

Balanced Models and Dynamics for the Large- and Mesoscale Circulation

MICHAEL D. MUNDT, GEOFFREY K. VALLIS, AND JIAN WANG

Department of Ocean Science, University of California, Santa Cruz, Santa Cruz, California

(Manuscript received 14 November 1995, in final form 19 November 1996)

ABSTRACT

A balanced model that incorporates the dynamics of both the quasigeostrophic and planetary geostrophic equations is examined via numerical simulations. The model is valid for large variations in the Coriolis parameter and layer thickness, typical of dynamics at the gyre scale and larger, as well as for mesoscale dynamics in which the advection of relative vorticity cannot be neglected. For length scales much larger than the deformation radius the dominant balances in the model equations are those of the planetary geostrophic equations, whereas for synoptic scales the dynamics of the model are asymptotically close to those of the quasigeostrophic equations. The model consists of the advection of a geostrophic potential vorticity, and is therefore conceptually simple and numerically easy to implement, requiring the solution of just one linear elliptic equation each time step. It is compared to the shallow-water equations and various other approximate models in order to assess its validity for modeling meso- and large-scale dynamics. Simulations are performed in both a periodic channel and also in a closed domain with wind stress forcing. The formulation is found to yield a substantial improvement in the qualitative nature of the circulation patterns, and in quantitative accuracy, over the planetary geostrophic and (especially) the quasigeostrophic equations, while still retaining much of the conceptual simplicity of those equations.

1. Introduction

Quasigeostrophic (QG) theory has been an enormously successful tool in helping to elucidate fundamental processes in synoptic-scale geophysical flows, including eddy generation by barotropic and baroclinic instability and the wind-driven circulation of the oceans. The QG equations, however, are formally valid only when the planetary vorticity is large compared to the relative vorticity of the fluid, when the variation in the Coriolis parameter over the domain is small compared to its mean value, and when free-surface and topographic fluctuations are both much smaller than the average fluid depth (e.g., Pedlosky 1987). The latter two constraints are not obviously, or in fact, requirements for the existence of balanced motion and, indeed, the planetary geostrophic equations (Welander 1959; Robinson and Stommel 1959) do allow large variations in both Coriolis parameter and layer depth. However, these equations are “inertia-free,” meaning that the advection of momentum is neglected in the momentum equations and hence are inappropriate for anything but the large-scale circulation. The primitive equations are not hampered by such constraints, but they allow short timescale processes, such as gravity waves, which are often con-

sidered to be largely unimportant for low Rossby number flows. Their presence complicates numerical implementation and, perhaps more importantly, inhibits theoretical understanding of the important dynamical processes.

Various intermediate models have been formulated in an attempt to extend the formal validity of balanced motion beyond the QG approximation while avoiding the presumably extraneous complications allowed by the primitive equations. There are two rather different ways in which the quasigeostrophic (or planetary geostrophic) equations can in principle be extended:

- 1) by proceeding to higher order in an asymptotic expansion (or related procedure) in Rossby number (or other small parameter)
- 2) by attempting to extend the regime of validity of the equations at lowest order.

The former approach was taken by, for example, Browning and Kreiss (1987), Allen (1993), and Warn et al. (1995). Specifically, Warn et al. showed how approximations of arbitrarily high order can be achieved by supposing that variables are slaved to a single master evolution variable—potential vorticity for example. Allen’s iterative approach is in many ways similar. The disadvantage of these approaches is that it does not extend the regime of validity of the equations so derived beyond that of the lowest order system, quasigeostrophy for example. Furthermore, because of the likely lack of a true globally slow manifold due to the spontaneous

Corresponding author address: Dr. Geoffrey K. Vallis, Department of Ocean Science, University of California, Santa Cruz, CA 95064. E-mail: vallis@cascade.ucsc.edu

emission of gravity waves (e.g., Warn and Menard 1986), it seems unlikely that any asymptotic approach will give significant improvements beyond leading order or possibly next order. Browning and Kreiss (1987) used the related method of bounded derivatives to construct self-consistent approximations. This method, while rigorous and powerful in guaranteeing the consistency of any approximations made, similarly does not allow for large variations in the height field.

The second approach does not necessarily seek to formally go to higher asymptotic order. Instead, it attempts to include terms that are neglected (or inaccurately approximated) in the quasigeostrophic or planetary geostrophic formulations and that thereby severely restrict their range of applicability. These include terms arising from large variations in height or Coriolis parameter that might be neglected in the QG approximation, or the advection of relative vorticity, neglected in a planetary geostrophic formulation. Vallis (1996) showed that a properly balanced model, based on the advection and inversion of potential vorticity, can be derived that is valid for both the planetary and synoptic scales, in contrast to either quasigeostrophy or planetary geostrophy, which are each valid for just one of these scales. Note that the presence of large height variations does not imply that balanced motion does not exist or that accurate approximate models cannot be constructed; the restriction is that any $O(1)$ height variations must occur on a scale larger than the Rossby deformation radius ($\sqrt{gh/f}$) or else the Rossby number will not be small (e.g., Pedlosky 1987). The planetary geostrophic equations can be asymptotically derived in this limit (e.g., Phillips 1963). The so-called frontal geostrophic equations of Cushman-Roisin are another set of balanced equations valid for finite— $O(1)$ —variations in layer thickness, although for small variations in Coriolis parameter. The frontal geostrophic equations may, in fact, be compactly derived from potential vorticity advection—see the appendix.

There are, of course, many so-called “intermediate” models. Notable are the semigeostrophic equations (Hoskins 1975) and the Charney balance model (Charney 1962). The linear balance model is a variant on this (Gent and McWilliams 1983). Allen et al. (1990a,b) and Barth et al. (1990) tested a large number of single-layer intermediate models against a shallow-water formulation for an f -plane geometry with large topographic variations, primarily with mesoscale and coastal applications in mind. They found that QG performed quite poorly and was nearly always the least accurate of all of the models studied. In contrast, the Charney balance (CB) equations, or slight variants thereof, consistently performed quite well in comparison to the shallow-water solutions. The Charney balance model is, in fact, of higher asymptotic order than QG (Warn et al. 1995); however, its evident numerical success may well be because its assumptions are not nearly as restrictive as QG [item 2) above], not its asymptotic order per se. Also,

for large variations of the Coriolis parameter it is, in fact, not first order in time and therefore not properly balanced (see section 2d).

A primary drawback of many balanced models is the complexity of their implementation. Boundary conditions are often difficult to pose properly, and for many models a nonlinear equation must be solved, usually iteratively at every time step. Moreover, it is sometimes difficult to discern why a particular balanced approximation succeeds or fails because its formulation is often complex and possibly somewhat ad hoc. Now, the large-scale midlatitude ocean circulation is certainly characterized by a low Rossby number, but layer thickness and Coriolis parameter may vary by an $O(1)$ amount. For example, outcropping of isopycnals, corresponding to an extreme height variation, occurs in the subpolar gyre. These factors formally preclude the use of quasigeostrophy to study these aspects of large-scale circulation. A planetary geostrophic model is, of course, valid for such large scales, yet it is by construction invalid for scales comparable to the deformation scale, or the oceanic mesoscale. Thus, neither model is appropriate for both meso- and large-scales, and proceeding asymptotically to higher order using either quasigeostrophic or planetary geostrophic scaling will, of course, still not give a model valid for both regimes. And yet on both the mesoscale and planetary scale the ocean is manifestly balanced, indeed in very good geostrophic balance except possibly close to boundaries.

Our primary goal in this paper is therefore to explore balanced models that are, in fact, valid for both the mesoscale and planetary scale, in particular a model based on the advection of geostrophic potential vorticity, explored in a preliminary way in Vallis (1996). The model is relatively simple to implement numerically in a closed domain. It is also conceptually simple and its validity in particular situations well defined because the asymptotically dominant dynamics of the model are those of the planetary geostrophic equations at very large scales and are those of the quasigeostrophic equations at synoptic scales. The performance of this “geostrophic potential vorticity” (GPV) model will be compared to both quasigeostrophic and shallow-water results in addition to those from a few other intermediate models. To a lesser extent, we will also examine the merits of a planetary geostrophic model, which is perhaps the simplest of all balanced models for large-scale dynamics. Our reasons for using single-layer, rather than stratified, models are twofold: First, such models exhibit the essential dynamics in the most transparent way. Second, multilevel primitive equation models almost always require thermodynamic forcing, and it is difficult to compare the adiabatic dynamics with those of balanced models (cf. McWilliams et al. 1990).

In section 2, we present the governing equations for the various models. Numerical results are presented in section 3 for both unforced simulations in a periodic channel and also for a wind-driven, closed-box domain.

The applicability of the various models is further examined in section 4 via the inversion of the potential vorticity field produced by the shallow-water model. In section 5, we examine the energetics of the GPV model and compare them to those from a shallow-water formulation. A discussion of the results is contained in section 6.

2. Governing equations

a. Shallow-water equations

The dimensional forms of the shallow-water (SW) momentum and continuity equations are, respectively,

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv = -g \frac{\partial \eta}{\partial x} + \nu \nabla^2 u \quad (2.1)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu = -g \frac{\partial \eta}{\partial y} + \nu \nabla^2 v \quad (2.2)$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0, \quad (2.3)$$

where the full layer thickness h is given by

$$h = H_0 + \eta(x, y, t) - h_b(x, y). \quad (2.4)$$

Here H_0 , η , and h_b are the mean layer depth, the free surface deviation, and the topographic height, respectively. The variables f , g , and ν are, respectively, the Coriolis parameter, the gravitational constant (or reduced gravity), and the kinematic viscosity. We use the β -plane approximation that $f = f_0 + \beta y$, where $\beta \equiv \partial f / \partial y$. We then nondimensionalize using the following scales: $\{u, v\} \sim U$, $\{x, y\} \sim L$, $t \sim L/U$, $\eta \sim f_0 UL/g$, $\{h, h_b\} \sim H_0$. This yields

$$\epsilon \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) - (1 + \epsilon \beta_0 y)v = -\frac{\partial \eta}{\partial x} + \frac{\epsilon}{\text{Re}} \nabla^2 u, \quad (2.5)$$

$$\epsilon \left(\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right) + (1 + \epsilon \beta_0 y)u = -\frac{\partial \eta}{\partial y} + \frac{\epsilon}{\text{Re}} \nabla^2 v, \quad (2.6)$$

$$\epsilon F \frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}(uh) + \frac{\partial}{\partial y}(vh) = 0, \quad (2.7)$$

where

$$h = 1 + \epsilon F \eta - h_b. \quad (2.8)$$

The variables u , v , η , h , and h_b are now nondimensional. The nondimensional parameters are the Rossby number:

$$\epsilon \equiv \frac{U}{f_0 L}, \quad (2.9)$$

the Froude number:

$$F \equiv \frac{f_0^2 L^2}{g H_0}, \quad (2.10)$$

the ‘‘beta’’ parameter:

$$\beta_0 \equiv \frac{\beta L^2}{U}, \quad (2.11)$$

and the Reynolds number:

$$\text{Re} \equiv \frac{UL}{\nu}. \quad (2.12)$$

We can take the curl of the momentum equations to get a vorticity equation, where the vorticity ζ is defined as

$$\zeta \equiv \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \quad (2.13)$$

The nondimensional vorticity equation is

$$\begin{aligned} \epsilon \left(\frac{\partial \zeta}{\partial t} + u \frac{\partial \zeta}{\partial x} + v \frac{\partial \zeta}{\partial y} \right) + D(1 + \epsilon \zeta + \epsilon \beta_0 y) \\ + \epsilon \beta_0 v = \frac{\epsilon}{\text{Re}} \nabla^2 \zeta, \end{aligned} \quad (2.14)$$

where $D \equiv \partial u / \partial x + \partial v / \partial y$. Similarly, we can obtain a divergence equation for D , that is,

$$\begin{aligned} \epsilon \left(\frac{\partial D}{\partial t} + u \frac{\partial D}{\partial x} + v \frac{\partial D}{\partial y} + D^2 - 2J(u, v) + \beta_0 u \right) \\ - \zeta(1 + \epsilon \beta_0 y) = -\nabla^2 \eta + \frac{\epsilon}{\text{Re}} \nabla^2 D, \end{aligned} \quad (2.15)$$

where $J(f, g) \equiv f_x g_y - f_y g_x$ and subscripts denote partial differentiation. Finally, combining (2.7) and (2.14) by eliminating D , and dividing through by ϵ , we obtain

$$\frac{DQ}{Dt} = \frac{1}{h} \frac{\nabla^2 \zeta}{\text{Re}}, \quad (2.16)$$

where the potential vorticity Q is defined as

$$Q \equiv \frac{\zeta + (1 + \epsilon \beta_0 y)/\epsilon}{h}, \quad (2.17)$$

and

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}. \quad (2.18)$$

In the absence of dissipation ($\nu = 0$), Q is conserved following fluid parcels.

