boldsymbol{\varphi}$.

Let us take the first approximation to the balanced flow to be $\boldsymbol{\varphi}_{\mathsf{f}_0}(\boldsymbol{x},t,\Delta_\tau)$. This field may be obtained from the actual PV, $\Pi(\boldsymbol{x},t)$, by a smooth forward integration f_0, starting from $\boldsymbol{\varphi}_{\mathsf{f}_0}(\boldsymbol{x},t,0)=\boldsymbol{0}$ and using the ramp function in the spatial description,

$$\Pi_{\mathsf{f}_0}(\boldsymbol{x},t,\tau)=\Pi_r(\boldsymbol{x})+\varpi(\boldsymbol{x},t)T(\tau)\,. \tag{2.8}$$

This is a 'frozen' initialization where the phase, or propagation, speed of vortical structures is zero. Let a PV contour be defined as the line joining particles having the same PV $\Pi(\boldsymbol{x},t)$ (or the same PV jump) in the actual flow. Then, PV contours are not advected in time τ during this first forward integration (f_0). This is the procedure used to obtain $\boldsymbol{\varphi}_{\mathsf{f}_0}(\boldsymbol{x},t,\Delta_\tau)$ in the examples below. An alternative is to set $\boldsymbol{\varphi}_{\mathsf{f}_0}(\boldsymbol{x},t,\Delta_\tau)=\boldsymbol{\varphi}(\boldsymbol{x},t)$, i.e. start with the actual flow.

The fields $\Pi(\boldsymbol{x},t)$, $\boldsymbol{r}(\boldsymbol{X},t)$, and $\boldsymbol{\varphi}_{\mathsf{f}_0}(\boldsymbol{x},t,\Delta_\tau)$ are then used to define the initial state of the first backward τ-integration (flow b_1), $\Pi_{\mathsf{b}_1}(\boldsymbol{x},t,\Delta_\tau)$, $\boldsymbol{r}_{\mathsf{b}_1}(\boldsymbol{X},t,\Delta_\tau)$, and $\boldsymbol{\varphi}_{\mathsf{b}_1}(\boldsymbol{x},t,\Delta_\tau)$, (lower-right box in table 1), in which the dynamical equations are integrated backwards from $\tau=\Delta_\tau$ to 0 using the ramp function $T(\tau)$. Since $T(0)=0$ at the end of this backward τ-integration then $\varpi_{\mathsf{b}_1}(\boldsymbol{x},t,0)=0$, and $\boldsymbol{\varphi}_{\mathsf{b}_1}(\boldsymbol{x},t,0)$ contains

only the IGWs produced by the imbalance between $\Pi_{\mathsf{b}_1}(\boldsymbol{x}, t, \Delta_\tau)$ and $\boldsymbol{\varphi}_{\mathsf{b}_1}(\boldsymbol{x}, t, \Delta_\tau)$. These IGWs are removed from the next forward integration (f_1) by simply setting $\boldsymbol{\varphi}_{\mathsf{f}_1}(\boldsymbol{x}, t, 0) = \boldsymbol{0}$. The particle positions are left unchanged, $\boldsymbol{r}_{\mathsf{f}_1}(\boldsymbol{X}, t, 0) = \boldsymbol{r}_{\mathsf{b}_1}(\boldsymbol{X}, t, \Delta_\tau)$, and the PV anomaly $\varpi_{\mathsf{f}_1}(\boldsymbol{x}, t, 0) = \varpi_{\mathsf{b}_1}(\boldsymbol{x}, t, \Delta_\tau) = 0$. In terms of PV contours the above procedure means that PV contours are left as they are in the transition from the backwards to the forward integration.