b. Quasigeostrophic equations

The quasigeostrophic (QG) equations can be obtained by an asymptotic expansion of the variables in terms of the presumed small parameter ϵ . The $O(1)$ balances of (2.5), (2.6), and (2.7) yield

$$v = \frac{\partial \eta}{\partial x} \quad (2.19)$$

$$u = -\frac{\partial \eta}{\partial y} \quad (2.20)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \quad (2.21)$$

The $O(\epsilon)$ momentum equations can be cast into a vorticity equation

$$\begin{aligned} \frac{\partial \zeta}{\partial t} + u \frac{\partial \zeta}{\partial x} + v \frac{\partial \zeta}{\partial y} + \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \\ \times (\zeta + \beta_0 y) + \beta_0 v + D^{(1)} = \frac{1}{\text{Re}} \nabla^2 \zeta, \end{aligned} \quad (2.22)$$

where

$$D^{(1)} \equiv \frac{\partial u^{(1)}}{\partial x} + \frac{\partial v^{(1)}}{\partial y}$$

and $u^{(1)}$ and $v^{(1)}$ are the $O(\epsilon)$ velocities. The $O(\epsilon)$ continuity equation is, after using (2.21),

$$F \frac{\partial \eta}{\partial t} + u \frac{\partial}{\partial x} \left(F \eta - \frac{h_b}{\epsilon} \right) + v \frac{\partial}{\partial y} \left(F \eta - \frac{h_b}{\epsilon} \right) + D^{(1)} = 0. \quad (2.23)$$

The $O(\epsilon)$ vorticity and continuity equation can be combined by elimination of $D^{(1)}$ and by using (2.19) and (2.21) to yield the quasigeostrophic potential vorticity equation,

$$\frac{\partial}{\partial t} (\nabla^2 \eta - F \eta) + J(\eta, q) = \frac{1}{\text{Re}} \nabla^4 \eta, \quad (2.24)$$

or

$$\frac{Dq}{Dt} = \frac{1}{\text{Re}} \nabla^4 \eta, \quad (2.25)$$

where q is the QG potential vorticity defined as

$$q \equiv \nabla^2 \eta - F \eta + h_b/\epsilon + \beta_0 y. \quad (2.26)$$

In order that all terms in (2.26) are $O(1)$, QG scaling requires that F , β_0 , and h_b/ϵ be $O(1)$ as well. Physically, this implies that the free surface and topographic variations must be small compared to the mean layer depth, and similarly that the Coriolis parameter variation be small compared to its mean value.

c. Geostrophic potential vorticity model equations

The geostrophic potential vorticity (GPV) model was presented by Vallis (1996), who treats (2.16) for potential vorticity as the fundamental equation of motion. To close the system, one needs to find a prescription for the advecting velocities. This is done simply by imposing local geostrophic balance such that

$$u = -\frac{1}{1 + \epsilon \beta_0 y} \frac{\partial \eta}{\partial y} \quad (2.27)$$

$$v = \frac{1}{1 + \epsilon \beta_0 y} \frac{\partial \eta}{\partial x}. \quad (2.28)$$

These relations should be contrasted with the quasigeostrophic balance given in (2.19), (2.20) that use a mean value of the Coriolis parameter. Using (2.13), the vorticity is then

$$\zeta = \frac{\nabla^2 \eta + \epsilon \beta_0 u}{1 + \epsilon \beta_0 y} \quad (2.29)$$

and the potential vorticity is

$$Q = \frac{(\nabla^2 \eta + \epsilon \beta_0 u)/(1 + \epsilon \beta_0 y) + (1 + \epsilon \beta_0 y)/\epsilon}{1 + \epsilon F \eta - h_b}. \quad (2.30)$$

In dimensional terms this is simply

$$Q = \frac{f + \zeta_g}{h - h_b} = \frac{f + \nabla \cdot (g/f) \nabla h}{h - h_b}. \quad (2.31)$$

This equation constitutes an inversion formula for the height field η , given the potential vorticity field Q , requiring the solution of an elliptic equation. The model is evolved by advecting the potential vorticity with the geostrophic wind given by (2.27) and (2.28); that is,

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + u \frac{\partial Q}{\partial x} + v \frac{\partial Q}{\partial y} = F - D, \quad (2.32)$$

where the right-hand side contains forcing and dissipative terms. The GPV model is similar to that proposed by Bleck (1973) for a continuously stratified flow, although in his formulation the Coriolis parameter is held constant in certain terms. It is also similar to the ‘‘L1’’ dynamics of Salmon (1983), except that the ageostrophic flow is not explicitly calculated.

The GPV equations reduce to the quasigeostrophic equations if quasigeostrophic scaling is imposed. That is, (2.30) asymptotically reduces to (2.26) if $\epsilon \beta_0 = O(\epsilon)$, $\epsilon F = O(\epsilon)$, and $h_b = O(\epsilon)$. The GPV equations reduce to a planetary geostrophic system if the relative vorticity is neglected in (2.30), which is appropriate for very large scale flow.

d. Charney balance equations

The Charney balance model (CB) for shallow water flow is obtained by retaining the full vorticity and height equations, but truncating the divergence equation. This is achieved by assuming that the velocities may be written as

$$u = -\frac{\partial \psi}{\partial y} + \epsilon \frac{\partial \chi}{\partial x} \quad (2.33)$$

$$v = \frac{\partial \psi}{\partial x} + \epsilon \frac{\partial \chi}{\partial y}, \quad (2.34)$$

which is simply a decomposition of the flow field into a nondivergent portion (ψ) and an irrotational component (χ). However, it is *presumed* for truncation purposes that the magnitude of the χ field is $O(\epsilon)$ in comparison to ψ . Equations (2.33) and (2.34) yield

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi \quad (2.35)$$

$$D = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \epsilon \nabla^2 \chi \quad (2.36)$$

for the vorticity ζ and the divergence D . We then substitute these relations into the vorticity equation (2.14) and keep terms of $O(\epsilon)$ and $O(\epsilon^2)$. This actually results in the retention of all terms so that

$$\frac{\partial \zeta}{\partial t} + u \frac{\partial \zeta}{\partial x} + v \frac{\partial \zeta}{\partial y} + \nabla^2 \chi (1 + \epsilon \zeta + \epsilon \beta_0 y) + \beta_0 v = \frac{1}{\text{Re}} \nabla^2 \zeta \quad (2.37)$$

after dividing through by ϵ . Doing the same for the divergence equation (2.15), we obtain

$$\nabla^2 \psi (1 + \epsilon \beta_0 y) - \epsilon \beta_0 \left(-\frac{\partial \psi}{\partial y} + \epsilon \frac{\partial \chi}{\partial x} \right) + 2\epsilon J(\psi_x, \psi_y) = \nabla^2 \eta. \quad (2.38)$$

If we wish to model flows with considerable variation in the planetary vorticity, that is, $\beta L/f = \epsilon \beta_0 \sim 1$, the $\epsilon^2 \beta_0 \partial \chi / \partial x$ term ought to be retained in the above equation. However, its inclusion allows the presence of high frequency waves in the problem, an entity we are attempting to eliminate from the possible solutions. Therefore, we neglect this term and the divergence equation becomes

$$\nabla^2 \psi (1 + \epsilon \beta_0 y) + \epsilon \beta_0 \frac{\partial \psi}{\partial y} + 2\epsilon J(\psi_x, \psi_y) = \nabla^2 \eta. \quad (2.39)$$

The full continuity equation, given by (2.7), is retained, and it can be expressed as

$$\epsilon F \frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} [u(\epsilon F \eta - h_b)] + \frac{\partial}{\partial y} [v(\epsilon F \eta - h_b)] + \epsilon \nabla^2 \chi = 0. \quad (2.40)$$

After substitution of (2.33) and (2.34) for the velocities, Eqs. (2.37), (2.39), and (2.40) form a closed system for the variables ψ , η , and χ . Note from the above argument that, in general, the CB model is not first order in time and, in that sense, is not a properly balanced model for geophysical flow.

e. Planetary geostrophic equations

The planetary geostrophic (PG) equations can be constructed from a potential vorticity approach, similar to the methodology used for the GPV formulation. Equations (2.16), (2.27), and (2.28) are again used, but the contribution to the potential vorticity by the relative vorticity ζ is completely neglected. This results in the

simple expression (in dimensional variables) $Q = f/h$ or in nondimensional form

$$Q = \frac{(1 + \epsilon \beta_0 y)/\epsilon}{h}. \quad (2.41)$$

Note that the equations are useful only when the variation of the Coriolis parameter is large—that is, when $\beta f/L \sim O(1)$ or $\beta_0 = O(1/\epsilon)$ in (2.41). (The quasigeostrophic scaling of beta, namely, $\beta_0 \equiv \beta L^2/U$, is used to keep notational consistency throughout the paper.) If the Coriolis parameter is constant or nearly so, there is no advection of the potential vorticity by the geostrophic wind at lowest order; the evolution therefore occurs on a timescale that is Rossby number slower than the advective timescale. A higher order velocity must be calculated to capture this evolution (see Cushman-Roisin 1986 and the appendix).

The PG equations are also strictly valid only for scales much larger than the deformation scale (Phillips 1963) and where planetary vorticity is much larger than the relative vorticity, or equivalently where $\epsilon \ll 1$. However, both the nondimensional height deviation and the divergence may be $O(1)$.

3. Numerical results

a. Numerical formulation

All the modeling is performed on a grid, with derivatives obtained via finite differencing. The QG and GPV models are defined on A grids (Mesinger and Arakawa 1976). For both models, the potential vorticity is marched forward in time and the height field is then obtained via an inversion using (2.26) for QG and (2.30) for GPV. Since the QG equation has constant coefficients, a fast Helmholtz solver is employed. The GPV equation, however, has coefficients that vary in x and y ; we consequently use a multigrid method to obtain the solution. Leapfrog time stepping is used, and at regular intervals an Euler step is employed to eliminate spurious computational modes. The Jacobians are computed with the Arakawa method (Arakawa 1966). The CB model is also defined on an A grid, and a leapfrog time-marching method is utilized. The integration scheme mimics that of Allen et al. (1990b), who give a detailed description of the methodology. The SW model uses a staggered C grid (Mesinger and Arakawa 1976). The nonlinear terms are computed with the energy- and enstrophy-conserving Arakawa–Lamb scheme (Arakawa and Lamb 1981), and time stepping is done via the leapfrog method with an Asselin filter (Asselin 1972) to eliminate computational instability.

b. Unforced periodic channel

In order to assess the fidelity of the GPV model to the shallow-water equations for synoptic- and large-scale flow, and to compare to other approximate models,

two different numerical tests are utilized: The first test is a comparison of eddy dynamics in a periodic channel. In addition to the SW and GPV models, we also include the QG and CB formulations. The latter is used for comparisons because its success in modeling f -plane flows over topography (Allen et al. 1990b) makes it a reasonable benchmark for balanced models in general.

The geometry utilized is a channel of equal length and width that is periodic in the zonal direction. The drawback of a channel geometry is that it is difficult or impossible to make the boundary conditions identical for each model. This proved to be problematic in several of our initial attempts, as differences at the sidewalls amplified quickly and rendered quantitative comparison difficult. We avoid this difficulty by considering flows whose solutions are very weak near the boundaries.

At the sidewalls, the condition of no normal flow ($v = 0$) is enforced. For the QG model, this requirement is satisfied as long as η is a constant (not necessarily the same) on each boundary. The values of η at the sidewalls can be determined by requiring that mass be conserved and that circulation at either wall be conserved in the absence of Laplacian or higher-order dissipation (McWilliams 1977). Conservation of circulation at the other sidewall is then also guaranteed. Similar requirements were placed on the GPV model, though the simultaneous conservation of mass and circulation is not necessarily always consistent. Consequently, we enforce mass conservation and circulation conservation at one of the boundaries and note that, in general, we do not observe circulation conservation at the other wall (especially when $\beta_0 \neq 0$). For the CB model, we largely follow the boundary conditions implemented by Allen et al. (1990b), who impose circulation and no-normal-flow requirements. They, however, set the total meridional velocity, that is, $v = \partial\psi/\partial x + \epsilon\partial\chi/\partial y$, to zero at the sidewalls. For arbitrary initial conditions, we sometimes encountered numerical instability when activity was present near the boundaries, and consequently we amended the condition so that $\partial\psi/\partial x$ and $\partial\chi/\partial y$ are set to 0 independently.

Dimensionally, we choose a square geometry with $L = 6000$ km, $H_0 = 10$ km, $f_0 = 1.03 \times 10^{-4}$ s $^{-1}$, $\beta = 1.62 \times 10^{-11}$ (m s) $^{-1}$, and $g = 2.5$ m s $^{-2}$. The values of f_0 and β are reasonable for the midlatitudes. The value of L was chosen in order to ensure $\beta L/f_0$ is $O(1)$. The actual values of H_0 and g are not crucial but are important in determining the Froude number and hence the size of the dynamic free-surface deviations. Non-dimensionally, the domain is defined by $0 \leq x \leq 1$ and $-1/2 \leq y \leq 1/2$. The resolution we use is 129^2 for all models, which is sufficient to resolve the scales of interest.

Numerical tests were performed in order to examine whether the GPV model is capable of improving upon QG dynamics when (i) variations in the Coriolis parameter f are as large as the mean value f_0 , and/or (ii) variations in layer thickness h are the same order as the

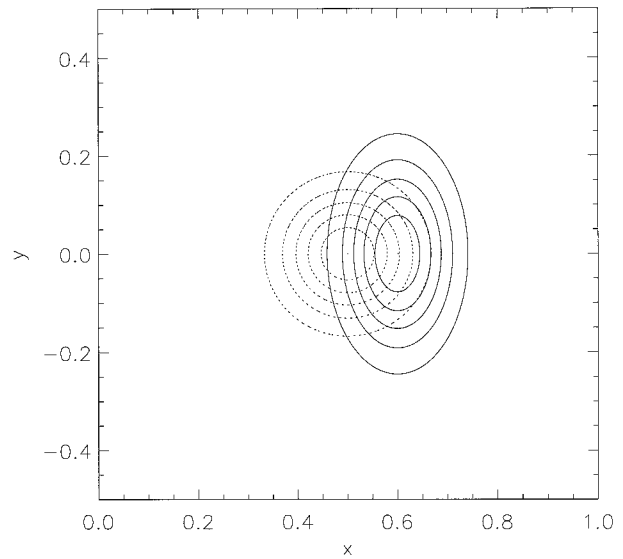


FIG. 1. Topography (dotted lines) and initial conditions (solid lines) for the unforced channel experiments.

mean thickness. The former is effected simply by employing a β -plane geometry since, for our choice of parameters, $\beta L/f_0 \approx 0.94$. To enforce the latter, we chose to incorporate topography rather than make the initial height deviations the same order as the mean layer height. We use for topography a Gaussian hump centered in the domain such that

$$h_b = A_b e^{-(8y)^2} e^{-[8(x-1/2)]^2}. \quad (3.1)$$

We initialize all the models by specifying the same height field, which is an anisotropic Gaussian vortex centered at $x = 0.6$, $y = 0$. A contour plot of the topography and the initial height field, denoted by dotted and solid lines respectively, is shown in Fig. 1. The height field is positive and thus initially has anticyclonic motion. For the SW model, we initialize the velocity fields by imposing local geostrophic balance as defined in (2.27) and (2.28). The simulations are run without forcing or dissipation. Although a lack of viscosity will eventually destabilize the numerical solutions as enstrophy accumulates in the high wavenumbers, we did not observe any detrimental effects over the integration time intervals used in this study. Another important issue that can be addressed in inviscid simulations is that of energy and enstrophy conservation. Further discussion of this issue is presented in section 4.

The initial peak velocity of the vortex is about 3 m s $^{-1}$, which yields a scaling Rossby number of $\epsilon = U/f_0 L = 0.005$. If we take the maximum of the relative vorticity and divide by the mean planetary vorticity, we achieve a maximum local Rossby number of approximately 0.15. The flow is therefore in a geostrophic regime. The initial height field is chosen so that the maximum perturbation height divided by the mean layer thickness is about 0.01. This ensures that any significant

TABLE 1. Numerical runs performed for channel geometry.

Expt	A_b	$\beta L/f_0$	Evolution time
c1	0.10	0	26.2
c2	0.10	0.94	20.9
c3	0.70	0	7.9
c4	0.70	0.94	7.9

layer thickness deviations are due to the topography only. Despite our effort to make the flow as geostrophic as possible, a small amount of gravity wave activity in the SW model is frequently generated from the initial conditions, not surprisingly because we do not perform a sophisticated initialization procedure. These do not alter the large-scale geostrophic flow by an appreciable amount, but due to the lack of viscosity they remain as small-amplitude disturbances observable in the height field.

We present four different simulations, listed in Table 1. The integration periods are generally different and were chosen so that the initial field could evolve substantially. The cases with larger topography evolve more quickly and therefore were run for shorter time intervals. Evolution times are given in inertial periods, where we define a dimensional inertial period as $2\pi/\Omega$ (i.e., one day). For each experiment, a correlation coefficient ρ , defined as

$$\rho = \frac{\sum_{ij} \eta(i, j) \eta_{sw}(i, j)}{\left[\sum_{ij} \eta(i, j)^2 \sum_{ij} \eta_{sw}(i, j)^2 \right]^{1/2}}, \quad (3.2)$$

is computed over time. In Eq. (3.2), η is the height deviation of a particular model, η_{sw} is the height deviation in the shallow-water model, and the sums are taken over all grid points. The results are shown in Fig. 2 for

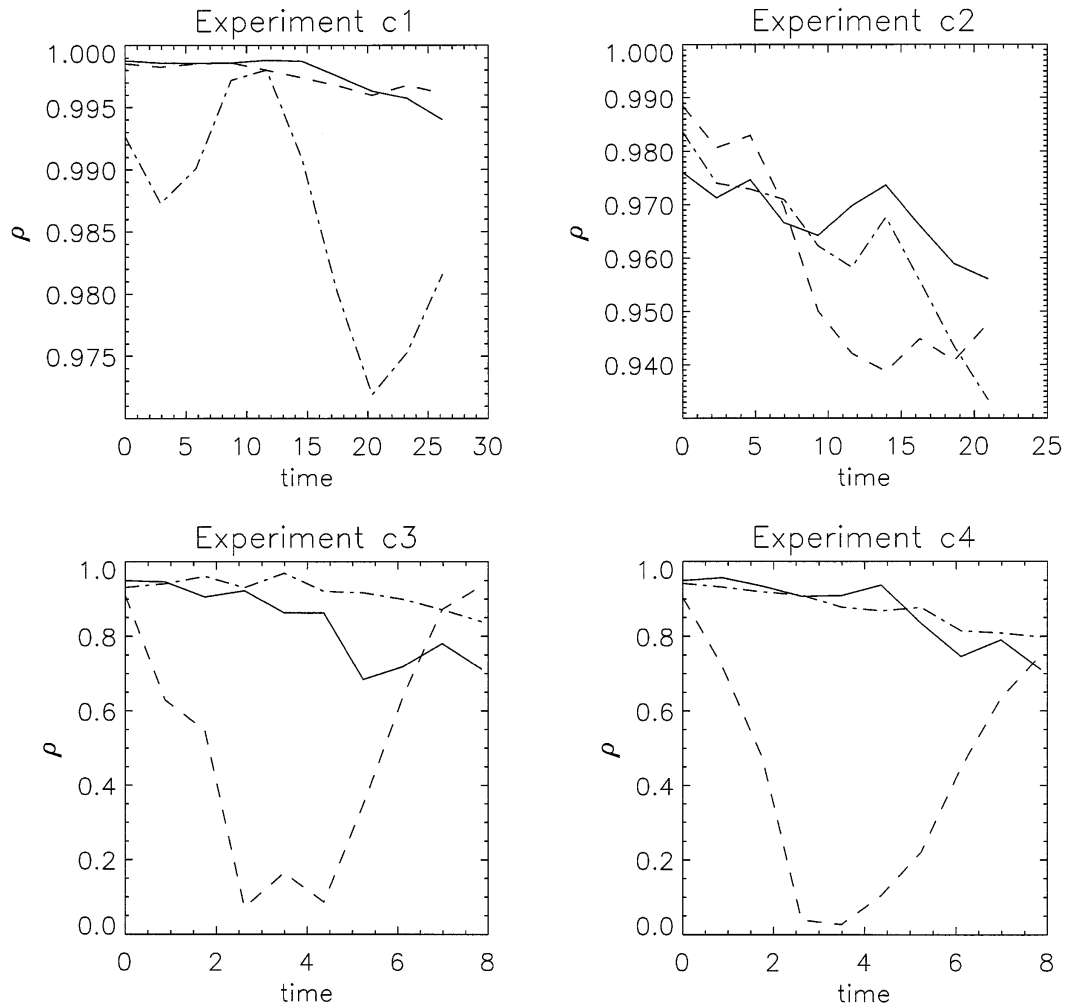


FIG. 2. Correlation coefficients vs time (in inertial periods) for the four numerical channel experiments. The solid, dashed, and dot-dashed lines display the SW/GPV, SW/QG, and SW/CB correlations, respectively. Note the offscale location of the zero in C1 and C2.

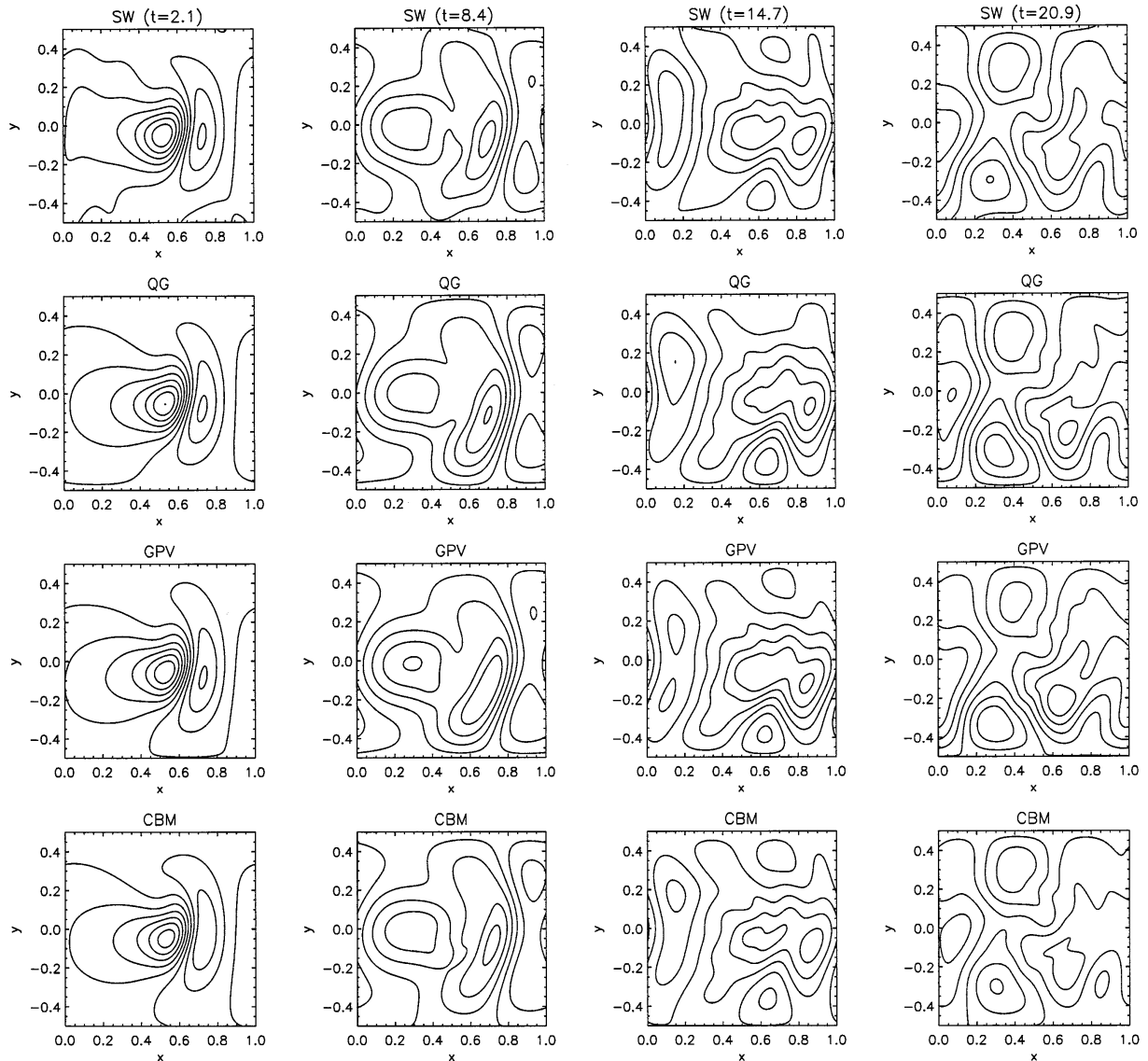


FIG. 3. Snapshots of height fields for experiment c2 at $t = 2.1, 8.4, 14.7,$ and 20.9 inertial periods. From top to bottom, the models are SW, QG, GPV, and CB. Time proceeds from left to right.

the four cases. The solid line shows the SW/GPV correlation, while the dashed and dot-dashed lines show the SW/QG and SW/CB correlations, respectively. In experiment c1, which is in a quasigeostrophic regime, with $\beta = 0$ and $A_b = 0.1$, all the models perform quite well, displaying very little reduction in correlation over 25 inertial periods. The CB model does marginally worse than either QG or GPV, though this is most likely due to the slightly different boundary conditions.