The next step is to advect the fluid particles toward their position in the actual flow at t. This is done by means of the forward τ-integration f_1 using again the ramp function $T(\tau)$ to increase the PV anomaly in the material description $\tilde{\varpi}_{\mathsf{f}_1}(\boldsymbol{X}, t, \tau) = \tilde{\varpi}(\boldsymbol{X}, t) T(\tau)$. The use of the ramp function guarantees a minimal generation of IGWs in the new approximation to the balanced state $\boldsymbol{\varphi}_{\mathsf{f}_1}(\boldsymbol{x}, t, \Delta_\tau)$, and a final PV distribution $\Pi_{\mathsf{f}_1}(\boldsymbol{x}, t, \Delta_\tau)$ closely similar to the one in the actual flow $\Pi(\boldsymbol{x}, t)$. The PV of the next backward τ-integration is then restored to its exact value $\Pi(\boldsymbol{x}, t)$, and the flow state $\boldsymbol{\varphi}_{\mathsf{f}_1}(\boldsymbol{x}, t, \Delta_\tau)$ is used for the next backward integration b_2, closing the iterative loop. Convergence is reached when the difference between two consecutive flow states, $\boldsymbol{\varphi}_{\mathsf{f}_k}(\boldsymbol{x}, t, \Delta_\tau)$ and $\boldsymbol{\varphi}_{\mathsf{f}_{k+1}}(\boldsymbol{x}, t, \Delta_\tau)$, becomes sufficiently small.

The OPV balance cycle may be interpreted as an iterative device which removes IGWs and restores PV in every iteration. The closer the balance between $\boldsymbol{\varphi}_{\mathsf{f}_{i-1}}(\boldsymbol{x}, t, \Delta_\tau)$ and $\Pi(\boldsymbol{x}, t)$ is at the end of every iteration, the smaller the amount of IGWs in the next backward integration b_i. This in turn implies that the PV contours at $\tau = 0$ become closer to the ones in the base configuration $\boldsymbol{r}_T(\boldsymbol{x}, t, 0)$ from which, due to (2.7), better balanced fields $\boldsymbol{\varphi}_{\mathsf{f}_i}(\boldsymbol{x}, t, \Delta_\tau)$ are obtained at the end of the forward τ-integration. Few iterations are in fact needed for practical purposes. Three to five may be good enough even for highly ageostrophic flows. The result depends only on Δ_τ (and on the form of the ramp $T(\tau)$), but only weakly so, based on a wide range of examples (see § 3).

3. Numerical results

The results presented next have been obtained with two different numerical models. We focus here on the OPV balance. Details of the numerical models, apart from some necessary remarks, are given in the cited references.

3.1. *Three-dimensional mesoscale baroclinic flow*

A three-dimensional baroclinic isochoric flow on the f-plane is simulated with a triply-periodic non-hydrostatic model whose prognostic variables are the horizontal ageostrophic vorticity and the PV. The latter is represented by contours in every isopycnal (constant density surface). The state variables are the components of the vector potential $\boldsymbol{\varphi}$ which provide the velocity $\boldsymbol{u} \propto \nabla \times \boldsymbol{\varphi}$ and the vertical displacement of isopycnals $D \propto \nabla \cdot \boldsymbol{\varphi}$. This approach, described in detail in Dritschel & Viúdez (2003), has been applied to the study of a variety of ageostrophic flows in the atmosphere and oceans (Viúdez & Dritschel 2003, 2004*a*,*b*).

Because density itself is materially conserved, only the horizontal positions $\boldsymbol{r}_h(\boldsymbol{X}, t)$ of the fluid particles in every isopycnal are required for the OPV balance. The vertical displacement of an isopycnal is a function of $\boldsymbol{\varphi}$. We may define a fluid particle by the three material invariants $\boldsymbol{X} = (X, Y, Z) = (\ell, \Pi, \rho)$, where ℓ labels the particle on the PV contour Π lying on the isopycnal ρ. Thus the position $\boldsymbol{r}_{\mathsf{b}_i}(\boldsymbol{X}, t, \Delta_\tau)$ is restored, not to $\boldsymbol{r}(\boldsymbol{X}, t)$, but to $\boldsymbol{r}_h(\boldsymbol{X}, t) + \tilde{z}_{\mathsf{f}_{i-1}}(\boldsymbol{X}, t, \Delta_\tau)\boldsymbol{k}$, where $\boldsymbol{k}$ is the unit vertical vector. The vertical location $\tilde{z}$ depends, through the vertical isopycnal displacement D, on the vector potential $\boldsymbol{\varphi}$.