For experiment c2, which also has $A_b = 0.1$ but has $\beta L/f_0 = 0.94$, the QG, GPV, and CB models again all yield solutions quite similar to the SW control run. The good performance of QG seems somewhat surprising, given that $\beta L/f_0 \approx 1$, because it indicates the Rossby wave dispersion relation in QG theory is approximately

valid for $\beta_0 \gg 1$. However, a numerical calculation of eigenfunctions and wave phase speeds using a linearized version of the GPV formulation yields results very similar to QG theory. The phase speeds are found to be only about 1% different than those obtained from the QG approximation, and the eigenfunctions are very similar to the sinusoidal QG modes. In this instance, the QG formulation does quite well far beyond its expected range of validity. Figure 3 shows snapshots of the height fields for the four models at selected times. Over the entire period of integration, a high degree of similarity to the SW solution is evident for all the models, though some differences are apparent in the fine structure.

For experiments c3 and c4, the QG model performs quite poorly, as is evident from Fig. 2. The correlation

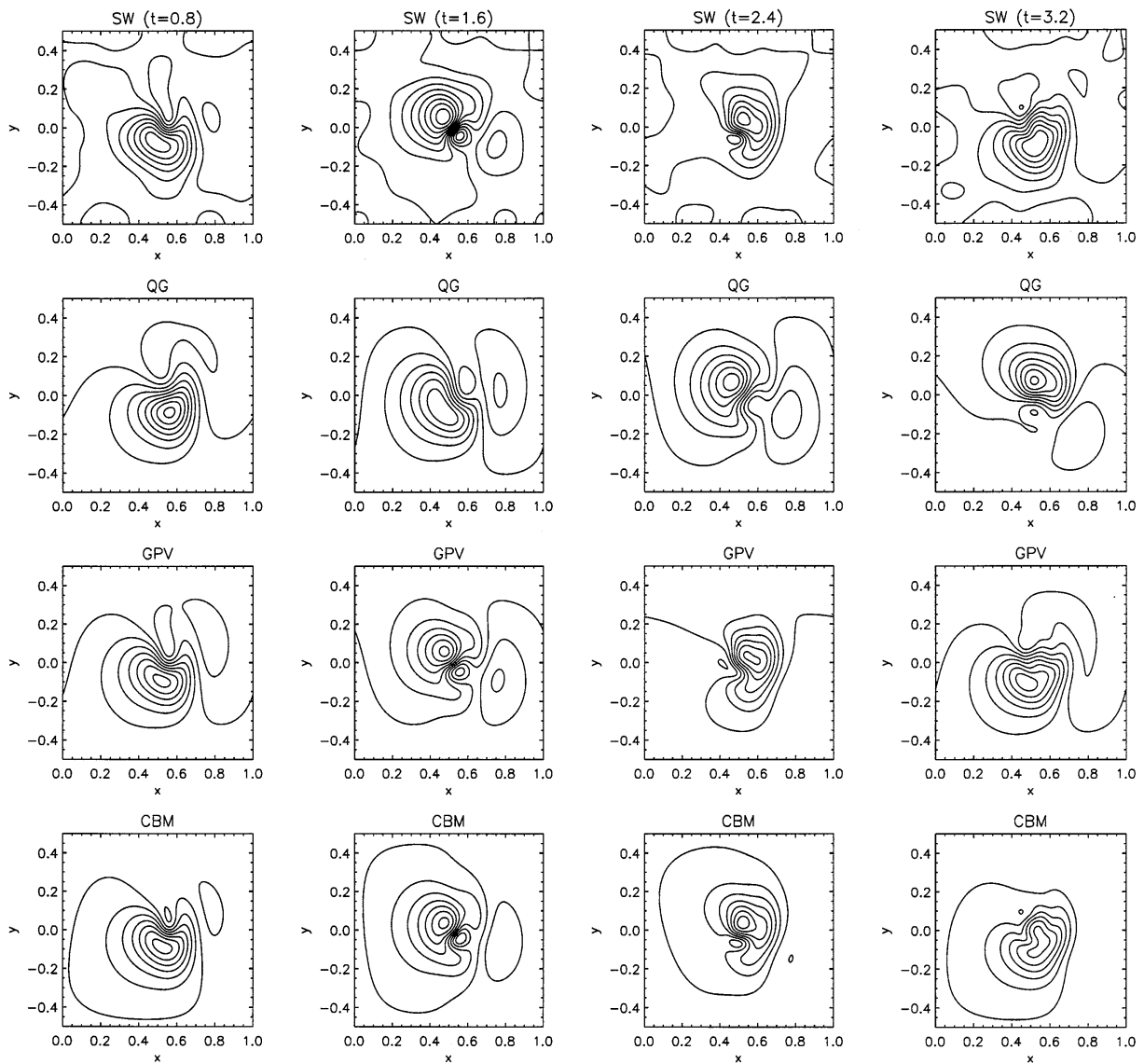


FIG. 4. Snapshots of height fields for experiment c3 at $t = 0.8, 1.6, 2.4,$ and 3.2 inertial periods. From top to bottom, the models are SW, QG, GPV, and CB. Time proceeds from left to right.

with the SW solution drops quite quickly, curiously rising to near unity again at the end of integration period. This is because the flow dynamics largely consist of the initial vortex rotating counterclockwise around the topography. The QG model performs about two rotations during the time period examined, while the remaining models perform three. The increase in correlation at later times is therefore completely spurious. This is evident in Fig. 4, which shows the height fields for the models near the beginning of experiment c3. By $t = 1.6$ inertial periods, the QG solution lags quite noticeably behind the other models. In addition, both the GPV and CB models capture portions of the finescale detail, such as the dipolelike structure evident for $t = 1.6$ and $t = 2.4$. In both experiments c3 and c4, the GPV solution appears

to rotate slightly too fast, while the CB model behavior appears more faithful to that of SW.

We note that the SW model appears to exhibit measurable height deviations near the sidewalls. This is due to gravity wave emission that occurs at the start of the integration. The gravity waves do not significantly alter the evolution of the vortex however, so no effort was made to eliminate them.

In that portion of parameter space where QG assumptions apply, all of the models yield results very similar to the SW results, as expected. In addition, the three balanced models (including QG) perform quite well when $\beta L/f_0$ is $O(1)$, which is outside of the quasigeostrophic regime of validity. This is a somewhat surprising result; however, the solution is dominated by

Rossby waves, and linear theory indicates that the wave speeds are nearly the same for the QG and GPV models for the parameters considered. When there are large layer thickness deviations, in this case due to topography, the GPV model performs significantly better than the QG approximation. Moreover, the GPV approximation performs comparably with the Charney balance model in these cases.

c. Wind-driven box geometry

To further examine the appropriateness of the GPV model for modeling large-scale, low Rossby number flows, we performed a series of numerical simulations in a closed, rectangular domain with wind stress forcing. Along with the GPV model, we tested the quasigeostrophic and planetary geostrophic (PG) formulations and compared the results to shallow-water integrations. PG was included because it is appropriate for scales much larger than the deformation scale. It proved prohibitively complex to obtain solutions of the Charney balance model or of a semigeostrophic model for this geometry. For both models severe difficulties arise in the proper implementation of boundary conditions, and this prevented a meaningful comparison of these models with the shallow-water or GPV model.

The domain used is a closed basin with sides $L = 4 \times 10^8$ cm. We choose $f_0 = 9.375 \times 10^{-5} \text{ s}^{-1}$ and $\beta = 1.75 \times 10^{-13} \text{ (cm s)}^{-1}$, corresponding to a central latitude of 40° . Nondimensionally, $\beta L/f_0 \approx 0.75$ so that the variation in f is nearly as large as the mean value. The equilibrium fluid depth is $H_o = 5 \times 10^4$ cm, while the reduced gravity is $g' = 3.0 \text{ cm s}^{-2}$. We use $U = 1.0 \text{ cm s}^{-1}$ as a scaling velocity, though the actual magnitude of the interior flow depends on the size of the forcing. The fluid is assumed to have a constant density of 1.0 g cm^{-3} . These parameters are similar to those used by Rhines and Schopp (1991) in a multilayer QG simulation. We again use the nondimensional form of the equations so that the domain is defined by $0 \leq x \leq 1$ and $-1/2 \leq y \leq 1/2$. The Rossby number is $\epsilon = 2.7 \times 10^{-5}$, so the flow is expected to be in geostrophic balance.

When comparing models such as these, small differences in the formulation of the forcing or boundary conditions can amplify and distort the solutions to such an extent that meaningful comparison becomes difficult. To minimize these the forcing between the models is made as similar as possible, and for the SW model this is facilitated by including the Ekman damping and the wind stress forcing in the continuity equation of the SW model. In nondimensional form this is written as

$$\frac{dh}{dt} + hD = \mathcal{F}, \quad (3.3)$$

where \mathcal{F} is a general forcing term. This differs from the conventional practice of including the forcing in the momentum equations (e.g., Jiang et al. 1995), but it is

convenient here so that the potential vorticity equation for the shallow-water implementation has the same form as the PV-based models. Moreover, it is physically justifiable if we treat the forcing as a mechanism for inducing vortex squeezing/stretching via the top and bottom Ekman layers. To obtain the potential vorticity equation, we substitute (3.3) into (2.14) and divide through by ϵ to yield

$$\frac{dQ}{dt} = \frac{1 \nabla^2 \zeta}{h \text{ Re}} - \frac{[\zeta + (1 + \epsilon \beta_0 y)/\epsilon] \mathcal{F}}{h^2}. \quad (3.4)$$

We choose

$$\mathcal{F} = r\zeta - A_r \mathbf{k} \cdot (\nabla \times \boldsymbol{\tau}), \quad (3.5)$$

where r is a bottom friction damping coefficient, $\mathbf{k} \cdot (\nabla \times \boldsymbol{\tau})$ is the vertical component of the curl of the wind stress, which is denoted as $\boldsymbol{\tau}$, and A_r is a nondimensional amplitude. For the GPV model, (3.4) is directly applicable, while for the PG formulation, we neglect relative vorticity so that the governing equation becomes

$$\frac{d}{dt} \left(\frac{(1 + \epsilon \beta_0 y)/\epsilon}{h} \right) = \frac{1 \nabla^2 \zeta}{h \text{ Re}} - \frac{[(1 + \epsilon \beta_0 y)/\epsilon] \mathcal{F}}{h^2}. \quad (3.6)$$

For the QG model, the governing equation is

$$\frac{d}{dt} (\nabla^2 \eta + \beta_0 y - F\eta) = \frac{\nabla^4 \eta}{\text{Re}} - \frac{\mathcal{F}}{\epsilon}. \quad (3.7)$$

The shallow-water model simply replaces (2.7) with (3.3).