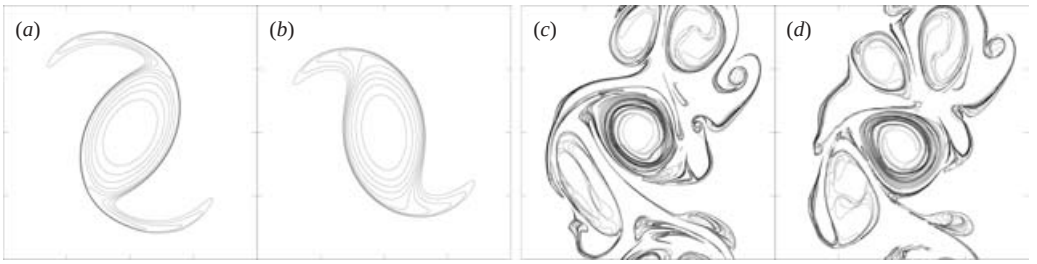

FIGURE 1. (a) The original PV contours $\varpi(\boldsymbol{x}, t)$, and ($b$) the PV contours after the first backward integration $\boldsymbol{r}_{\mathsf{b}_1}(\boldsymbol{X}, t, 0)$, in the ellipsoid simulation ($t = 10$ i.p., and contour interval $\Delta = 0.05$); (c, d) same as (a, b) but in the dipole simulation ($t = 20$ i.p., $\Delta = 0.075$). In both cases $z = 0$, and the domain extent is $10^2([-\pi, \pi] \times [-\pi, \pi])$.

Two numerical simulations are analysed. The ratio between the background Brunt–Väisälä and Coriolis frequencies is set, without loss of generality to $N/f = 10^2$, so that 1 inertial period (i.p.) is 10^2 buoyancy periods (b.p.). The first simulation (figure 1a, b) consists of a single baroclinic anticyclone initially defined by a three-dimensional ellipsoidal distribution of PV with $\min\{\varpi\} = -0.5$ at the centre of the vortex. The domain has $(n_X, n_Y, n_Z) = (64, 64, 128)$ grid points, and $n_\rho = 128$ isopycnals. At $t = 10$ i.p., the diagnostic period, the extreme Rossby and Froude numbers are $\min\{Ro\} = -0.38$ and $\max\{Fr\} = 0.19$, respectively. The second simulation (figure 1c,d) consists initially of a northward baroclinic jet, with maximum and minimum PV anomalies $\varpi = \pm 0.75$, which breaks up to form dipolar vortices. In this case $(n_X, n_Y, n_Z) = (64, 64, 64)$ grid points, and $n_\rho = 64$ isopycnals. At the diagnostic period $t = 20$ i.p. $(\min\{Ro\}, \max\{Ro\}) = (-0.67, 0.63)$ and $\max\{Fr\} = 0.39$.

The PV contours (figure 1) exhibit significant displacement from the beginning to the end of the first backward integration b_1 with an initialization period $\Delta_\tau = 5$ i.p. The PV contours at the end of the forward integrations, $\varpi_{\mathsf{f}_i}(\boldsymbol{x}, t, \Delta_\tau)$, are visually indistinguishable from $\varpi_{\mathsf{b}_1}(\boldsymbol{x}, t, \Delta_\tau) = \varpi(\boldsymbol{x}, t)$ and are not shown.