For both the QG and GPV models, the requirements of no normal flow at the boundary and of mass conservation can be satisfied with appropriate boundary conditions. For the PG model, however, it is not generally possible to conserve mass in the forced case. (Conservation of mass is not a problem *per se* with the PG equations; the adiabatic equations conserve mass, energy, and potential vorticity. The problem only arises here through using a forcing that is analogous to that used in the other models.) We choose to assess the effect of lack of mass conservation on the subsequent solutions. In fact, given a forcing in the continuity equation, the shallow-water equations will not generally conserve mass unless the integral of \mathcal{F} over the domain is 0. For double-gyre wind forcing, $\int \mathbf{k} \cdot (\nabla \times \boldsymbol{\tau}) dA$ is zero, but $\int \zeta dA$ does not necessarily vanish if free-slip boundary conditions are used. Nonetheless, for these simulations the mass fluctuates by only 0.1%, so that we do not expect lack of strict mass conservation to have a noticeable effect on the dynamics.

A series of numerical experiments were performed with double-gyre wind forcing such that

$$\boldsymbol{\tau} = -\cos[2\pi(y + 1/2)] \mathbf{i} \quad (3.8)$$

so that

$$\mathbf{k} \cdot (\nabla \times \boldsymbol{\tau}) = -2\pi \sin[2\pi(y + 1/2)]. \quad (3.9)$$

Table 2 shows the parameters used in the various

TABLE 2. Numerical runs performed for wind-driven closed domain.

Expt	A_τ^* (dyn cm^{-2})	A_τ	r	Re
b1	0.20	1600	0.004	8
b2	0.40	3200	0.004	8
b3	0.60	4800	0.004	8
b4	0.80	6400	0.004	8
b5	0.90	7200	0.004	8
b6	0.875	7000	20	40

numerical experiments, along with the corresponding dimensional wind stress amplitude for comparison [denoted as A_τ^* , where $A_\tau^* = (U^2 H_o/L)A_\tau$]. For all the runs performed, the bottom drag parameter r was effectively set to zero, leaving A_τ and Re as the two adjustable parameters. For most of the runs, Re was set to 8 ($\nu = 5 \times 10^7 \text{cm}^2 \text{s}^{-1}$) and A_τ was successively increased in order to test the models in a regime of larger layer-thickness variations. The largest value of wind stress ($A_\tau = 7200$) brings the layer very near to outcropping (i.e., zero layer thickness) in the shallow-water model. In one run (expt b6), Re was set to 40 ($\nu = 1 \times 10^7 \text{cm}^2 \text{s}^{-1}$) in order to induce time-dependent solutions.

For these experiments we are mainly interested in examining the larger-scale structure of the solutions, and simple measures of model performance serve to illustrate any differing results. We average the layer thickness zonally over the left quarter of the domain, where the gyres are most prominent, in order to obtain a mean height versus y . An average thickness is used in lieu of pointwise extrema in order to yield a more robust estimate of the gyre strengths. Figure 5 summarizes the behavior of the four models for the five numerical simulations. The average maximum and minimum layer thicknesses, denoted respectively as $\langle h \rangle_{\max}$ and $\langle h \rangle_{\min}$, are plotted in for each of the numerical runs.

In experiments b1 through b5, all of the model solutions exhibit steady-state behavior owing to the relatively large viscosity. Figure 5 shows that, as the forcing is increased, the SW subtropical gyre height is well predicted by both the QG and GPV models. The GPV formulation, however, yields a much better estimate of the true subpolar gyre height (as given by the SW model) than does the QG approximation. In experiments b1–b3, PG also does quite well and is nearly as accurate as GPV. However, in experiments b4 and b5, PG predicts outcropping and thus no results were obtained. The failure of the PG model is likely due to its lack of mass conservation. There is a systematic loss of mass as A_τ is increased, which results in a bias toward smaller h .

Another shortcoming of the QG solutions is that, for these runs, the height field is antisymmetric about the mean value of y , so that the subtropical and subpolar gyres possess equal strengths. Cessi and Ierley (1995) have found that for strong forcing and/or small dissipation, stable asymmetric solutions may exist. In our

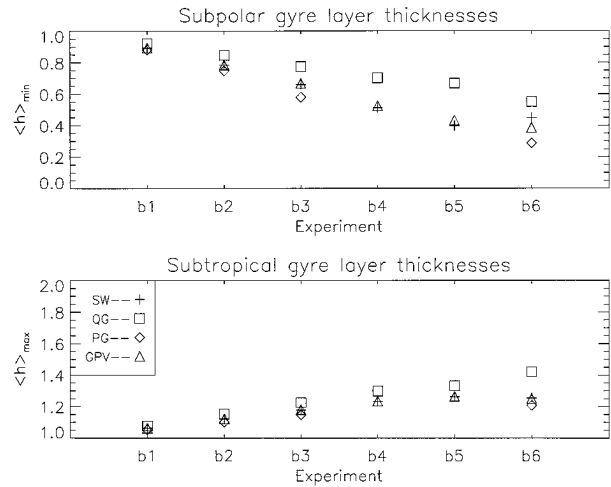


FIG. 5. Average minimum and maximum layer thicknesses for the numerical experiments listed in Table 2. The plus signs, squares, diamonds, and triangles denote solutions from the shallow-water, quasigeostrophic, planetary geostrophic, and GPV formulations, respectively. No planetary geostrophic solution was obtained in experiments b4 and b5.

study, however, only experiment b6 yields an asymmetric solution, and even in this situation the symmetry breaking is quite weak. Conversely, the PG and GPV formulations exhibit the asymmetric gyre solution observed in the SW model for all the numerical runs.

Figure 6 displays contour plots of the height fields for the models when $A_\tau = 4800$ (expt b3). There is very good agreement between the SW and GPV solutions. The PG solution exhibits the gyre asymmetry but otherwise captures only the gross features of the full solution. In particular, mesoscale features such as the meandering of the intergyre jet are not present—this is not unexpected since the PG is formally valid only for large-scale flow. The QG solution, which is symmetric about the mean value of y , shows reasonable agreement in the subtropical gyre but is missing key features present in the SW subpolar gyre. Figure 7 shows the SW, QG, and GPV height fields when $A_\tau = 7200$ (expt b5) and the subpolar gyre in the SW model is close to outcropping. The gross structure of the GPV solution again looks very similar to the SW result, though there are some noticeable differences near the western boundary. In particular, the confluence point at which the two gyres meet near the western edge is moved south of the meridional midpoint in the GPV solution. This has also been observed in SW simulations (Jiang et al. 1995) and occurs as a result of an imperfect pitchfork bifurcation in parameter space. Most likely the bifurcation structures of the SW and GPV models are slightly different, thus predisposing the GPV model toward a less symmetric solution for a given forcing. The QG solution again exhibits symmetric gyre strengths and is quite dissimilar from the SW field.

For experiment b6 at higher Reynolds number, all four

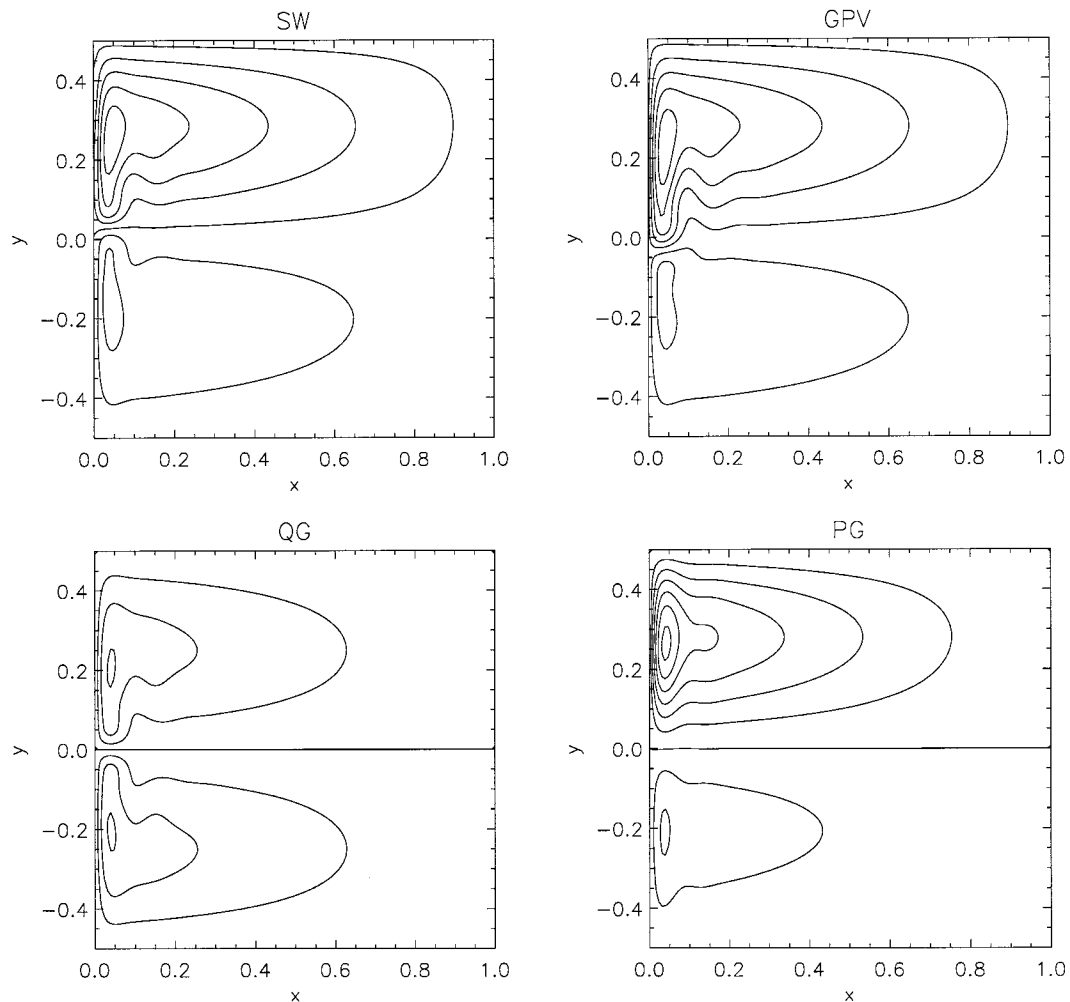


FIG. 6. Steady-state height fields of SW, GPV, PG, and QG model solutions for $A_v = 4800$ and $Re = 8$ (expt b3).

models exhibit time-dependent, aperiodic behavior. Accordingly, we plot $\langle \bar{h} \rangle_{\min}$ and $\langle \bar{h} \rangle_{\max}$, where \bar{h} is a time-averaged height field. Figure 5 shows that the PG and GPV formulations both give good estimates of the subtropical gyre strength, while both slightly underestimate the thickness in the subpolar region. The QG solution is again quite poor but now gives a better estimate of the subpolar gyre, grossly overestimating the subtropical circulation. Figure 8 shows the time-averaged height fields for all four models from experiment b6. Again, the large-scale features of the SW and GPV solutions are in very good agreement, while the mesoscale features are similar but do exhibit some differences. The PG formulation captures the large-scale gyre strengths but has no mesoscale features. The QG solution misrepresents the large-scale circulation, and its mesoscale dynamics consist of strong recirculation gyres reminiscent of a Fofonoff solution, which is not observed in the SW case. The kinetic energy evolution was examined for each model in order to ascertain the dominant frequency in each case. The Fourier spectra of the time

series are rather broadband, reflecting the quasi-turbulent nature of the solutions, so the timescales obtained are gross estimates of the dominant period of the flow. The SW model has a dominant period of about 2.4 years, while the GPV and PG models oscillate with periods of about 1.7 and 1.5 years, respectively. Moreover, the strength of the SW oscillation is about 50% larger than that of the GPV and PG models. The QG solution possesses weak oscillations on similar timescales, but its dominant period of vacillation is about 42 years. This appears to arise as a result of competition between the “Fofonoff” solution and another solution with strong mesoscale eddy dynamics. The GPV and SW models, while exhibiting similar behavior, do however show some differences as noted above. Further simulations reveal that the behavior of both models is quite sensitive to the magnitude of the dissipation, indicating that we are likely near bifurcation points in parameter space. We therefore speculate that differences in the bifurcation structure of the two models result in slightly different dynamics, and this is probably exacerbated by the dif-