No IGWs are generated during the first integration with steady PV contours (f_0) because the vortices do not move (spectra are shown in figure 2). However, the large-amplitude balanced components at frequencies smaller than f are missing. During integration f_0 the displacement, or phase speed of the vortices, is zero, and hence the velocity at each spatial location always increases (or decreases). However, in the subsequent backward and forward integrations the vortices move, so that the local rate of change of the velocity (at each location) may change sign, which explains the large amplitude of the balanced components at frequencies smaller than f. The first backward integration (b_1) includes these low-frequency components but develops relatively large-amplitude IGWs due to the imbalance between $\boldsymbol{\varphi}_{\mathsf{f}_0}$ and $\Pi(\boldsymbol{x}, t)$. The next forward integration (f_1) includes the low-frequency balanced components and avoids the generation of large-amplitude IGWs caused by any initial imbalance. The small-amplitude IGWs still present are due to the spontaneous generation occurring during the ramp period Δ_τ, which necessarily have to be considered part of the balanced flow in an IGW-permitting dynamics. Successive forward and backward integrations have spectra very similar to f_1. For practical purposes, the spectra have converged after only a few iterations.

The balanced vertical velocity field for the second simulation $w_{\mathsf{f}_3}(\boldsymbol{x}, t, \Delta_\tau)$ (figure 3a) adjusts to the large scale of the vortices, while the imbalanced part $w(\boldsymbol{x}, t) - w_{\mathsf{f}_3}(\boldsymbol{x}, t, \Delta_\tau)$ (figure 3b) exhibits a more homogeneous distribution of small-scale patterns with maximum amplitudes one to two orders of magnitude smaller

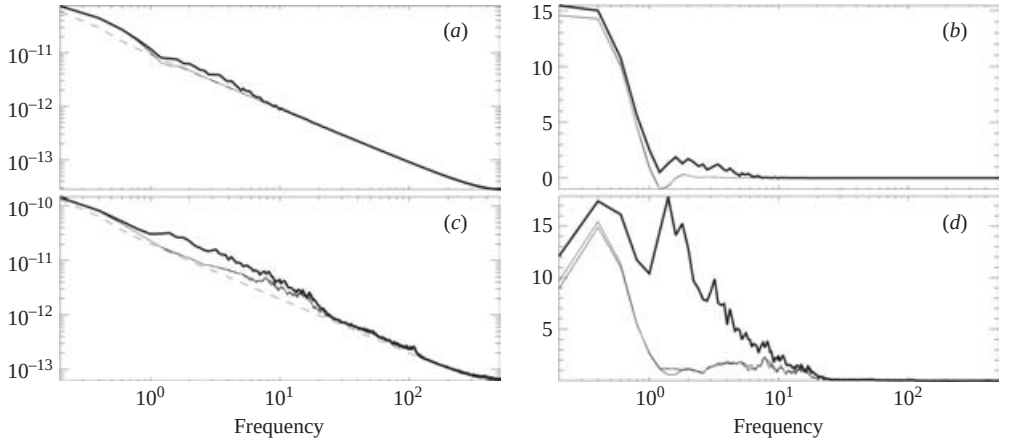

FIGURE 2. (*a*) Averaged frequency spectra of w^2 in the ellipsoid simulation: f_0 (dashed line), b_1 (thick line), and f_1, b_2, f_2, b_3, and f_3 (thin lines, indistinguishable except at few locations); (*b*) the difference spectra $\mathsf{b}_1 - \mathsf{f}_0$ (thick line), and $\mathsf{f}_i - \mathsf{f}_0$, $\mathsf{b}_i - \mathsf{f}_0$ for the other flows (thin lines), in units of 10^{-12}cycles/b.p.); (*c*, *d*), same as (*a, b*) but for the dipole simulation. Frequency is in cycles/i.p. Every spectrum is the average of 8^3 spectra each computed from the time series from $\tau = 0$ to Δ_τ.

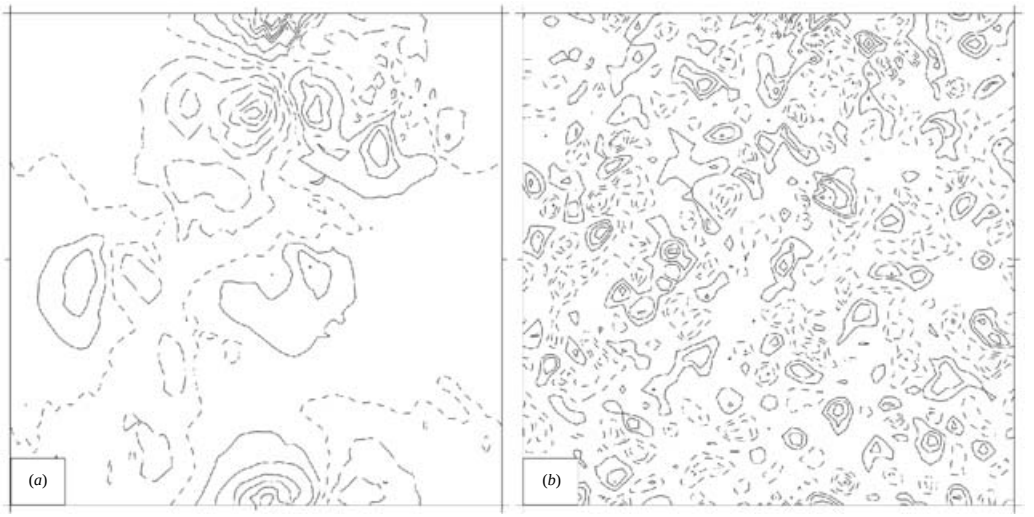

FIGURE 3. Horizontal distribution of (*a*) $w_{\mathsf{f}_3}(\boldsymbol{x}, t, \Delta_\tau)$, at the depth of maximum vertical velocity (contour interval $\Delta = 5 \times 10^{-5}$; long-dashed/short-dashed/solid contours are used for negative/zero/positive values). (*b*) The imbalance $w(\boldsymbol{x}, t) - w_{\mathsf{f}_3}(\boldsymbol{x}, t, \Delta_\tau)$ ($\Delta = 5 \times 10^{-6}$; the zero contour is omitted).

than the balanced part. The difference between $w_{\mathsf{f}_3}(\boldsymbol{x}, t, \Delta_\tau)$ and $w_{\mathsf{f}_2}(\boldsymbol{x}, t, \Delta_\tau)$ has a grid-size scale pattern and is two orders of magnitude smaller than the balanced vertical velocity.

3.2. *Two-dimensional shallow-water flow on the sphere*

A series of numerical studies has demonstrated the advantages of using different sets of prognostic variables for the accurate simulation of shallow-water (SW) flows in planar geometry ($f =$ constant; Mohebalhojeh & Dritschel 2000, 2001). A particularly

robust choice across the full range of Froude numbers for which the SW equations are valid is the set (δ, γ, Π), where $\delta = \nabla_h \cdot \boldsymbol{u}_h$ is the velocity divergence ($\boldsymbol{u}_h$ is the horizontal velocity field), $\gamma = \nabla_h \cdot \boldsymbol{a}_h$ is the acceleration divergence (proportional to the ageostrophic vertical vorticity), and Π is the SW PV. The use of these variables has been found to strongly reduce the artificial numerical generation of IGWs that occurs when using standard variable sets such as (u, v, h) and (δ, ζ, h), where h is the height (or mass) anomaly of the fluid layer and ζ is the vertical vorticity.