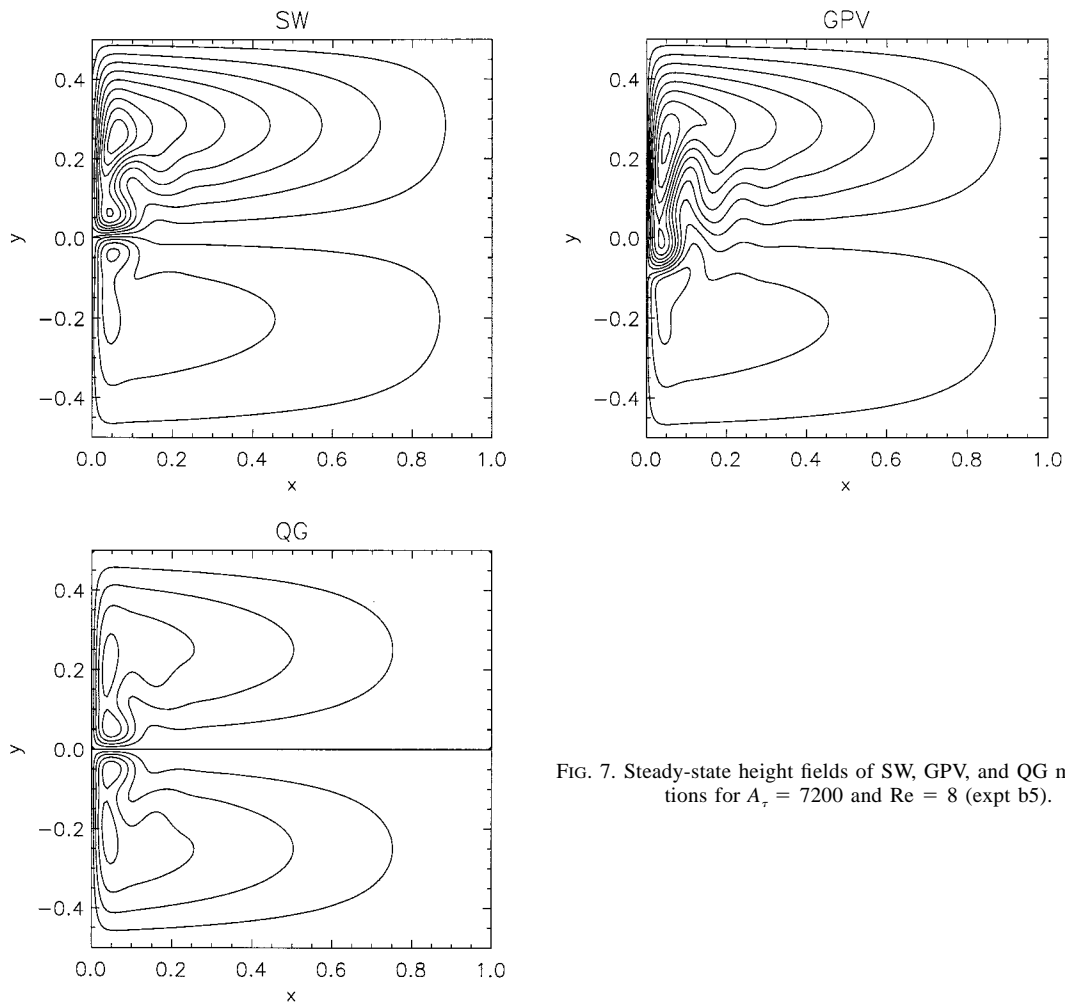


FIG. 7. Steady-state height fields of SW, GPV, and QG model solutions for $A_\tau = 7200$ and $Re = 8$ (expt b5).

ferent methods for specifying the height field at the boundaries.

4. PV inversion

The GPV model and, to a somewhat lesser degree, the PG model perform significantly better than quasi-geostrophy in nearly all the wind-driven experiments done. We can better determine the cause of this and also further examine the relative merits of the various balanced models by inverting the potential vorticity of the shallow-water model. This isolates the accuracy of the inversion formula, thereby avoiding any problems that the time-dependent dynamics might have. Thus, given the potential vorticity Q (2.17) from the SW model, we “invert” this quantity to obtain the corresponding height field using various assumptions regarding how the fields are balanced. Specifically, for the GPV and PG formulations, we solve (2.30) and (2.41) respectively (with $h_b = 0$) for the interface deviation η . The full quasi-geostrophic potential vorticity is similar to (2.26) and is defined as

$$Q = \nabla^2 \eta - F\eta + \frac{1 + \epsilon\beta_0 y}{\epsilon}. \tag{4.1}$$

In each case, the Q field is given from the SW model. The various η fields obtained from the inversions can then be compared to the corresponding SW result.

To more closely examine the GPV formulation we define two other inversions. The first one, denoted the first modified GPV model (MGPV1), ignores local geostrophy but makes no approximations in the height field. The potential vorticity is thus

$$Q = \frac{\nabla^2 \eta + (1 + \epsilon\beta_0 y)/\epsilon}{1 + \epsilon F \eta}. \tag{4.2}$$

The second inversion, which is called the second modified GPV model (MGPV2), retains local geostrophy but linearizes the height field, as does the QG approximation. In this case, the potential vorticity is

$$Q = \frac{\nabla^2 \eta}{1 + \epsilon\beta_0 y} - \frac{\epsilon\beta_0}{(1 + \epsilon\beta_0 y)^2} \frac{\partial \eta}{\partial y} + \frac{1 + \epsilon\beta_0 y}{\epsilon} - (1 + \epsilon\beta_0 y)F\eta. \tag{4.3}$$

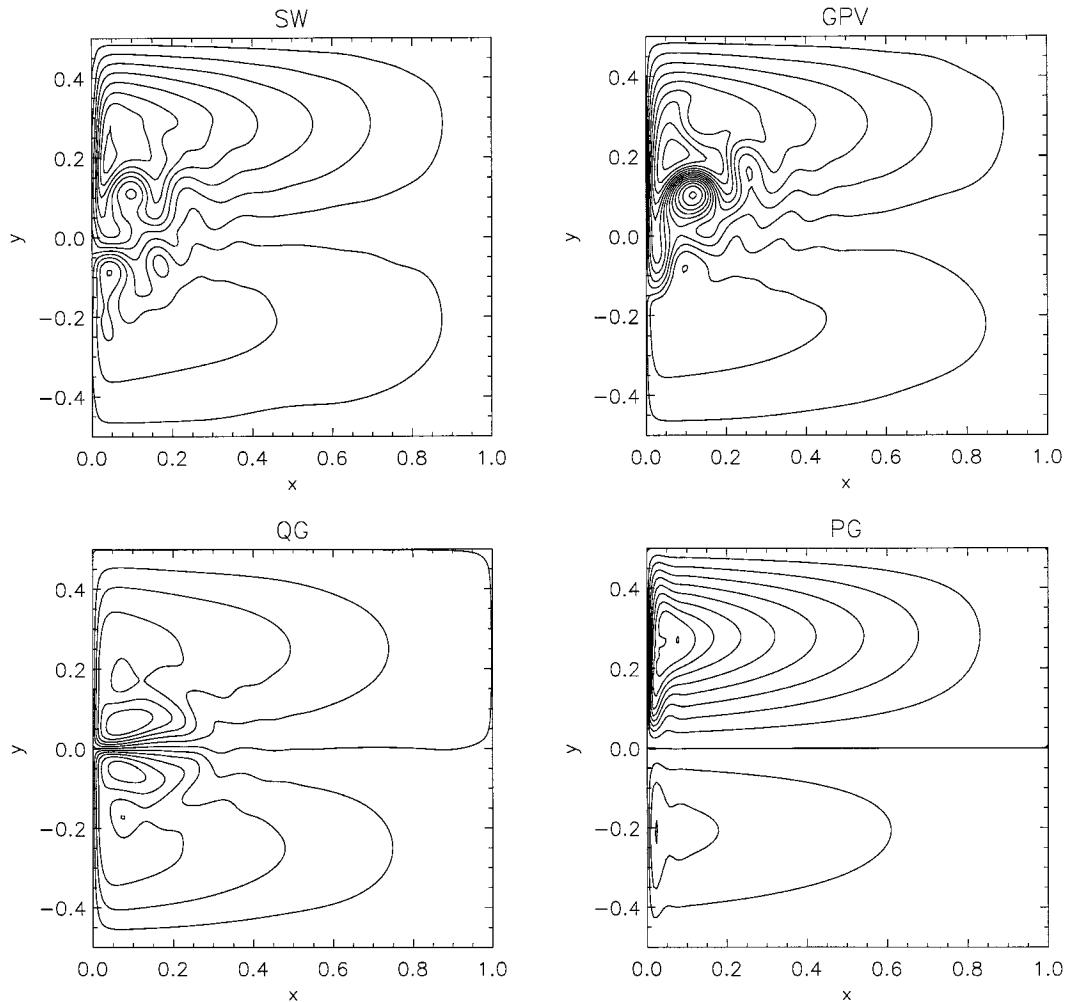


FIG. 8. Time-averaged height fields of SW, GPV, QG, and PG model solutions for $A_r = 7000$ and $Re = 40$ (expt b6).

By independently relaxing the QG assumptions, MGPV1 and MGPV2 allow a direct assessment of the importance of both local geostrophy and a full height field for properly modeling both mesoscale and large-scale dynamics.

The five inversions (PG, GPV, QG, MGPV1, MGPV2) were performed on selected shallow-water Q fields from the experiments discussed in section 3c. Differences between the inverted fields and the SW height field were then computed. Table 3 displays the rms errors in height and vorticity for experiments b1, b3, b5, and b6. The vorticity is obtained a posteriori by differentiating the computed height field.

The GPV approximation consistently performs extremely well in all of the inversions. There is no degradation even for expt b6, where the flow is turbulent and there exists mesoscale as well as large-scale flow. The MGPV1 inversion, which has no local geostrophic balance, does only slightly worse than the GPV formulation. However, the vorticity errors are proportionately higher than the height errors when compared to

the GPV results. Therefore, the inclusion of local geostrophic balance makes a small but measurable difference in the height field, and the effect is amplified for higher-order quantities such as vorticity. The PG approximation also performs quite well. For the steady cases (b1, b3, and b5), where the solution mainly consists of large-scale gyres, the height and vorticity errors are larger but on the same order as the GPV model results. In the turbulent case, the errors are noticeably larger because the PG inversion cannot well represent the mesoscale activity, an important feature of these runs. For all the cases considered, the QG inversion yields poor results. In addition, for the steady solutions the errors grow as the SW height deviations become larger, becoming greater than unity for large forcing (e.g., expt b5). This indicates that the QG inversion is failing largely because of its linearization of the height field in the PV formulation. For experiment b6, the error in the vorticity field is somewhat smaller than that for experiment b3 because the QG inversion is capable of capturing the mesoscale activity near the western bound-

TABLE 3. Rms errors in height and vorticity (in parentheses) for PV inversions of experiments b1, b3, b5, and b6.

Inversion	Expt b1	Expt b3	Expt b5	Expt b6
PG	5.81×10^{-2} (2.70×10^{-1})	5.53×10^{-2} (2.61×10^{-1})	5.23×10^{-2} (2.56×10^{-1})	1.20×10^{-1} (8.04×10^{-1})
GPV	1.90×10^{-2} (1.12×10^{-1})	1.75×10^{-2} (1.05×10^{-1})	1.55×10^{-2} (9.35×10^{-2})	2.10×10^{-2} (7.97×10^{-2})
QC	2.84×10^{-1} (4.41×10^{-1})	6.02×10^{-1} (1.07×10^0)	2.02×10^0 (6.89×10^0)	1.53×10^0 (2.47×10^0)
MGPV1	2.07×10^{-2} (1.56×10^{-1})	1.92×10^{-2} (1.33×10^{-1})	1.72×10^{-2} (1.21×10^{-1})	2.39×10^{-2} (9.04×10^{-2})
MGPV2	8.74×10^{-2} (1.59×10^{-1})	3.72×10^{-1} (6.94×10^{-1})	1.63×10^0 (5.49×10^0)	1.25×10^0 (2.15×10^0)

ary. Finally, the MGPV2 inversion provides a modest improvement to the QG results. When the wind forcing is small, the improvement is largest. This is to be expected, since in this case the linearization of the height field is less suspect than the global geostrophy condition. For large wind forcing (expt b5), there is very little difference in the errors.

Figure 9 shows the inverted, time-mean height fields for experiment b6. The GPV and MGPV1 inversions are visually indiscernible from the SW height field. The PG inversion captures the large-scale features well but exhibits some differences in the mesoscale eddies. The QG and MGPV2 inversions both look quite poor on the large scale, as is evident from the excessive height deviation near $x = 0$, $y = 0.2$. However, the mesoscale eddies are fairly well replicated by both approximations.

The inversions clearly demonstrate the importance of the various terms in the balanced models. Retention of the full height field in the governing equation is the most important factor in accurately modeling flow with large height deviations. The inclusion of local geostrophic balance, while less critical, also provides measurable quantitative improvement in the solutions.