As in the non-hydrostatic model above, all fields can be expressed in terms of a vector potential $\boldsymbol{\varphi}$, here given by $\boldsymbol{\varphi} = (\psi, \chi, h)$, in terms of which

$$\boldsymbol{u}_h = \boldsymbol{k} \times \nabla_h \psi + \nabla_h \chi, \quad \boldsymbol{a}_h = -f \boldsymbol{k} \times \boldsymbol{u}_h - g \nabla_h h\,, \tag{3.1}$$

where $f(\vartheta)$ is the Coriolis frequency at latitude ϑ, and g is the acceleration due to gravity. From $\delta = \nabla_h \cdot \boldsymbol{u}_h$ and $\gamma = \nabla_h \cdot \boldsymbol{a}_h$, together with the definition of SW PV

$$\Pi \equiv (\zeta + f)/(1 + h/H)\,, \tag{3.2}$$

where H is the mean height of the fluid layer, one can find $\boldsymbol{\varphi}$ by inversion (here via linear elliptic operators). This procedure has been successfully adapted to spherical geometry and verified against standard test cases. Here we note only that the method uses contour advection for the PV evolution, as in the non-hydrostatic model above.

Results are next described for a simulation of rotating SW turbulence on a 128×128 latitude–longitude grid. The initial conditions were generated from a random isotropic PV anomaly ϖ spatially correlated over the scale $L_c = 0.1$ on a unit sphere. A highly nonlinear flow regime was set up by taking the short-scale wave speed $c = \sqrt{gH}$ to be $4\pi/3$, the planetary rotation rate Ω to be 2π (so the unit of time is a day), and the r.m.s. value of ϖ to be $4\pi/3$. After a five-day ramped PV initialization the flow was evolved for a further 20 days. Over this period $\max\{|Ro|\} = \max\{|\zeta|/2\Omega\} = 1.47$ and $\max\{Fr\} = \max\{|\boldsymbol{u}_h|/c(1 + h/H)^{1/2}\} = 0.47$.

The diagnosis of the OPV balance was carried out at times $t = 10$ and 20 after initialization and using a variety of ramp periods Δ_τ ranging from 1 to 10 days. Here, only results for $t = 10$ and for a ramp period of $\Delta_\tau = 4$ are described. Note that $\max\{|Ro|\} = 0.83$ and $\max\{Fr\} = 0.38$ at $t = 10$.

The PV contours $\Pi(\boldsymbol{x}, t)$ in the actual flow at $t = 10$ (figure 4*a*) are greatly distorted from their base configuration $\boldsymbol{r}_{\mathrm{b}_6}(\boldsymbol{X}, t, 0)$ (figure 4*b*). In this example the r.m.s. differences $|\boldsymbol{\varphi}_{\mathrm{f}_{k+1}}(\boldsymbol{x}, t, \Delta_\tau) - \boldsymbol{\varphi}_{\mathrm{f}_k}(\boldsymbol{x}, t, \Delta_\tau)|$, decrease by more than an order of magnitude at each iteration k.

The divergence frequency spectra, for various iterations in the OPV balance cycle and averaged over 64 points on the sphere (figure 5*a*) are strikingly similar to the vertical velocity spectra found in the non-hydrostatic case (figure 2). All the forward spectra converge rapidly, as seen by the differences from the f_0 spectrum in figure 5(*b*). The actual divergence, the OPV-balanced divergence, and the difference (or imbalance) fields are shown as a function of longitude and latitude in figure 6. The imbalance is greatest near the equator, where geostrophic balance breaks down. A small-scale wave-train pattern is evident near the centre of the domain, but there is also a global scale (wave 1) pattern superposed. In the height field (not shown), a wave 1 pattern is clearly visible near the equator (evidence of a Kelvin wave). Quantitatively similar results are found for other ramp periods with $\Delta_\tau \geqslant 2$ days. Furthermore, similar patterns of imbalance are found at $t = 20$ and in other simulations analysed.