5. Energetics

One potential drawback of the GPV model, and indeed of many balanced models, is the lack of exact energy conservation in the adiabatic case (i.e., in the absence of forcing and dissipation). For example, neither the balanced model of Charney (1962) nor the linear balance model (Gent and McWilliams 1983) conserve energy. The shallow-water equations and the quasi-geostrophic and planetary geostrophic equations do conserve energy in the adiabatic case. We monitored the kinetic and potential energy for the channel runs of section 3b. The kinetic and potential energies of the GPV model are evaluated in the same way as for the shallow-water model, except that the geostrophic velocities are used in the expression for kinetic energy. Thus,

$$KE = \frac{1}{2} \int h(u^2 + v^2) dA \quad (5.1)$$

$$PE = \frac{1}{2} \int \eta^2 dA, \quad (5.2)$$

where h is the total layer thickness, η the deviation height, and u and v are the velocities, taken as geostrophic for the GPV model. [See Gent (1993) for a discussion of the consistent energetics of the height-forced shallow-water equations.] In experiment c1, both the GPV and CB models conserve energy to within a few percent of the mean value. In the remainder of the experiments, the CB formulation tended to lose kinetic energy (and thus overall energy) with time. The GPV model, on the other hand, gains kinetic and total energy in experiments c2 and c4, and in experiment c3 the total energy oscillates with an amplitude about 10% of the mean value. Figure 10 shows the kinetic, potential, and total energy versus time for the GPV model in experiment c2. In this run, the potential energy is approximately constant with time, but the kinetic energy increases by about 50% during the integration. For this experiment, the CB model *loses* a similar amount of kinetic energy while its potential energy also remains constant.

These findings seem somewhat incongruous given the good correspondence of both models with the SW results. Further examination reveals that the anomalies seem to be generated at the north/south walls where boundary conditions must be applied. In the absence of dissipation, the momentum equations imply that the circulation at each wall must be conserved. For the QG model, this allows the zonally averaged height field to be uniquely determined at the boundaries (McWilliams 1977). However, it is not clear that this condition should apply to models that are not based on momentum equations. Also, when lateral viscosity is present (which is usually the case), circulation is not conserved at the walls and a different boundary condition must be employed.

That boundary effects may be important to energy conservation was tested by performing an unforced in-

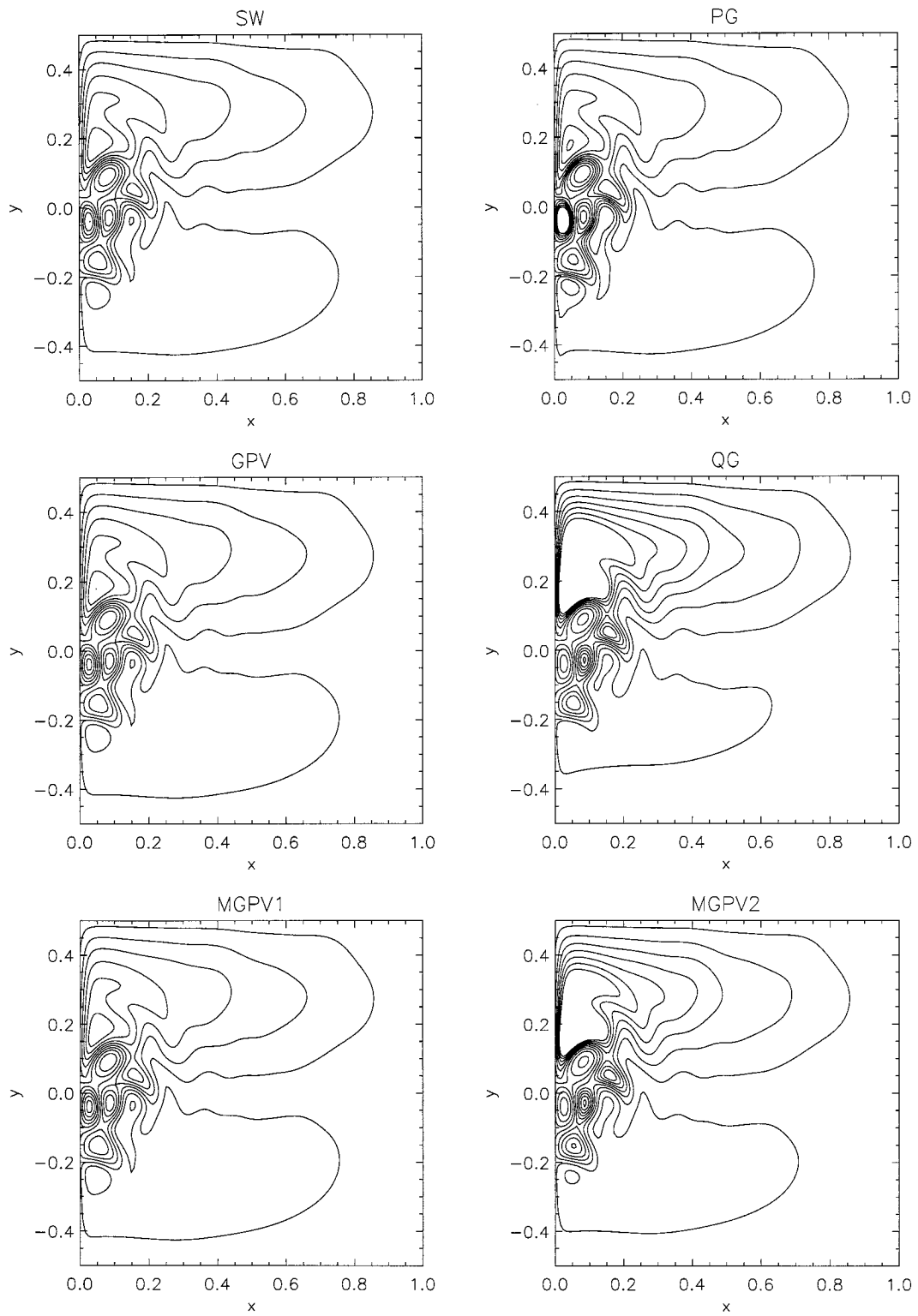


FIG. 9. Inversions for experiment b6.

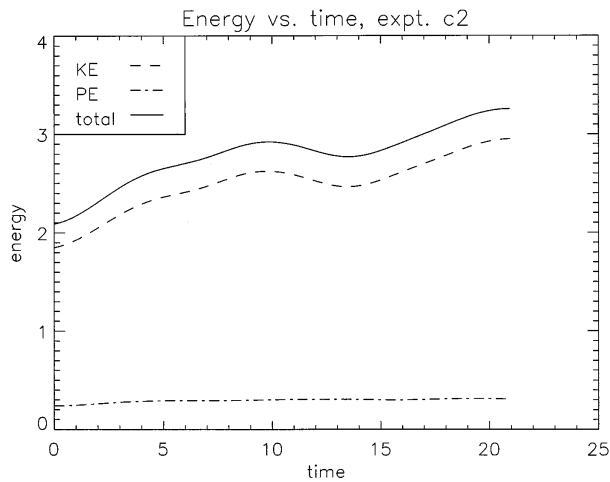


FIG. 10. Energy vs time for GPV model, experiment c2.

viscid simulation in the closed box domain. Two vortices with heights equal to about one-fourth of the mean fluid depth were placed in the northeast and southeast corners of the domain. These propagated westward until they collided with the western boundary, at which point they broke up rapidly into grid-scale noise. Figure 11 shows the total energy versus time for this case. The arrow shows the time at which the vortices first collide with the western boundary. The total energy is very nearly conserved until the collision occurs, and afterward the fluctuations are only about 10% of the mean value, even after the grid-scale noise becomes prominent. Therefore, while the GPV formulation does not strictly conserve total energy, it approximately does so. This result also indicates that the energy conservation (or lack thereof) is likely in part related to the applied boundary conditions.

We also monitored the energetics for the wind-driven simulations. When the GPV model is forced and dissipated, its energy cycle behaves very similarly to that of SW. Figure 12 shows the kinetic and potential energy versus time for the SW and GPV models in experiment b1. The models are initialized with $\eta = 0$. It is evident that, in this case, the equilibration of the GPV model behaves very much like its shallow-water counterpart, even though only the latter has an exact energy conservation property. In cases of more practical interest, then, where forcing and dissipation are important ingredients of the system, energy diagnostics can be performed and an energy budget evaluated, even in the absence of an exact energy conservation principle. The growth and decay of energy in the GPV model, even in time-dependent calculations, is very similar to that of the shallow-water model.

6. Discussion

The numerical results indicate that the geostrophic potential vorticity formulation is an appropriate and ac-

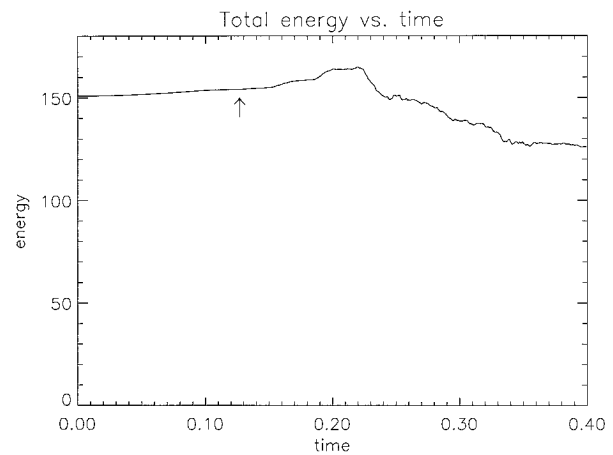


FIG. 11. Total energy vs time for unforced, inviscid simulation of GPV model in closed box domain. The arrow marks the time at which the vortices first collide with the western boundary.

curate model for describing both large- and synoptic-scale balanced flows. It encompasses quasigeostrophic (QG) dynamics and also extends to situations outside the formal range of validity of quasigeostrophic theory, namely, where the Coriolis parameter and layer thicknesses are allowed to vary by an $O(1)$ amount. Equivalently, it encompasses planetary-geostrophic (PG) dynamics and extends into a parameter regime where relative vorticity is important, such as the oceanic meso-scale, outside of the range of planetary-geostrophic theory.

By performing inversions of the SW potential vorticity field, it is found that for low Rossby number large-

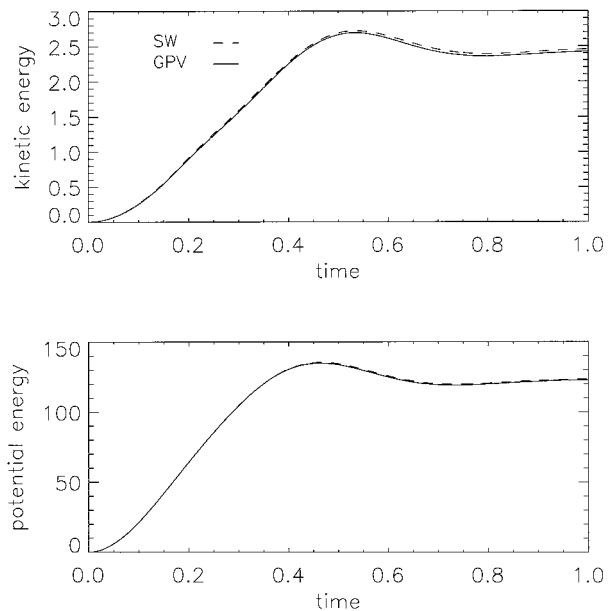


FIG. 12. Growth of potential and kinetic energy in experiment b1 for SW and GPV models. The models are initially started from rest.

scale flow typifying planetary gyres, it is particularly important that a balanced model should allow full height variations in the potential vorticity equation (whereas the height variation is linearized in the QG formulation). Secondly, local geostrophic balance provides an increase in accuracy over a global geostrophic (constant Coriolis parameter) form. We note that the GPV model does not, in our numerical simulations, exactly conserve energy in the absence of external forcing and dissipation. However, for most cases of practical interest, when forcing and dissipation are present, even at small levels, the GPV potential and kinetic energies closely mimic the SW results, implying that lack of energy conservation is not a substantial practical drawback to utilizing potential vorticity-based balanced models. The theoretical lack of energy-based diagnostics (including wave activities and energy cycles) may be a more serious issue, although appropriate approximations to these may be constructed. One other topic deserving of a further exploration is the treatment and effects of boundary conditions. For example, the consistency boundary condition of quasigeostrophic dynamics (McWilliams 1977) seems to effect a parameterization of boundary-trapped gravity waves (Milliff and McWilliams 1994); whether an analogous effect occurs in the model presented here is a topic of further research.

Perhaps the single most important overall conclusion to be drawn from this study is that demonstrably balanced flow exists, and can be modeled, that is well outside of the parameter regime described by just the quasigeostrophic, or just the planetary geostrophic equations. The dynamics of both the midlatitude oceanic mesoscale and large-scale can be accurately described by the advection of a geostrophic potential vorticity by a geostrophic velocity field: A model containing these dynamics and little else (i.e., the GPV model) gives a very accurate rendition of the shallow-water equations in the low Rossby number limit. This implies that the dynamics of some fundamental but still ill-understood low Rossby number processes, such as western boundary current separation or the recirculation regime of the subtropical gyre, can be addressed with such a model.

Note that the GPV model is of no formal higher asymptotic order than either PG or QG, and it reduces to these models in appropriate limits—the planetary scale and mesoscale, respectively. In some sense the model might be considered a higher-order correction of PG since it does keep a higher-order term, namely the relative vorticity, in the expression for potential vorticity. However, this term is, in fact, leading order for synoptic-scale flow, and furthermore the advecting velocity is not expanded beyond leading order at either scale. See Vallis (1996, §§ 2d and 2e) for further discussion. Formally more accurate models, such as discussed by Browning and Kreiss (1987), Allen (1993), and Warn et al. (1995) will give improved results provided the dynamics stays in a regime in which quasigeostrophic dynamics is already appropriate. Our (complementary) goal in this

paper is to extend the parameter regime wherein a balanced model may be used, rather than to achieve higher accuracy in a particular dynamical regime. No assumption save that Rossby number is small is used in deriving the dynamics of the GPV model, and thus its range of applicability is relatively large. The Charney balance model does have higher formal accuracy than quasigeostrophy and GPV in a synoptic regime, as well as allowing large height variations. However, its numerical and analytic complexity, as well as its unbalanced nature when the beta effect is large, make it, in our opinion, much less attractive than the GPV model. Salmon's "L1" dynamics (Salmon 1983) are in many ways similar to the GPV model. They include an ageostrophic advection, which allows energy to be conserved, at the cost of a more complex inversion process. It seems likely that their dynamics will be at least as good for large-scale and mesoscale flows.

The efficiency of balanced models versus various flavors of primitive equations is obviously a consideration. If we define the time required for one QG time step in the wind-driven case to be unity, then, for the simulations presented in section 3, the PG, GPV, and SW models require 0.45, 4.4, and 1.4 time units, respectively. The PG model is, not surprisingly, the fastest to implement. The SW model is about three times faster *per time step* than the GPV model, but the time step itself tends to be about an order of magnitude smaller, resulting in a speedup in GPV of about a factor of 3. The GPV model is plainly slower than a comparable QG formulation, mainly because the QG models are able to use a faster elliptic solver to invert potential vorticity, but this ratio is likely to tend toward unity in future as faster elliptic solvers (e.g., parallelizing multigrid) for problems with variable coefficients become available.

What are the implications for large-scale ocean/atmosphere modeling? For the large-scale oceanic flow, the PG model results are, in fact, very good—considerably better than quasigeostrophy, for example—although the PG dynamics fail to capture the eddy motion of the western boundary current, and the mesoscale variability of the ocean cannot be properly captured by such models. But see Samelson and Vallis (1997) for an example of their utility. (The GPV model can be extended in a reasonably straightforward manner to a multilayered configuration; this is particularly easy in isopycnal coordinates.) For the atmosphere, variability of synoptic-scale motion cannot be neglected, since both midlatitude variability and poleward heat transport in midlatitudes are dependent on eddy motion. For the large-scale atmospheric flow, therefore, a GPV model is to be preferred over either a PG model, which has no proper synoptic variability, or a QG model, which incorrectly models large-scale changes in potential vorticity.

Acknowledgments. This work was funded by NSF (ATM 93-9317485 and OCE 94-15512). We are grateful for informative comments from Ralph Milliff.

APPENDIX

Frontal Geostrophic Dynamics from Potential Vorticity Slaving¹

So-called frontal geostrophic dynamics (Cushman-Roisin 1986) are the leading order nontrivial dynamics for shallow-water motions at a scale large compared to the deformation radius but small compared to the scale over which the Coriolis parameter varies by an order-one amount. The layer thickness is allowed to vary by an order-one amount. Here we give a compact derivation of these equations by supposing that the motion is determined by potential vorticity advection and inversion. To obtain the inversion formula, all the variables are expanded in terms of the small parameter (Rossby number) except the slaving field which, because it is given, is lowest order. The nondimensional evolution equation is

$$\frac{Dq}{Dt} = 0, \tag{A.1}$$

where

$$q = \frac{\epsilon \zeta + f_0}{h} \tag{A.2}$$

in which ϵ is the Rossby number and $\zeta(x, y, t)$, f_0 , and $h(x, y, t)$ are the nondimensional vorticity, constant Coriolis parameter, and layer thickness, respectively (the parameter f_0 could be taken as unity).

If ζ and h are expanded in powers of ϵ , then (A.2) becomes

$$\begin{aligned} q &= \frac{f_0}{h_0} \\ &= \frac{f_0}{h_0} + \epsilon \left\{ \frac{\zeta_0}{h_0} - \frac{h_1 f_0}{h_0^2} \right\} \end{aligned} \tag{A.3}$$

to zeroth and first order respectively. Since q is a lowest-order quantity and is equal to f_0/h_0 , the coefficient of ϵ in (A.3) must vanish (Warn et al. 1995), implying

$$h_1 = \frac{\zeta_0 h_0}{f_0} = \frac{h_0 \nabla^2 h_0}{f_0^2}, \tag{A.4}$$

using $\zeta_0 = \nabla^2 h_0 / f_0$.

The evolution equation (A.1) then becomes

$$\frac{\partial f_0}{\partial t h_0} + \mathbf{v}_0 \cdot \nabla \frac{f_0}{h_0} + \mathbf{v}_1 \cdot \nabla \frac{f_0}{h_0} = 0. \tag{A.5}$$

The first advective term $\mathbf{v}_0 \cdot \nabla (f_0/h_0)$ vanishes identically. Thus, the resulting evolution occurs on a timescale $O(\epsilon)$ slower than an advective timescale. To obtain \mathbf{v}_1 , use the leading-order momentum equation:

$$\mathbf{v}_0 \cdot \nabla \mathbf{v}_0 + f_0 \mathbf{k} \times \mathbf{v}_1 = -\nabla h_1, \tag{A.6}$$

where the partial time derivative is neglected because the evolution timescale is $O(\epsilon)$ smaller than the advective timescale. Then casting the advection term in Bernoulli form, and using (A.4), gives

$$f_0 \mathbf{v}_1 = \mathbf{k} \times \nabla \left(\frac{v_0^2}{2} + \frac{\zeta_0 h_0}{f_0} \right) - \zeta_0 \mathbf{v}_0. \tag{A.7}$$

Using (A.7) in (A.5), we obtain

$$\frac{\partial q}{\partial t} + \frac{1}{f_0^3} J \left(\frac{1}{2} (\nabla h_0)^2 + h_0 \nabla^2 h_0, q \right) = 0, \tag{A.8}$$

where $q = f_0/h_0$.

The corresponding height evolution equation may be written

$$\frac{\partial h_0}{\partial t} + \frac{1}{f_0^3} J \left(\frac{1}{2} (\nabla h_0)^2 + h_0 \nabla^2 h_0, h_0 \right) = 0. \tag{A.9}$$

The inclusion of a beta-effect is straightforward, although the scaling allows only $\beta L/f = O(\epsilon)$. This is because if β is not small the advection by the geostrophic wind does not vanish, and the PG equations result at lowest order.

REFERENCES

Allen, J. S., 1993: Iterated geostrophic intermediate models. *J. Phys. Oceanogr.*, **23**, 2447–2461.

—, J. A. Barth, and P. A. Newberger, 1990a: On intermediate models for barotropic continental shelf and slope flow fields. Part I: Formulation and comparison of exact solutions. *J. Phys. Oceanogr.*, **20**, 1017–1042.

—, —, and —, 1990b: On intermediate models for barotropic continental shelf and slope flow fields. Part III: Comparison of numerical model solutions in periodic channels. *J. Phys. Oceanogr.*, **20**, 1949–1973.

Arakawa, A., 1966: Computational design for long-term numerical integrations of the equations of atmospheric motion. *J. Comput. Phys.*, **1**, 119–143.

—, and V. R. Lamb, 1981: A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Wea. Rev.*, **109**, 18–36.

Asselin, R. A., 1972: Frequency filter for time integrations. *Mon. Wea. Rev.*, **100**, 487–490.

Barth, J. A., J. S. Allen, and P. A. Newberger, 1990: On intermediate models for barotropic continental shelf and slope flow fields. Part II: Comparison of numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.*, **20**, 1044–1076.

Bleck, R., 1973: Numerical forecasting based on the conservation of potential vorticity on isentropic surfaces. *J. Appl. Meteor.*, **12**, 737–752.

Browning, G., and H. Kreiss, 1987: Reduced systems for the shallow water equations. *J. Atmos. Sci.*, **44**, 2813–2822.

Cessi, P., and G. R. Ierley, 1995: Symmetry-breaking multiple equilibria in quasigeostrophic, wind-driven flows. *J. Phys. Oceanogr.*, **25**, 1196–1205.

Charney, J. G., 1962: Integration of the primitive and the balance equations. *Proc. Int. Symp. Numerical Weather Prediction*, Tokyo, Japan, Meteor. Soc. Japan, 131–152.

Cushman-Roisin, B., 1986: Frontal geostrophic dynamics. *J. Phys. Oceanogr.*, **16**, 132–143.

¹ This appendix was written with Dr. T. Warn of McGill University, Montreal, Canada.

- Gent, P. R., 1993: The energetically consistent shallow-water equations. *J. Atmos. Sci.*, **50**, 1323–1325.
- , and J. C. McWilliams, 1983: Consistent balanced models in bounded and periodic domains. *Dyn. Atmos. Oceans*, **7**, 67–93.
- Hoskins, B. J., 1975: The geostrophic momentum approximation and the semigeostrophic equations. *J. Atmos. Sci.*, **32**, 233–242.
- Jiang, S., F-F Jin, and M. Ghil, 1995: Multiple equilibria, periodic, and aperiodic solutions in a wind-driven, double-gyre shallow-water model. *J. Phys. Oceanogr.*, **25**, 764–786.
- McWilliams, J. C., 1977: A note on a consistent quasi-geostrophic model in a multiply connected domain. *Dyn. Atmos. Oceans*, **1**, 427–441.
- , N. J. Norton, P. R. Gent, and D. B. Haidvogel, 1990: A linear balance model of wind-driven midlatitude ocean circulation. *J. Phys. Oceanogr.*, **20**, 1349–1378.
- Mesinger, F., and A. Arakawa, 1976: Numerical methods used in atmospheric models. GARP Publ., 64 pp.
- Milliff, R., and J. McWilliams, 1994: The evolution of boundary pressure in ocean basins. *J. Phys. Oceanogr.*, **24**, 1317–1338.
- Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. Springer-Verlag, 710 pp.
- Phillips, N. A., 1963: Geostrophic motion. *Rev. Geophys.*, **1**, 123–176.
- Rhines, P. B., and R. Schopp, 1991: The wind-driven circulation: Quasi-geostrophic simulations and theory for nonsymmetric winds. *J. Phys. Oceanogr.*, **21**, 1438–1469.
- Robinson, A. R., and H. Stommel, 1959: The oceanic thermocline and the associated thermohaline circulation. *Tellus*, **11**, 295–308.
- Salmon, R., 1983: Practical use of Hamilton's principle. *J. Fluid Mech.*, **132**, 431–444.
- Samelson, R., and G. K. Vallis, 1997: Large-scale circulation with small diapycnal diffusivity: The two-thermocline limit. *J. Mar. Res.*, **55**, 1–54.
- Vallis, G. K., 1996: Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. *Quart. J. Roy. Meteor. Soc.*, **122**, 291–322.
- Warn, T., and R. Menard, 1986: Nonlinear balance and gravity-inertial wave saturation in a simple atmospheric model. *Tellus*, **38A**, 285–294.
- , O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby number expansions, slaving, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121**, 723–739.
- Welander, P., 1959: An advective model of the ocean thermocline. *Tellus*, **11**, 309–318.