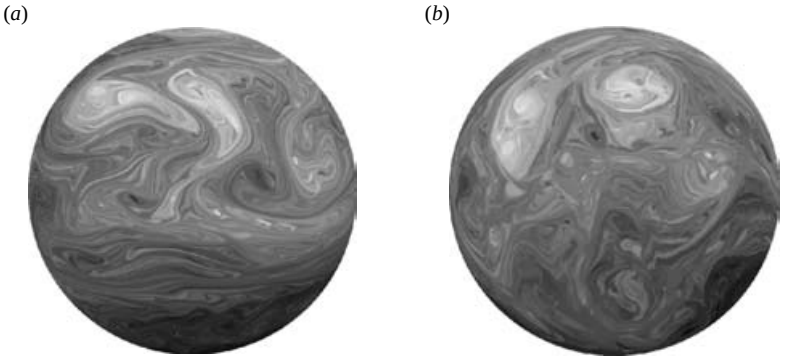

FIGURE 4. (*a*) The original PV contours, $\Pi(\boldsymbol{x}, t)$, at $t = 10$, and (*b*) the PV contours at the beginning of the 4-day ramp, $\boldsymbol{r}_{\mathsf{f}_6}(\boldsymbol{X}, t, 0)$, which evolve into those in (a) from $\boldsymbol{\varphi}_{\mathsf{f}_6}(\boldsymbol{x}, t, 0) = \boldsymbol{0}$ while ramping up the PV anomaly. The contours are separated by uniform shades of grey, with an intensity proportional to the PV (white is used for the maximum PV). The actual contours are too dense to plot directly. The view is orthographic from 0° longitude and 45° latitude.

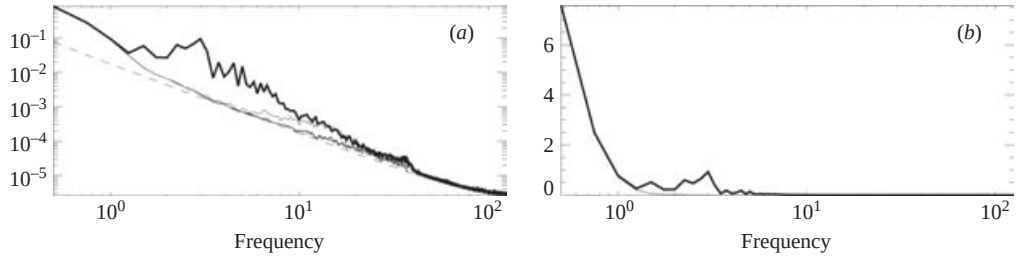

FIGURE 5. (*a*) Divergence frequency spectra, $\hat{\delta}^2$, for various iterations of the OPV balance cycle (as in figure 2*a, c*). (*b*) Difference spectra (in units of 10^{-1}) relative to that for iteration f_0 (as in figure 2*b, d*). The forward iterations f_i, for $i \geqslant 1$ are virtually indistinguishable in both plots.

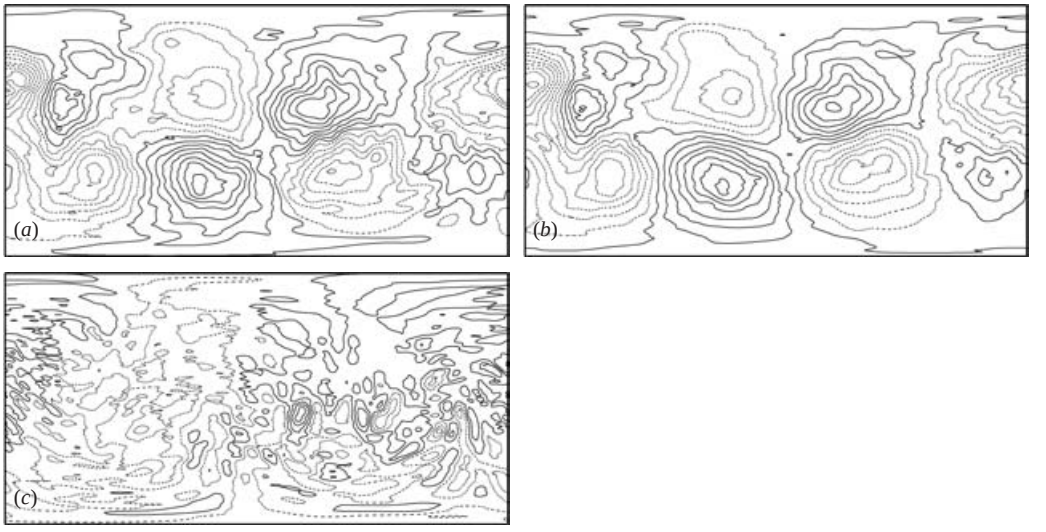

FIGURE 6. Longitude–latitude maps of (*a*) divergence δ, (*b*) balanced divergence $\delta_{\mathrm{bal}} = \delta_{\mathsf{f}_6}(\boldsymbol{x}, t, \Delta_\tau)$, and (*c*) the difference $\delta - \delta_{\mathrm{bal}}$ as diagnosed by the OPV balance at $t = 10$ (contour interval $\Delta = 0.05$ except in (*c*) where $\Delta = 0.025$; negative values are dashed).

4. Concluding remarks

The OPV balance is, conceptually, a solution of an IGW-permitting dynamics in which the amount of IGWs is minimal. The residual IGWs are those spontaneously generated by the vortical flow during a forward integration in which the PV anomaly grows slowly over a sufficiently long ramp period toward a prescribed PV field. In practice, the OPV balanced flow is obtained, iteratively, in a cycle of backward and forward integrations where IGWs are removed and PV is restored to its prescribed value in every loop. Practical applications of this balance include its use in forecast initialization, the generation and evolution of the imbalanced part of the flow, in particular the dynamics of spontaneous generated IGWs, and the interaction between IGWs and the balanced flow.

Support for this research has come from the UK Engineering and Physical Sciences Research Council (grant number XEP294), the Spanish program *Ramón y Cajal 2001* and the *Ministerio de Ciencia y Tecnología* (grant number REN2002-01343). We thank Professor Michael McIntyre, Dr Ali Mohebalhojeh, and two reviewers for their comments.

REFERENCES

Daley, R. 1991 *Atmospheric Data Analysis*. Cambridge University Press.

Dritschel, D. G. & Viúdez, A. 2002 A balanced approach to modelling rotating stably-stratified geophysical flows. *J. Fluid Mech.* **488**, 123–150.

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

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.

Leith, C. E. 1980 Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.* **37**, 958–968.

Lorenz, E. N. 1980 Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.* **37**, 1685–1699.

Lorenz, E. N. & Krishnamurthy, V. 1987 On the nonexistence of a slow manifold. *J. Atmos. Sci.* **44**, 2940–2950.

Lynch, P. & Huang, X.-Y. 1992 Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.* **120**, 1019–1034.

McWilliams, J. C. & Yavneh, I. 1998 Fluctuation growth and instability associated with a singularity of the balance equations. *Phys. Fluids* **10**, 2587–2596.

Medvedev 1999 The slow manifold for the shallow water equations on the f plane. *J. Atmos. Sci.* **56**, 1050–1054.

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.

Vanneste, J. 2004 Inertia-gravity wave generation by balanced motion: revisiting the Lorenz-Krishnamurthy model. *J. Atmos. Sci.* **61**, 224–234.

Vanneste, J. & Yavneh, I. 2004 Exponentially small inertia-gravity waves and the breakdown of quasigeostrophic balance. *J. Atmos. Sci.* **61**, 211–223.

Viúdez, A. 2003 The multi-flow framework and the generalized Lagrangian mean theory. *Fluid Dyn. Res.* **32**, 141–177.

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

Viúdez, A. & Dritschel, D. G. 2004*a* Potential vorticity and the quasigeostrophic and semigeostrophic mesoscale vertical velocity. *J. Phys. Oceanogr.* **34**, 865–887.

Viúdez, A. & Dritschel, D. G. 2004*b* Dynamic potential vorticity initialization and the diagnosis of mesoscale motion. *J. Phys. Oceanogr.*, in press.

Wang, X. & Zhang, D.-L. 2003 Potential vorticity diagnosis of a simulated hurricane. Part I: Formulation and quasi-balanced flow. *J. Atmos. Sci.* **60**, 1593–1607.